

Pengenalan Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan Arsitektur Mobilenet

Alfian Rudiyanto*¹, Yufis Azhar¹
Universitas Muhammadiyah Malang
alfianrudiyanto1@webmail.umm.ac.id*

Abstrak

Sistem Isyarat Bahasa Indonesia atau lebih dikenal dengan SIBI merupakan Bahasa resmi yang digunakan penyandang tuli dalam berkomunikasi. Namun, permasalahan muncul karena sebagian besar orang tidak memahami bahasa isyarat. Untuk itu diperlukan perantara alternatif yang dapat menjadi penerjemah antara penyandang tuli dengan masyarakat biasa. Penelitian ini bertujuan untuk mengklasifikasikan SIBI menggunakan metode Convolution Neural Network dengan MobileNetV2. Data memiliki total 24 kelas, dengan pengecualian huruf J dan Z. Permasalahan lain muncul ketika memiliki data yang sedikit, yaitu 220 gambar setiap kelas. Karena itu diperlukan data augmentasi sehingga dataset menjadi lebih banyak. Pada penelitian ini akan menerapkan rotasi, shear, dan flip pada setiap gambar sehingga membuat gambar baru. Hasil penelitian ini menunjukkan dengan penambahan data, model menjadi lebih optimal. Untuk pengujian akan menggunakan data test, yang menghasilkan akurasi sebesar 98,54 melalui penggunaan batchsize sebesar 32 dan menjalankan proses training sebanyak 25 epoch.

Kata Kunci: Bahasa Isyarat, Sistem Isyarat Bahasa Indonesia, Convolutional Neural Network, Klasifikasi, Data Augmentasi, MobileNetV2

Abstract

The Indonesian Sign Language System or better known as SIBI is the official language used by deaf people in communication. However, problems arise because most people do not understand sign language. For this reason, an alternative intermediary is needed that can become a translator between deaf people and ordinary people. This research aims to classify SIBI using the Convolution Neural Network method with MobileNetV2. The data has a total of 24 classes, with the exception of the letters J and Z. Another problem arises when having little data, which is 220 images per class. Therefore, augmentation data is needed so that the dataset becomes more. This research will apply rotation, shear, and flip to each image to create a new image. The results of this research show that with the addition of data, the model becomes more optimal. For testing, we will use test data, which produces an accuracy of 98.54 through the use of a batchsize of 32 and running the training process for 25 epochs.

Keywords: Sign Language, Indonesian Sign Language System, Convolutional Neural Network, Classification, Data Augmentation, MobileNetV2

1. Pendahuluan

Manusia merupakan makhluk yang selalu melakukan interaksi dengan manusia lainnya. Seorang manusia tidak dapat mencapai apa yang ia inginkan tanpa bantuan dari manusia lain, karena itu komunikasi adalah hal terpenting [1]. Komunikasi mengarah pada proses pertukaran informasi, pikiran, ide, atau emosi antara seseorang. Itu sebabnya kita perlu memiliki kemampuan komunikasi yang baik. Namun, tidak semua dari kita mampu berkomunikasi dengan baik karena sebagian dari kitamengalami kesulitan dalam berkomunikasi [1]. Baik itu karena penyakit atau memang kurang sempurnanya indra mereka, yaitu masyarakat penyandang tuna rungu.

Bahasa isyarat diciptakan untuk membuat hidup masyarakat tuna rungu lebih mudah dalam berkomunikasi di antara mereka. Dalam Bahasa isyarat segala bentuk gestur tubuh bisa digunakan sebagai alat komunikasi seperti kombinasi Gerakan jari, orientasi dan Gerakan tangan, tubuh, ataupun ekspresi wajah [2]. Namun hanya sedikit orang yang mengerti bagaimana cara berkomunikasi dengan bahasa isyarat. Kursus atau pengajaran bahasa isyarat tidak dilakukan secara intensif karena memang menurut data WHO pada tahun 2018, penyandang

tuna rungu hanya memiliki 5% populasi manusia atau 466 juta orang dan 34 juta diantaranya adalah anak-anak [3]. Bahasa isyarat dasar yang mudah dipahami adalah alfabet, namun itu hanya sebagian kecil dari banyaknya bahasa isyarat.

