

DETEKSI HURUF VOKAL BAHASA ISYARAT INDONESIA (BISINDO) MENGGUNAKAN CNN ARSITEKTUR MOBILENETV2

Rangga Tansa Trisna^{1*}, Shofiya Syidada²

^{1,2}Program Studi Informatika, Fakultas Teknik, Universitas Wijaya Kusuma Surabaya
Email: rtt023@mhs.uwks.ac.id¹, shofiya@uwks.ac.id²

*Penulis Korespondensi

Abstrak

Penelitian ini bertujuan untuk mendeteksi huruf vokal pada Bahasa Isyarat Indonesia (BISINDO) dengan memanfaatkan metode Convolutional Neural Network (CNN) menggunakan arsitektur MobileNetV2 dan EfficientNet sebagai ekstraktor fitur. Penelitian difokuskan pada lima huruf vokal, yaitu A, I, U, E, dan O, karena jumlah kelas yang lebih sedikit dinilai dapat membantu model dalam mempelajari pola dengan lebih baik. Data yang digunakan berupa gambar tangan yang diperoleh dari berbagai sumber, kemudian diproses melalui augmentasi untuk menambah variasi data. Selanjutnya, data dibagi secara stratified menjadi data pelatihan, validasi, dan pengujian dalam dua skenario pembagian. Model dibangun menggunakan lapisan Dense dengan fungsi aktivasi ReLU dan Softmax. Penelitian ini juga melakukan perbandingan performa antara arsitektur MobileNetV2 dan EfficientNet berdasarkan nilai accuracy, loss, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa MobileNetV2 memiliki performa yang lebih baik dibandingkan EfficientNet dalam mengenali huruf vokal BISINDO dengan tingkat akurasi yang lebih tinggi. Penelitian ini diharapkan dapat menjadi dasar pengembangan teknologi pendeteksi bahasa isyarat yang mendukung komunikasi antara penyandang tunarungu dan masyarakat umum.

Kata kunci: Sistem Deteksi, Bahasa Isyarat, Machine Learning, Deep Learning, CNN

Abstract

This study aims to detect vowels in Indonesian Sign Language (BISINDO) using the Convolutional Neural Network (CNN) method using the MobileNetV2 architecture as a feature extractor. The study focused on five vowels, namely A, I, U, E, and O, because a smaller number of classes is considered to help the model learn patterns better. The data used were hand images obtained from various sources, then processed through augmentation to increase data variation. Next, the data was stratified into training, validation, and testing data in two division scenarios. The model was built using a Dense layer with ReLU and Softmax activation functions. Model performance was evaluated using a classification report to see the level of accuracy and classification performance. The results of the study are expected to show that the model is able to recognize BISINDO vowels well and has the potential to be developed as a technology that supports communication between deaf people and the general public.

Keywords: Detection Systems, Sign Language, Machine Learning, Deep Learning, CNN

I. PENDAHULUAN

Di zaman sekarang, teknologi tidak dapat dihindari; hampir semua kegiatan manusia bergantung pada teknologi. Dengan bantuan teknologi canggih ini, manusia telah melakukan banyak kemajuan. Mungkin lebih mudah untuk melakukan pekerjaan di era modern. seperti belanja, belajar, dll. Di era globalisasi saat ini, teknologi telah menjadi nilai dan tanda kemajuan. Sebuah negara dapat dianggap maju jika memiliki penguasaan teknologi canggih atau tinggi. Orang ini telah menikmati banyak manfaat dari teknologi yang telah membuat hidup lebih mudah bagi penggunanya [1]. Oleh karena itu, pemerataan pemahaman terhadap teknologi ini merupakan dasar berpijak bagi totalitas masyarakat untuk dapat memasuki era globalisasi [2].

