

Pengembangan Deteksi Realtime untuk Bahasa Isyarat Indonesia dengan Menggunakan Metode *Deep Learning Long Short Term Memory* dan *Convolutional Neural Network*

Eka Altiarika^{1)}, Winda Purnama Sari²⁾

¹⁾ Fakultas Teknik dan Sains, Universitas Muhammadiyah Bangka Belitung

²⁾ Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Bangka Belitung

*Correspondence Author: eka.altiarika@unmuhbabel.ac.id, Pangkalanbaru, Indonesia

DOI: <https://doi.org/10.37012/jtik.v9i1.1272>

Abstrak

Kehadiran *Computer Vision* mampu mempengaruhi bidang kajian *Sign Language Recognition System* (SLRS). Adapun penelitian dibidang SLRS terhadap dua standar bahasa isyarat di Indonesia yaitu standar SIBI (Sistem Bahasa Isyarat Indonesia) dan BISINDO (Bahasa Isyarat Indonesia). Tantangan dalam penelitian ini adalah kendala dalam memproses gambar dinamis dan gambar statis ketika setelah melalui preprocessing rekognisi. Perlakuan yang berbeda saat recognisi awal pada gambar bergerak dengan gambar statis mempengaruhi waktu memunculkan hasil dengan cepat sehingga dibutuhkan model dengan *training* yang baik dan cepat. Tujuan penelitian ini adalah untuk mengetahui faktor akurasi yang mempengaruhi tingkat akurasi penerapan objek deteksi dan klasifikasi gambar maupun video secara *realtime* untuk BISINDO (Bahasa Isyarat Indonesia) dengan menggunakan metode *Deep Learning Long Short Term Memory* (LSTM) dan *Convolution Neural Network* (CNN). Pentingnya penelitian ini karena hasilnya dapat dijadikan dasar untuk mempercepat pengembangan lebih lanjut aplikasi *sign language recognition* khusus untuk BISINDO yang bisa dimanfaatkan oleh penyandang disabilitas maupun masyarakat agar komunikasi dua arah lebih mudah dilakukan dimasa depan secara *real-time*.

Kata kunci: Bahasa Isyarat Indonesia, *Deep Learning*, *Convolution Neural Network*, *Long Short Term Memory*

Abstract

The presence of *Computer Vision* is able to influence the field of study of *Sign Language Recognition System* (SLRS). As for research in the field of SLRS on two sign language standards in Indonesia, namely the SIBI standard (Indonesian Sign Language System) and BISINDO (Indonesian Sign Language). The challenge in this study is the constraints in processing dynamic images and static images when after going through recognition preprocessing. Different treatment during initial recognition of moving images and static images affects the time to produce results quickly, so a model with good and fast training is needed. The purpose of this study was to determine the accuracy factor that affects the accuracy of the application of object detection and classification of images and videos in real time for BISINDO (Indonesian Sign Language) using the *Deep Learning Long Short Term Memory* (LSTM) and *Convolution Neural Network* (CNN) methods. This research is important because the results can be used as a basis for accelerating further development of a sign language recognition application specifically for BISINDO that can be utilized by persons with disabilities and the public so that two-way communication can be carried out more easily in the future in real-time.

Keywords: Indonesian Sign Language, *Deep Learning*, *Convolution Neural Network*, *Long Short Term Memory*

PENDAHULUAN

Perkembangan kajian terhadap bidang *Computer Vision* berkembang pesat tidak hanya untuk penggunaannya di bidang robotik, interaksi manusia dengan komputer, autentikasi iris mata dan sidik jari, deteksi wajah, dan lainnya tetapi mampu membantu

mengembangkan aplikasi dengan kemampuan teknologi deteksi untuk bahasa isyarat yang dikenal dengan *teknologi Sign Language Recognition (SLR)*. Pengembangan SLR Bahasa yang diperuntukkan bagi komunitas tunarungu atau tuli khususnya di Indonesia yang mencapai 146.987 pada tahun 2018, untuk memudahkan berkomunikasi satu sama lain melalui berbagai bahasa isyarat dan gerak tubuh (Irwanto et al., 2010; Subburaj & Murugavalli, 2022).

Indonesia mengenal Bahasa isyarat yaitu SIBI (Sistem Bahasa Isyarat Indonesia) dan BISINDO (Bahasa Isyarat Indonesia) yang menggunakan bentuk terstruktur dari gerakan tangan komunikatif untuk mengekspresikan ide (Sharma & Singh, 2021; Suharjito et al., 2019) bersifat statis dan dinamis. Bidang rekognisi bahasa isyarat tangan dilakukan untuk menginterpretasikan cara memproses gambar statis dan gambar dinamis namun masih mengalami kendala karena kesulitan dalam memproses gambar dinamis/video (Rakun et al., 2013; Supria et al., 2016). Sistem pengenalan atau penerjemahan abjad BISINDO ke dalam sebuah teks yang diharapkan dapat membantu penyandang tuna rungu untuk berkomunikasi dua arah difokuskan untuk menghadapi *dataset* yang kecil (Aljabar & Suharjito, 2020; Damatraseta et al., 2021).