Setiap negara memiliki panduan baku bahasa isyarat [4]. Indonesia menggunakan Sistem Isyarat Bahasa Indonesia (SIBI) yang merujuk pada bahasa isyarat Amerika yaitu American Sign Language (ASL). American Sign Language adalah bahasa isyarat yang paling sering digunakan di seluruh dunia [5]. Bahasa Isyarat Indonesia (BISINDO) merupakan pendoman bahasa isyarat lainnya yang lebih sering digunakan penyandang tuna rungu dari pada SIBI. Namun penggunaan SIBI lebih dikhususkan pada instansi pendidikan khususnya sekolah luar biasa [6]. Perbedaan dari antara SIBI dan BISINDO yaitu BISINDO memerlukan 2 tangan sedangkan SIBI hanya 1 tangan. Bahasa isyarat cukup penting untuk diajarkan dalam pendidikan formal, namun jarang diterapkan dalam lembaga pendidikan formal [6]. Hal tersebut menyebabkan penyandang tuna rungu dan tuna wicara kesulitan dalam berkomunikasi secara normal. Oleh karena itu penting untuk melakukan pengenalan pada bahasa isyarat dengan metode deep learning sehingga informasi bisa tersampaikan dengan baik dan benar.

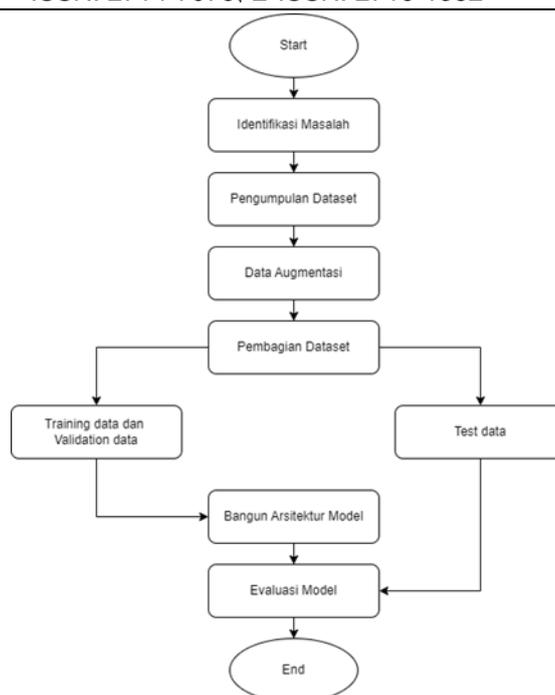
Deep Learning adalah salah satu cabang dari Machine Learning Dengan algoritma penyusun yang terstruktur seperti pola pikir manusia. Metode ini cukup kompleks karena dapat beradaptasi dengan jumlah data yang besar dan dapat memecahkan permasalahan yang tidak dapat diselesaikan dengan Machine Learning. Deep Learning sangat baik untuk klasifikasi gambar [7]. Salah satu metode Deep Learning terbaik yaitu Convolutional Neural Network (CNN) [8]. CNN merupakan salah satu bagian dari algoritma Artificial Neural Network (ANN) yang kemampuannya sangat baik dalam klasifikasi gambar. CNN memiliki beberapa model arsitektur, yaitu AlexNet, Suny, Stanford, Rao, LeNet, VGG16, MobileNet, dan masih banyak lagi [9]. Pada penelitian ini akan mengimplementasikan salah satu model CNN yaitu MobileNet.

Penelitian terkait bahasa isyarat menggunakan Deep Learning bukan permasalahan yang baru. Salah satu penelitian yang membahas Bahasa isyarat yaitu penelitian yang dilakukan oleh Thira, IndraJiwana dkk [10], yang mana menggunakan MobileNetV2, MobileNetV3 small, dan MobileNetV3 Large. Hasil penelitian tersebut yaitu accuracy sebesar 97,62% dan loss 0.1193 dengan dataset 840 gambar. Namun dataset tersebut terbilang sedikit sehingga membuat model bisa overfitting atau underfitting. Maka dari itu data augmentasi merupakan hal yang penting karena membuat data lebih heterogen.