Tidak dapat dihindari bahwa komunikasi adalah komponen penting dari kehidupan manusia. Baik secara sadar maupun tidak, semua orang terlibat dalam proses komunikasi untuk berbagi ide, perasaan, dan informasi. Komunikasi efektif menjadi semakin penting sebagai alat strategis untuk membangun pemahaman, kepercayaan, dan kerjasama yang berkelanjutan dalam berbagai aspek kehidupan di dunia yang kompleks dan dinamis saat ini. Ini karena tantangan komunikasi bukan hanya masalah teknis, tetapi juga terkait dengan meningkatnya konflik sosial, kerusakan hubungan, dan rendahnya produktifitas di lingkungan keluarga, pendidikan, dan organisasi [3]. Tujuan komunikasi dalam pendidikan adalah

menyampaikan informasi tentang materi pelajaran yang akan diajarkan, agar materi pelajaran yang disampaikan dapat dimengerti dan dipahami sebagai komunikasi yang kita laksanakan dapat tercapai [4].

Bahasa adalah cara orang berinteraksi satu sama lain. Meskipun beberapa orang dapat berkomunikasi secara verbal dalam kehidupan sehari-hari, beberapa orang tidak dapat berkomunikasi secara verbal. Dalam berkomunikasi, orang tunarungu dan tunawicara menggunakan bahasa isyarat. Dibutuhkan perantara alternatif yang dapat berfungsi sebagai penerjemah antara penderita tunarungu dan tunawicara dan orang normal karena bahasa isyarat yang digunakan orang tunarungu dan tunawicara umum tetapi tidak digunakan orang normal [5]. Menurut World Federation of the Deaf, terdapat lebih dari 300 bahasa isyarat di seluruh dunia dan 70 juta tunarungu menggunakannya [6].

Bahasa isyarat mampu menunjukkan identitas seorang tunarungu [7]. Hasil penelitian menunjukkan bahwa SIBI dan BISINDO berbeda dalam hal komunikasi tunarungu. Pada BISINDO, hanya 3% yang sangat tidak setuju, 6% yang cukup setuju, 43% yang setuju, dan 48% yang sangat setuju. Sebaliknya, pada SIBI, 31% yang sangat tidak setuju, 37% yang tidak setuju, 24% yang cukup setuju, 6% yang setuju, dan 2% yang sangat setuju. Menurut data, BISINDO lebih diterima dibandingkan SIBI. Sekitar 91% responden menggunakan BISINDO dalam komunikasi sehari-hari, sedangkan hanya 9% menggunakan SIBI. Ini menunjukkan bahwa BISINDO lebih efektif, lebih mudah dipahami, dan lebih sesuai dengan kebutuhan komunikasi penyandang tunarungu di Indonesia saat ini [8]. Mengingat gerakan dan ekspresi yang berbeda antar bahasa isyarat, teknologi yang dikembangkan untuk satu bahasa isyarat seringkali tidak dapat diadaptasi untuk bahasa isyarat lain [9].

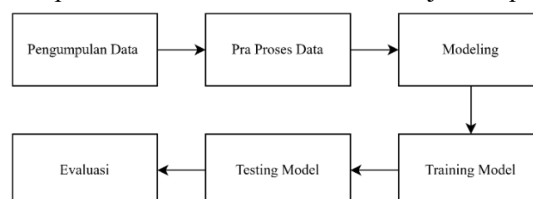
Deep Learning merupakan bagian dari bidang pengolahan citra digital atau image processing. Dengan adanya sistem pengolahan citra, tujuannya adalah untuk membantu manusia dalam mengenali atau mengklasifikasi objek dengan efisien, yaitu dengan cepat, tepat, dan mampu melakukan proses dengan jumlah data yang besar secara bersamaan [10]. Deep Learning merupakan bagian dari Machine Learning yang dapat mempelajari metode komputasinya sendiri. Dalam perkembangan Deep Learning saat ini, Convolutional Neural Network (CNN) telah menunjukkan hasil yang paling signifikan dalam tugas pengenalan citra [11].

CNN merupakan metode yang efektif untuk mengklasifikasikan, mengidentifikasi, dan mengenali pola di dalam citra. [12]. Beberapa peneliti telah mengembangkan teknologi self-checkout system dalam mengatasi permasalahan tersebut. Salah satu arsitektur yang sering digunakan dalam deteksi objek citra ialah Convolutional Neural Network (CNN) [13]. Arsitektur model CNN MobileNetV2 juga dikembangkan, hal ini dikarenakan akurasi yang lebih tinggi serta jumlah komputasi yang lebih rendah [14].