Berdasarkan permasalahan yang telah dijelaskan, maka dalam penelitian ini diusulkan model arsitektur MobileNet. MobileNet adalah model yang berfokus pada periode waktu atau latensi [11]. MobileNet memiliki desain sebagai arsitektur yang sangat efisien dengan 2 set hyper-parameters. Hal tersebut bertujuan untuk membangun model dengan latensi rendah agar mudah mengimplementasikan kebutuhan MobileNet. Model ini dibuat untuk mengurangi komputasi layer awal dengan dasar depthwise separable convolutions [12]. MobileNet mampu untuk mengatasi kebutuhan komputasi dari sumber yang sangat banyak. Seperti Namanya model ini dapat digunakan pada smartphone. Penelitian ini bertujuan untuk mengimplementasikan model MobileNet untuk meminimalisir waktu dan meningkatkan akurasi.

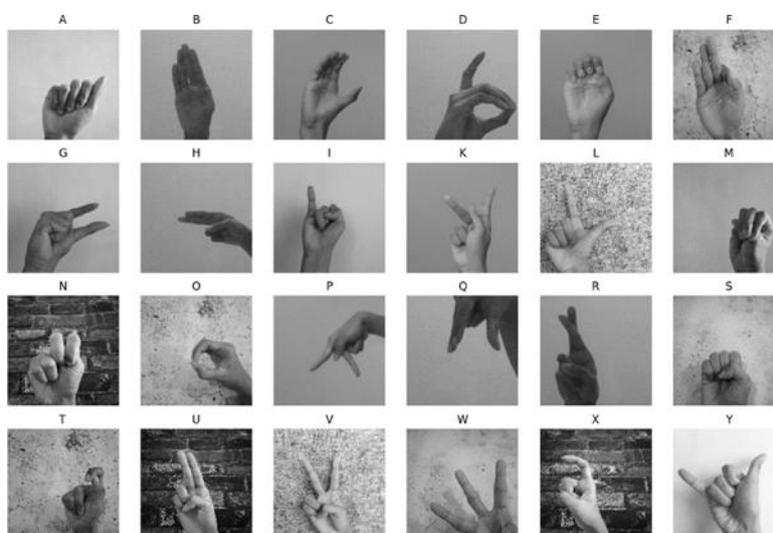
2. Metodologi Penelitian

Pada penelitian ini topik yang akan dibahas adalah mengenali gambar atau mendeteksi gambardengan menerapkan model arsitektur yang dipilih. Seperti yang ada pada Gambar 1 dijelaskan bahwa ada beberapa tahapan yang dilakukan agar penelitian ini dapat berhasil. Pertama mengidentifikasi masalah yang akan dibahas yaitu pengenalan pada suatu objek. Kedua mengumpulkan dataset yang digunakan untuk penelitian. Untuk meningkatkan keberhasilan penelitian maka perlu ditambah jumlah dataset dengan proses augmentasi. Ketiga melakukan pencarian arsitektur model yang akan digunakan untuk penelitian ini. Terakhir yaitu menerapkan model yang terpilih dan mengevaluasi agar mengetahui kemampuan model agar dapat diperbaiki pada penelitian selanjutnya.



Gambar 1. Rancangan Penelitian

Data primer yang akan digunakan pada penelitian ini adalah data gambar SIBI yang diambil dari Kaggle.com. Dalam dataset ini terdapat 5280 gambar yang terbagi 3 kategori yaitu data training, data validation, dan data testing yang terdiri dari 24 kelas, dengan pengecualian huruf J dan Z. Semua gambar pada dataset akan diubah menjadi format Joint Photographic Experts Group (JPEG) dengan ukuran gambar 128 x 128 pixels dan grayscale. Pada gambar 2 menunjukkan contoh gambar masing- masing kelas pada dataset.