Arsitektur CNN MobileNetV2 dirancang untuk digunakan pada perangkat mobile. Arsitektur ini memiliki kelebihan pada efisiensi dan kecepatan dibandingkan dengan MobileNetV2 karena menggunakan convolutions yang terpisah secara kedalaman, yang mengurangi jumlah parameter dan operasi komputasi dibandingkan dengan convolutions standar. Ini memungkinkan model berjalan lebih cepat dengan menggunakan lebih sedikit memori. Meskipun demikian, MobileNetV2 memiliki beberapa kekurangan. MobileNetV2 masih kurang akurat jika dibandingkan dengan model yang lebih kompleks dan berat seperti ResNet atau Inception, meskipun lebih efisien daripada model CNN konvensional. Tugas-tugas yang membutuhkan akurasi tinggi dapat mengalami penurunan kinerja jika jumlah parameter dan komputasi dikurangi. Selain itu, meskipun lebih ringan, penerapan MobileNetV2 pada perangkat dengan keterbatasan yang sangat ekstrem masih dapat menjadi masalah [15].

II. METODE PENELITIAN

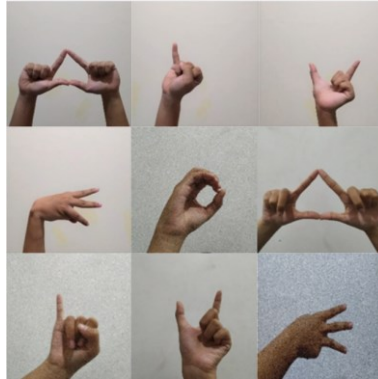
Penelitian ini menggunakan pendekatan *computer vision* berbasis *deep learning* dengan metode *Convolutional Neural Network* (CNN) untuk melakukan klasifikasi bahasa isyarat BISINDO. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, perancangan model, pelatihan (*training*), serta pengujian (*testing*) model. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

2.1 Pengumpulan Data

Berdasarkan Gambar 1, proses penelitian dimulai dari pengumpulan data hingga tahap evaluasi untuk mengetahui performa model dalam melakukan klasifikasi. Dataset yang digunakan berupa citra tangan bahasa isyarat BISINDO yang diperoleh dari berbagai sumber, seperti dataset publik dan pengambilan gambar secara langsung oleh peneliti. Data yang digunakan terdiri dari beberapa kelas huruf vokal (A, I, U, E, O). Dataset kemudian dibagi menjadi data training, validasi, dan testing menggunakan dua skenario pembagian, yaitu 70:15:15 dan 80:10:10 untuk menganalisis pengaruh jumlah data terhadap performa model.

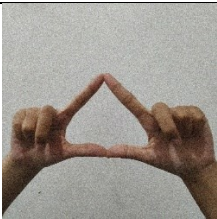
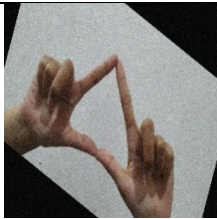


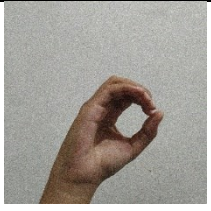
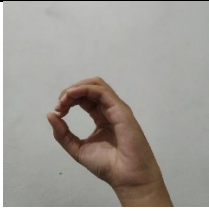
Gambar 2. Contoh Dataset BISINDO

2.2 Pra Proses Data

Pada tahap pra-pemrosesan, dilakukan augmentasi data untuk meningkatkan variasi dataset, seperti rotasi, pergeseran, *shearing*, *zooming*, dan pembalikan horizontal. Selain itu, citra dinormalisasi dengan rentang nilai piksel 0–1 serta diubah ukurannya menjadi 128×128 piksel agar sesuai dengan input model.