Gambar 2. Contoh Gambar Dataset

3. Hasil Penelitian dan Pembahasan

3.1 Proses Augmentasi dan Pembagian dataset

Tahapan augmentasi merupakan tahapan setelah pengumpulan dataset yang memiliki 220 gambar setiap kelas dengan total 24 kelas. Augmentasi yang diterapkan pada setiap gambar yaitu rotasi, shear, dan flip sehingga setiap gambar akan membuat 3 gambar baru. Total dari hasil augmentasi ialah 21120 gambar yang setiap kelasnya memiliki 880 gambar. Pembagian dataset dilakukan setelah pengambilan gambar untuk data testing. Data testing akan mengambil 20

gambar acak pada setiap kelas. Tabel 1 menunjukkan rencana pembagian dataset yang akan dilakukan dengan rasio 8:2 untuk data training dan data validation.

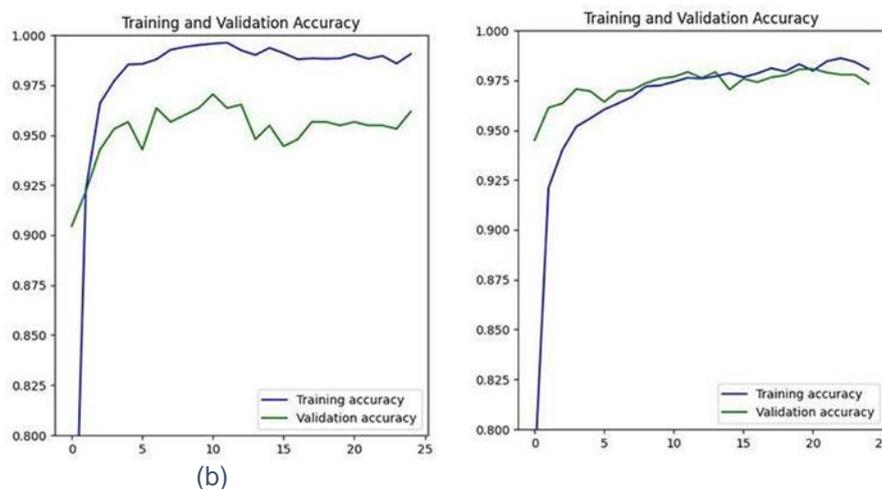
Tabel 1. Pembagian Dataset

Total Kelas	Data Training	Data Validation	Data Testing	Total Gambar
24	16896	3744	480	21120

3.2 Pelatihan model *MobileNetV2*

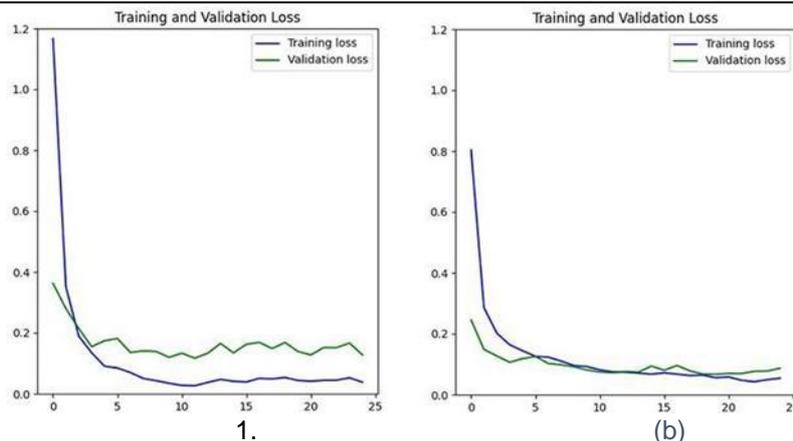
Model yang dihasilkan oleh *Convolutional Neural Network (CNN)* diperlukan 3 parameter untuk *compile* model, yaitu optimizer, loss, dan metrics. *Optimizer* pada penelitian ini akan menggunakan *Adam* dengan *learning rate* 0.001. *Optimizer Adam* digunakan untuk mempercepat pelatihan model dalam mencapai loss minimum. Sedangkan pada *loss* akan menggunakan *categorical_crossentropy*. Nilai semakin kecil menunjukkan performa model yang lebih baik. Parameter terakhir yaitu *metrics accuracy* yang akan menunjukkan hasil akurasi pada saat melatih model.

Tahapan selanjutnya adalah pelatihan model atau *fit* model yang bertujuan melatih data gambar ke dalam model. Parameter yang digunakan antara lain *batch_size* dan *epoch*. *Batch_size* akan menggunakan nilai sebesar 32 yang merupakan jumlah contoh pelatihan dalam 1 *forward/backward pass*. Sedangkan pada bagian *epoch* dilakukan dengan 25 *epoch* saja. Setelah parameter sudah ditentukan maka data *training* dan data *validation* akan dimasukkan pada bagian *fit* juga.



Gambar 3. Perbandingan Grafik Akurasi dengan Dataset Asli (a) dengan Dataset setelah Augmentasi (b)

Berdasarkan Gambar 3, hasil akurasi antara dataset asli (a) dengan dataset setelah augmentasi memiliki perbedaan secara signifikan. Grafik metrik akurasi model mendapatkan akurasi terbaik dari *epoch* ke-3 dan terus bergerak sampai akhir proses pelatihan. Hasil akurasi *training* pada dataset asli (a) sebesar 99,20% serta akurasi *validation* sebesar 94,27%. Sedangkan hasil akurasi *training* pada dataset setelah augmentasi (b) sebesar 98,25% serta akurasi *validation* sebesar 97,65%.



Gambar 4. Perbandingan Grafik Loss Antara Dataset Asli (a) dengan Dataset setelah Augmentasi (b)

Pada Gambar 4 menunjukkan hasil *loss* antara dataset asli (a) dengan dataset setelah augmentasi(b). Grafik metrik *loss* model mendapatkan *loss* terendah pada *epoch* ke-4 dan terus bergerak sampai pada akhir proses pelatihan. Hasil *loss training* pada dataset asli (a) sebesar 0,0360 serta *loss validation* sebesar 0,1639. Sedangkan hasil *loss training* pada dataset setelah augmentasi (b) sebesar 0.0505 serta *loss validation* sebesar 0,0784.

3.3 Evaluasi Model dan Prediksi hasil

Setelah semua proses pelatihan selesai maka perlu dilakukan evaluasi untuk mengetahui efektifitas dan efisiensi model yang sudah diterapkan. Pada tahapan ini model diukur performanya dengan melihat parameter *accuracy*, *recall*, dan *score f1*. Selain itu tahapan ini akan *confusion matrix* juga. Proses *score accuracy* dan *score loss* akan menggunakan *library Keras* yaitu *evaluate* pada data test. Hasil *evaluate* tersebut menghasilkan *accuracy* sebesar 98,75% serta *loss* sebesar 0,0592. Sedangkan untuk *score precision*, *recall*, dan *f1* akan menggunakan *classification report* dan mendapatkan nilai 0,99 atau 99% seperti pada Gambar 5.

```

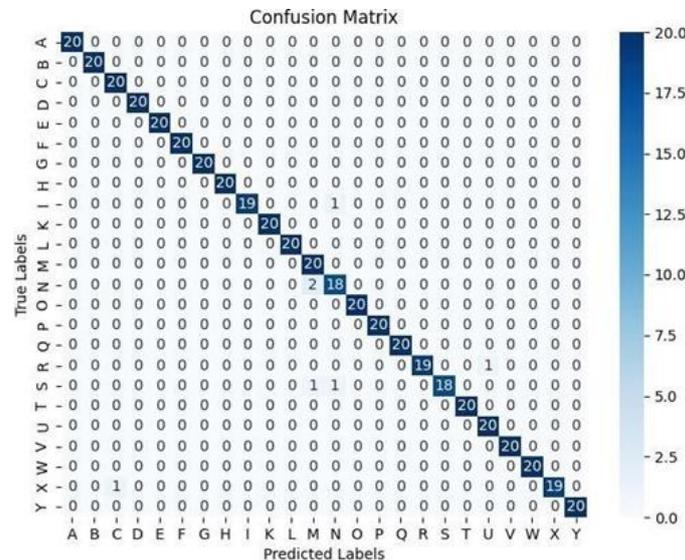
Classification Report:
      precision    recall  f1-score   support

 A           1.00        1.00        1.00         20
 B           1.00        1.00        1.00         20
 C           0.95        1.00        0.98         20
 D           1.00        1.00        1.00         20
 E           1.00        1.00        1.00         20
 F           1.00        1.00        1.00         20
 G           1.00        1.00        1.00         20
 H           1.00        1.00        1.00         20
 I           1.00        0.95        0.97         20
 K           1.00        1.00        1.00         20
 L           1.00        1.00        1.00         20
 M           0.87        1.00        0.93         20
 N           0.90        0.90        0.90         20
 O           1.00        1.00        1.00         20
 P           1.00        1.00        1.00         20
 Q           1.00        1.00        1.00         20
 R           1.00        0.95        0.97         20
 S           1.00        0.90        0.95         20
 T           1.00        1.00        1.00         20
 U           0.95        1.00        0.98         20
 V           1.00        1.00        1.00         20
 W           1.00        1.00        1.00         20
 X           1.00        0.95        0.97         20
 Y           1.00        1.00        1.00         20

 accuracy          0.99          0.99          0.99         480
 macro avg         0.99          0.99          0.99         480
 weighted avg      0.99          0.99          0.99         480
    
```