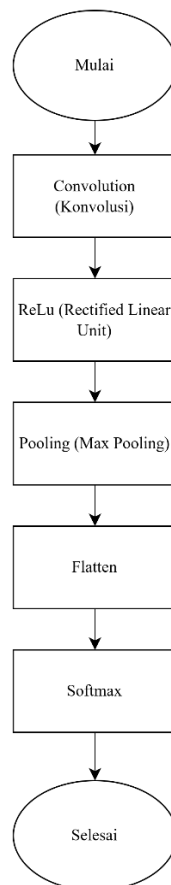
Tabel 1. Image Processing

Dataset Asli	Image Processing	Hasil Image Processing
	Rotasi	
	Pergeseran	
	Shearing (Kemiringan Bentuk)	

	Zooming (Perbesar/Perkecil)	
	Pembalikan Horizontal	

2.3 Modelling

Model yang digunakan dalam penelitian ini adalah CNN dengan pendekatan *transfer learning*. Dua arsitektur yang digunakan yaitu MobileNetV2 dan EfficientNet sebagai *feature extractor*, yang kemudian dikombinasikan dengan lapisan *Dense*, fungsi aktivasi ReLU, serta *Dropout* untuk mengurangi *overfitting*. Pada lapisan output digunakan fungsi aktivasi Softmax untuk menghasilkan probabilitas klasifikasi.



Gambar 3. Flowchart MobileNetV2

2.4 Training Model

Proses pelatihan dilakukan menggunakan fungsi loss *categorical_crossentropy* dan optimizer Adam. Selama proses pelatihan, performa model dievaluasi menggunakan metrik *accuracy* dan *loss* pada data training dan validasi.

2.5 Testing Model

Pengujian dilakukan untuk mengevaluasi performa dan kemampuan generalisasi model CNN dalam mengklasifikasikan bahasa isyarat BISINDO. Dataset dibagi menggunakan dua skenario, yaitu rasio 70:15:15 dan 80:10:10 untuk data training, validasi, dan testing.

2.6 Evaluasi

Evaluasi dilakukan menggunakan data testing yang telah dipisahkan sebelumnya. Performa model diukur menggunakan classification report yang meliputi precision, recall, dan F1-score pada setiap kelas. Selain itu, metrik *accuracy* dan *loss* juga digunakan untuk membandingkan performa model pada kedua skenario pembagian dataset. Pengujian ini bertujuan untuk memastikan bahwa model mampu melakukan klasifikasi secara akurat dan memiliki kemampuan generalisasi yang baik terhadap data baru.

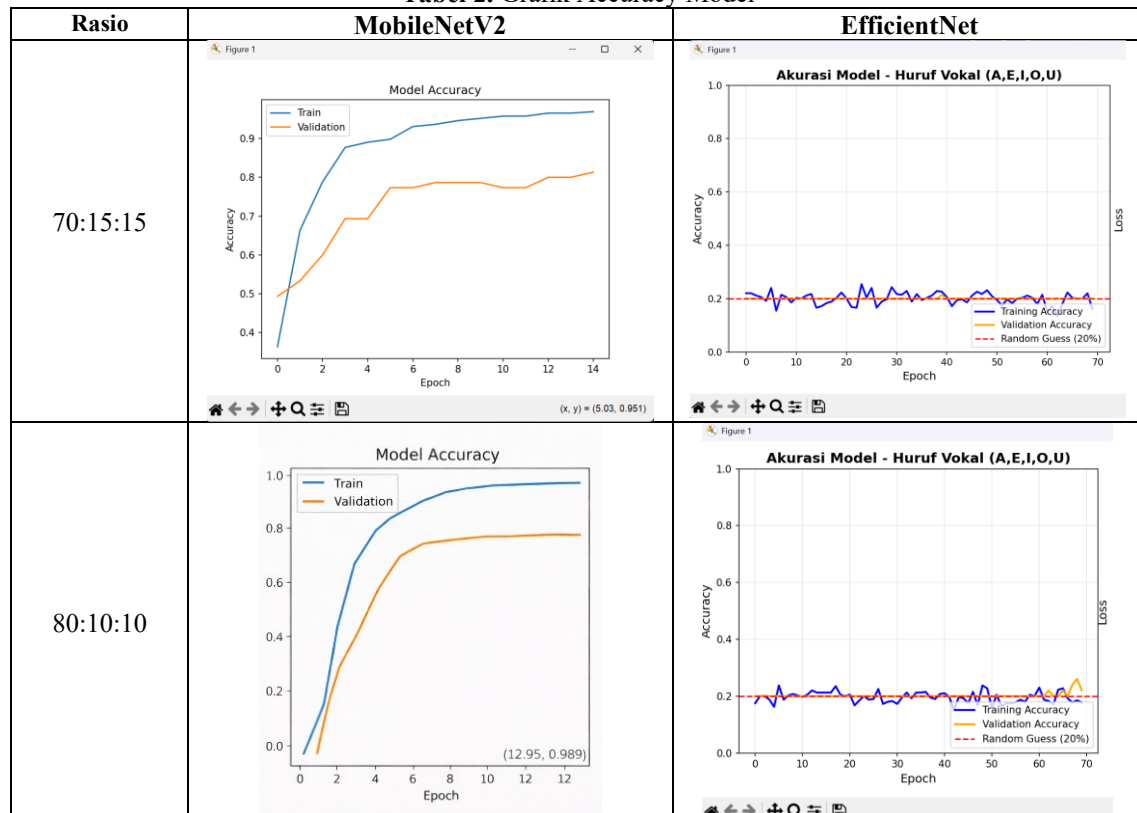
III. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan model klasifikasi bahasa isyarat BISINDO menggunakan metode *Convolutional Neural Network* (CNN) dengan perbandingan arsitektur MobileNetV2 dan EfficientNet. Pengujian dilakukan menggunakan dua skenario pembagian dataset, yaitu rasio 70:15:15 dan 80:10:10 untuk menganalisis performa model.