Gambar 5. Classification Report dari Model MobileNetV2

Sedangkan pada evaluasi *confusion matrix* akan memprediksi gambar pada data test, kemudian akan dilihat apakah pada setiap kelas dapat memprediksi dengan benar secara menyeluruh atau ada kesalahan dalam memprediksi gambar. Model mampu memprediksi sebanyak 473 gambar dengan benar serta 7 gambar yang salah dari 480 gambar data test yang dapat dilihat pada Gambar 6.



Gambar 6. Confusion Matrix dari Model MobileNetV2

4. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan, maka terdapat beberapa kesimpulan yang dapat diambil. Metode *Deep Learning* cukup bagus dengan arsitektur *MobileNetV2* yang terbilang ringan komputasinya mendapatkan hasil akurasi sebesar 98% dan hasil *score precision, recall*, serta *f1* yang nilainya sama yaitu 99% hanya waktu pelatihan 373 detik dalam 25 epoch. Dapat juga dilihat pada grafik plot akurasi dan loss pada masing-masing plot training atau plot validation mampu menghasilkan nilai yang baik. Lalu pada confusion matrix dapat disimpulkan dari 480 gambar yang diuji dapat menebak 473 gambar dengan benar serta 7 gambar gagal karena beberapa faktor, salah satunya kemiripan. Selain itu dengan dataset yang lebih heterogen, nilai akurasi, loss, presisi, recall dan *f1* dapat meningkat secara signifikan.

Referensi

- [1] Nurmansyah, A., Rhamadhani, N. R., Alfarissy, S., Hakim, N., Agustin, A., & Hamidah, S. (n.d.). Permasalahan Komunikasi Yang Kerap Terjadi Pada Penyandang Disabilitas Bahasa Dan Budaya, 2(2). <https://doi.org/10.55606/jpbb.v2i2.1515>
- [2] Farnell, B. M., & Davies Brenier, L. (2021). Sign Language. In Jackson Jr John L (Ed.), Oxford Bibliographies in Anthropology. Oxford University Press. <https://doi.org/10.1093/obo/9780199766567-0285>
- [3] Goud, V., Gupta, R., Suresh Babu, A., Das, D., Kulkarni, G., & Swathi, K. (2021). Oral health status and treatment needs among deaf, mute and visually impaired children of Gulbarga district – A population based cross sectional study. *Journal of Family Medicine and Primary Care*, 10(10), 3664. https://doi.org/10.4103/jfmpc.jfmpc_291_21
- [4] Rauf Yulian, S. (n.d.). Pengenalan Bahasa Isyarat Huruf Abjad Menggunakan Metode Learning Vector Quantization (LVQ). In Suhartono *Jurnal Masyarakat Informatika* (Vol. 8, Issue 1).
- [5] Kusters, A. (2021). International sign and American sign language as different types of global deaflingua francas. In *Sign Language Studies* (Vol. 21, Issue 4, pp. 391–426). Gallaudet University Press. <https://doi.org/10.1353/sls.2021.0005>
- [6] Sosialisasi Jurnal Hasil Pemikiran, J., Pengembangan Keilmuan Sosiologi Pendidikan Vol, dan, Aisyah Muhammad Amin, N., Pribadi, F., & Kunci, K. (n.d.). Urgensi Bahasa Isyarat dalam Pendidikan Formal sebagai Media Komunikasi dan Transmisi Informasi Penyandang Disabilitas Rungu dan Wicara.