3.1 Hasil Training Model

Pada tahap ini dilakukan proses pelatihan model menggunakan arsitektur MobileNetV2 dan EfficientNet. Evaluasi performa model dilakukan menggunakan metrik *accuracy* dan *loss* pada data training dan validasi.

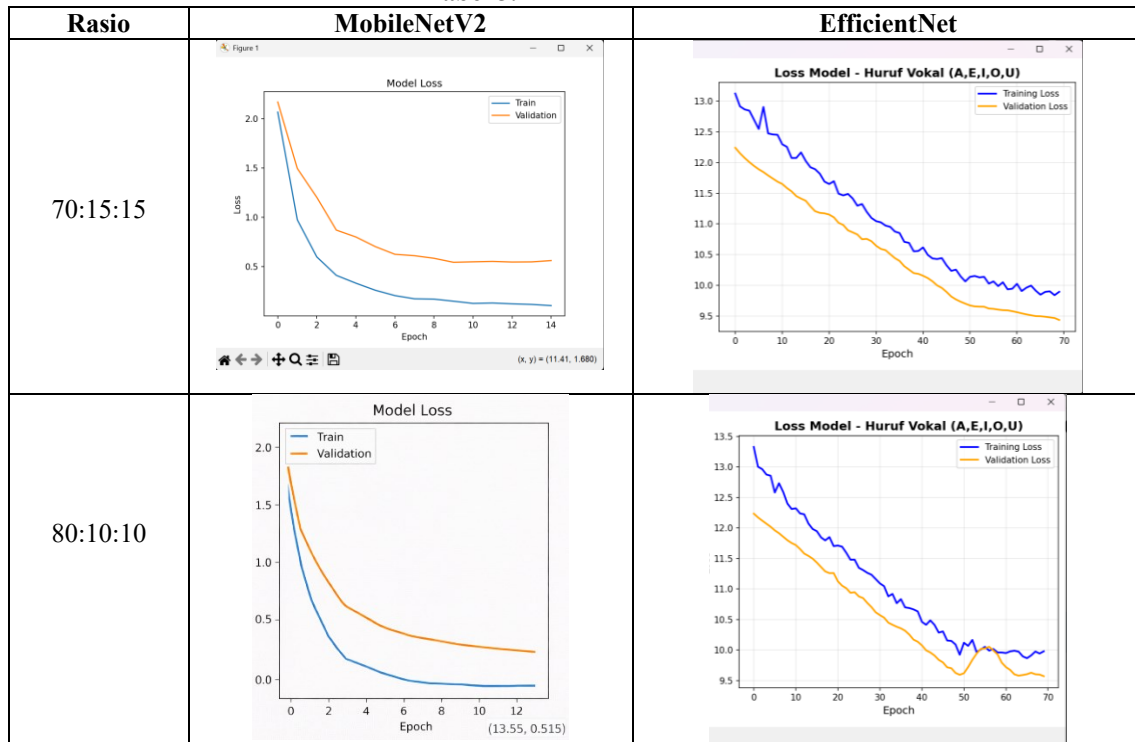
Tabel 2. Grafik Accuracy Model



Berdasarkan **Tabel 2**, grafik accuracy, model MobileNetV2 menunjukkan peningkatan akurasi yang signifikan hingga mencapai lebih dari 90% pada data training dan sekitar 75–80% pada data validasi. Hal ini menunjukkan bahwa model mampu mempelajari pola data dengan baik, meskipun terdapat indikasi *overfitting* ringan.

Sebaliknya, model EfficientNet menunjukkan performa yang rendah dengan akurasi yang cenderung stagnan di sekitar 20%, yang mengindikasikan bahwa model mengalami *underfitting* dan tidak mampu mempelajari pola dari dataset secara optimal.

Tabel 3. Grafik Loss Model



Berdasarkan **Tabel 3**, grafik loss, model MobileNetV2 menunjukkan penurunan loss yang signifikan pada data training dan validasi. Sementara itu, EfficientNet juga mengalami penurunan loss, namun tidak diikuti dengan peningkatan akurasi, yang memperkuat indikasi *underfitting*.

3.2 Hasil Pengujian Model

Evaluasi model dilakukan menggunakan data testing untuk mengetahui kemampuan generalisasi model.

Tabel 4. Classification Report

Rasio	MobileNetV2	EfficientNet
70:15:15	==== CLASSIFICATION REPORT =====	precision recall f1-score support
	75/75 4s 27ms/step	
	precision recall f1-score support	
	A 1.00 0.87 0.93 15	A 0.0000 0.0000 0.0000 15
	E 0.68 0.87 0.76 15	E 0.2000 1.0000 0.3333 15
	I 1.00 0.27 0.42 15	I 0.0000 0.0000 0.0000 15
	O 0.62 1.00 0.77 15	O 0.0000 0.0000 0.0000 15
	U 0.73 0.73 0.73 15	U 0.0000 0.0000 0.0000 15
	accuracy 0.75	accuracy 0.2000
	macro avg 0.81 0.75 0.72 75	macro avg 0.0400 0.2000 0.0667 75
	weighted avg 0.81 0.75 0.72 75	weighted avg 0.0400 0.2000 0.0667 75

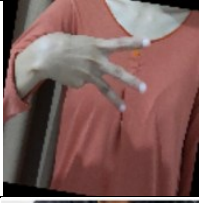
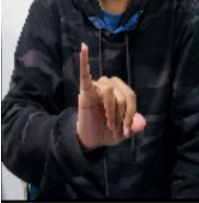
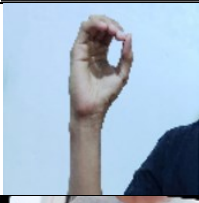
80:10:10	==== CLASSIFICATION REPORT =====					precision	recall	f1-score	support	
	75/75	3s 25ms/step								
		precision	recall	f1-score	support					
	A	1.00	1.00	1.00	15	A	0.2000	0.4000	0.2667	10
	E	0.72	0.87	0.79	15	E	0.0000	0.0000	0.0000	10
	I	1.00	0.27	0.42	15	I	0.0000	0.0000	0.0000	10
	O	0.68	1.00	0.81	15	O	0.0000	0.0000	0.0000	10
	U	0.69	0.73	0.71	15	U	0.2000	0.6000	0.3000	10
						accuracy			0.2000	50
						macro avg	0.0800	0.2000	0.1133	50
						weighted avg	0.0800	0.2000	0.1133	50
		accuracy			0.77	75				
		macro avg	0.82	0.77	0.75	75				
	weighted avg	0.82	0.77	0.75	75					