- [7] Chen, A., Li, C., Rahaman, M. M., Yao, Y., Chen, H., Yang, H., Zhao, P., Hu, W., Liu, W., Zou, S., Xu, N., & Grzegorzec, M. (2023). A Comprehensive Comparative Study of Deep Learning Methods for Noisy Sperm Image Classification: from Convolutional Neural Network to Visual Transformer. *Intelligent Medicine*. <https://doi.org/10.1016/j.imed.2023.04.001>
- [8] Hemalatha, B., Karthik, B., Krishna Reddy, C. v., & Latha, A. (2022). Deep learning approach for segmentation and classification of blood cells using enhanced CNN. *Measurement: Sensors*, 24. <https://doi.org/10.1016/j.measen.2022.100582>
- [9] Rafly Alwanda, M., Putra, R., Ramadhan, K., & Alamsyah, D. (2020). Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle. In *Jurnal Algoritme* (Vol. 1, Issue 1).
- [10] Thira, I. J., Riana, D., Ilhami, A. N., Rizky, B., Dwinanda, S., & Choerunisya, H. (n.d.). Pengenalan Alfabet Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan Convolutional Neural Network. www.github.com.
- [11] Edel, G., & Kapustin, V. (2022). Exploring of the MobileNet V1 and MobileNet V2 models on NVIDIA Jetson Nano microcomputer. *Journal of Physics: Conference Series*, 2291(1). <https://doi.org/10.1088/1742-6596/2291/1/012008>.
- [12] Prasetyo, E., Purbaningtyas, R., Adityo, R. D., Suciati, N., & Fatichah, C. (2022). Combining MobileNetV1 and Depthwise Separable convolution bottleneck with Expansion for classifying the freshness of fish eyes. *Information Processing in Agriculture*, 9(4), 485–496. <https://doi.org/10.1016/j.inpa.2022.01.002>
- [13] Wahyu Farosa, A., Studi Magister Desain, P., Desain, J., & Seni Rupa dan Desain, F. (n.d.). Perancangan Mobile Apps Kamus Sebagai Media Dokumentasi Bahasa Isyarat Khas Bandung Dengan Peraga Animasi 3D. *Jurnal Seni Rupa*, 12.
- [14] Sharada, K., Alghamdi, W., Karthika, K., Alawadi, A. H., Nozima, G., & Vijayan, V. (2023). Deep Learning Techniques for Image Recognition and Object Detection. *E3S Web of Conferences*, 399. <https://doi.org/10.1051/e3sconf/202339904032>
- [15] Ali, A. H., Yaseen, M. G., Aljanabi, M., Abed, S. A., & GPT, C. (2023). Transfer Learning: A New Promising Techniques. *Mesopotamian Journal of Big Data*, 29–30. <https://doi.org/10.58496/mjbd/2023/004>
- [16] Sulistya, Y. I., Br Bangun, E. T., & Tyas, D. A. (2023). CNN Ensemble Learning Method for Transfer learning: A Review. *ILKOM Jurnal Ilmiah*, 15(1), 45–63. <https://doi.org/10.33096/ilkom.v15i1.1541.45-63>
- [17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," Apr. 2017. arXiv:1704.04861. [Online]. Available: <https://arxiv.org/abs/1704.04861>