 Akurasi Testing Keseluruhan: 0.2000 (20.00%)

Berdasarkan **Tabel 4.** hasil *classification report*, model MobileNetV2 memperoleh akurasi sekitar 75–77%. Model mampu mengenali sebagian besar huruf dengan baik, namun masih terdapat kesalahan pada huruf yang memiliki kemiripan bentuk, seperti huruf I dan O. Sebaliknya, model EfficientNet menunjukkan performa yang sangat rendah dengan akurasi sekitar 20%, di mana model hanya mampu mengenali satu kelas secara dominan.

3.3 Hasil Perbandingan Testing

Tabel 5. Hasil Perbandingan Testing

Data Testing	Huruf Asli	MobileNet	EfficientNet
	A	A	U
	E	E	E
	I	U	U
	O	O	U
	U	E	U

Berdasarkan **Tabel 5.** menunjukkan hasil perbandingan testing antara model MobileNet dan EfficientNet dalam mengenali huruf vokal BISINDO, yaitu A, E, I, O, dan U. Berdasarkan hasil pengujian tersebut, dapat dilihat bahwa model MobileNet mampu mengenali beberapa huruf dengan cukup baik, seperti huruf A, E, dan O yang berhasil diprediksi sesuai dengan huruf aslinya. Namun, masih terdapat kesalahan prediksi pada huruf I yang diprediksi menjadi U serta huruf U yang diprediksi menjadi E. Hal

ini menunjukkan bahwa beberapa bentuk gestur memiliki kemiripan sehingga model masih mengalami kesulitan dalam membedakan pola tertentu.

Sementara itu, model EfficientNet menunjukkan hasil yang berbeda. Model ini berhasil mengenali huruf E dan U dengan benar, tetapi mengalami kesalahan pada beberapa data lainnya. Huruf A diprediksi menjadi U, huruf I juga diprediksi menjadi U, dan huruf O diprediksi menjadi U. Dari hasil tersebut terlihat bahwa EfficientNet cenderung lebih sering mengklasifikasikan beberapa gestur ke dalam kelas U. Kondisi ini dapat terjadi karena bentuk tangan pada beberapa huruf memiliki karakteristik yang hampir mirip, sehingga model belum mampu menangkap perbedaan fitur secara optimal.

Secara keseluruhan, berdasarkan sampel data testing pada tabel tersebut, MobileNet menunjukkan performa yang lebih baik dibandingkan EfficientNet karena menghasilkan prediksi yang lebih banyak sesuai dengan huruf asli. Hasil ini menunjukkan bahwa arsitektur MobileNet lebih efektif dalam mengenali pola gestur huruf vokal BISINDO pada penelitian yang dilakukan.

3.4 Analisis Hasil

Hasil penelitian menunjukkan bahwa MobileNetV2 lebih efektif digunakan untuk klasifikasi huruf vokal BISINDO dibandingkan EfficientNet. Berdasarkan hasil pengujian pada Tabel 4, MobileNetV2 memperoleh accuracy sekitar 75–77%, sedangkan EfficientNet hanya mencapai sekitar 20%. Selain itu, pada Tabel 5 MobileNetV2 berhasil mengenali huruf A, E, dan O dengan benar, sementara EfficientNet lebih sering memprediksi data ke kelas U.

Performa MobileNetV2 yang lebih baik dipengaruhi oleh arsitekturnya yang ringan dan lebih sesuai untuk dataset berukuran terbatas. Selain itu, skenario pembagian dataset 80:10:10 juga menghasilkan performa yang lebih stabil dibandingkan 70:15:15 karena jumlah data training yang lebih banyak membantu model mempelajari pola gestur dengan lebih optimal.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan metode Convolutional Neural Network (CNN) untuk mendeteksi huruf vokal Bahasa Isyarat Indonesia (BISINDO), yaitu A, I, U, E, dan O. Berdasarkan hasil pengujian, arsitektur MobileNetV2 menunjukkan performa yang lebih baik dibandingkan EfficientNet dalam proses klasifikasi citra gesture tangan BISINDO. Selain itu, jumlah data training berpengaruh terhadap performa model, di mana pembagian dataset dengan data training yang lebih banyak menghasilkan hasil yang lebih optimal. Dengan demikian, MobileNetV2 dapat menjadi salah satu arsitektur yang efektif untuk pengembangan sistem deteksi bahasa isyarat berbasis deep learning.

REFERENSI

- [1] F. T. and S. M. , "Teknologi Sebagai Kegiatan Manusia Dalam Era Modern Kehidupan Masyarakat," *Jurnal Teknologi dan Sistem Informasi* , pp. 66-67, 2024.
- [2] A. Ahmad, "PERKEMBANGAN TEKNOLOGI KOMUNIKASI DAN INFORMASI: AKAR REVOLUSI DAN BERBAGAI STANDARNYA," *Perkembangan Teknologi Komunikasi Dan Informasi* , vol. 13, pp. 137-149, 2012.
- [3] I. G. A. S. R. Y. W. and S. K. J. , "Peran Komunikasi Efektif dalam Kehidupan Sehari-hari," *Jurnal Ilmu Komunikasi*, vol. 4, pp. 1508-1517, 2025.
- [4] E. N. Inah, "PERANAN KOMUNIKASI DALAM PENDIDIKAN," *Jurnal Al-Ta'dib* , vol. 6, pp. 176 - 188, 2013.
- [5] D. A. S. V. D. T. and H. , "Bahasa Isyarat Indonesia," *Jurnal Medical Profession*, vol. 4, no. Vol. 4 No. 1 (2022): Februari 2022, pp. 8-14, 2022.
- [6] J. Murray, "70 million deaf people. 200+ sign languages.," 2018.
- [7] N. M. and H. , "IMPLEMENTASI BISINDO DALAM BERKOMUNIKASI PADA SESAMA ANAK TUNARUNGU," *Jurnal UNM*, pp. 1-10, 2022.
- [8] R. A. Mursita, "RESPON TUNARUNGU TERHADAP PENGGUNAAN SISTEM BAHASA ISYARAT INDONESIA (SIBI) DAN BAHASA ISYARAT INDONESIA (BISINDO) DALAM KOMUNIKASI," *Jurnal Of Disability Studies*, vol. 2, no. Vol. 2 No. 2 (2015), pp. 221-232, 2015.
- [9] O. D. N. D. E. and M. H. T. , "SISTEM ISYARAT BAHASAINDONESIA(SIBI) METODECONVOLUTIONALNEURAL NETWORKSEQUENTIAL SECARA REAL TIME," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 9, pp. 819-828, 2022.

-
- [10] F. F. M. and N. R. , "Klasifikasi Citra Buah Menggunakan Convolutional Neural Network," *Journal of Informatics and Computer Science*, vol. Volume 01 Nomor 02, pp. 104-108, 2019.
 - [11] P. A. N. I. F. and R. A. , "IMPLEMENTASI DEEP LEARNING MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA EKSPRESI MANUSIA," *JURNAL ALGOR*, vol. VOL.2NO.1, no. Vol. 2 No. 1 (2020): Data And System, pp. 12-21, 2020.
 - [12] I. W. S. E. A. Y. W. and R. S. , "Klasifikasi Citra Menggunakan Convolutional Neural Network (Cnn) pada Caltech 101," *JURNAL TEKNIK ITS*, Vols. Vol. 5, No. 1, no. 2337-3539, pp. 65-69, 2016.
 - [13] A. R. P. . M. P. and S. R. , "Application of Deep Learning for Object Detection," *Procedia Computer Science*, vol. 132, pp. 1706-1717, 2018.
 - [14] M. S. A. H. M. Z. A. Z. and L.-C. C. , "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Arxiv Computer Vision and Pattern Recognition*, pp. 1-14, 2018.
 - [15] A. A. and A. V. H. , "Review of deep learning: Convolutional Neural NetworkAlgorithm.," *Babylonian Journal of Machine Learning*, no. 3006–5429, pp. 19-25, 2023.