Deep learning menggunakan segmentasi data gambar (Hesamian et al., 2019) dapat terlihat saat sebuah video disajikan dan dilakukan recognisi awal deteksi karena video merupakan rangkaian gambar dan lebih susah untuk diklasifikasi karena mengandung *temporal* dan *spatial features*. Beberapa peneliti telah mencoba untuk membuat penelitian dengan video tetapi dengan bantuan alat khusus seperti sensor sarung tangan dan sensor *Kinect*. Metode *deep learning* yang telah digunakan dalam Penelitian yang berkaitan adalah menggunakan *Convolutional Neural Network (CNN)*. Keuntungan dari CNN adalah mudah untuk melakukan proses training dan memiliki parameter yang lebih sedikit dibandingkan arsitektur yang lainnya. Adapun berbagai tantangan dalam penelitian *deep learning* dengan pengenalan bahasa isyarat sebagai objek seperti pelatihan data, posisi objek, pose, pencahayaan, dan latar belakang objek (Sharma & Singh, 2021; Suharjito et al., 2021).

Dasar dari model LSTM-CNN adalah baseline LSTM untuk data penyematan untuk setiap token di label sebagai input. Intuisinya adalah bahwa keluaran token akan menyertakan informasi dari token sebelumnya dan saat ini. Artinya, lapisan LSTM menyediakan kode baru untuk input yang tidak diautentikasi. Output dari layer LSTM kemudian dikirim ke fitur lokalisasi yang diperkirakan akan mengalami gangguan. Keluaran lapisan konvolusi pada akhirnya akan mengecil menjadi ukuran yang lebih kecil dan diberi label baik secara positif maupun negatif pada saat itu. (Chevtchenko et al., 2018). Fondasi

dari model ini adalah *convolutional initial* lapisan yang akan menerima *embedding data* untuk setiap token pada input label. Output dari lapis CNN akan dikurangi menjadi ukuran yang lebih mudah diatur dan kemudian ditransfer ke lapis *Long Short Term Memory* (LSTM), yang diharapkan dapat meningkatkan fungsionalitas lokal. Akhirnya, itu diklasifikasikan sebagai label positif atau negatif. (Suharjito et al., 2019).

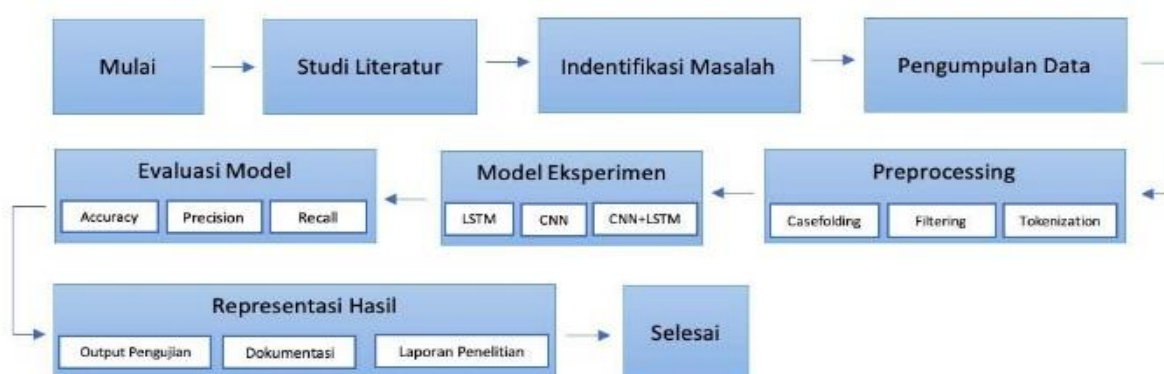
Berbagai inovasi tentang deteksi tangan dan pengenalan gerakan saat ini meningkat melalui berbagai aplikasi potensialnya, seperti interaksi manusia-komputer, pengenalan bahasa isyarat, analisis tindakan tangan, pemantauan perilaku tangan pengemudi, dan realitas virtual. Beberapa pendekatan telah diusulkan dengan tujuan mengembangkan algoritma yang kuat yang berfungsi dalam lingkungan yang kompleks dan berantakan. Meskipun beberapa peneliti telah mengatasi masalah yang menantang ini, sistem yang kuat masih sulit dipahami (Mohammed et al., 2019). Pengembangan lain melalui algoritma segmentasi kulit manusia dengan deteksi otomatis berdasarkan informasi warna. Ruang warna YCbCr digunakan dalam pengkodean video dan menyediakan penggunaan informasi *chrominance* yang efektif untuk memodelkan warna kulit manusia. Performa algoritma diilustrasikan dengan simulasi yang dilakukan pada gambar yang menggambarkan orang dari etnis yang berbeda. Kemudian *Convolutional Neural Network* (CNN) digunakan untuk mengekstrak fitur dari gambar dan Metode *Deep Learning* digunakan untuk melatih *classifier* untuk mengenali Bahasa Isyarat (Shahriar et al., 2019).

Kekurangan dari penelitian yang pernah dilaksanakan sebelumnya diketahui bahwa belum adanya jumlah dan variasi dataset pada *neural network* agar prediksi *neural network* dapat menjadi lebih baik, juga perlu pengembangan proses segmentasi dalam menghilangkan *background* dibelakang objek dapat dihasilkan *output* yang lebih baik, dan mencoba pengujian lainnya agar mendapat model yang lebih baik lagi sehingga bisa membangun pengembangan lanjut (Aljabar & Suharjito, 2020; Ikram et al., 2018). Pengembangan deteksi yang disarankan dengan pendekatan *machine learning* yang handal saat ini yaitu *deep learning*, sebuah metode pembelajaran mesin berbasis pada pembelajaran data yang diperluas (Papastratis et al., 2021). Mengubah bahasa isyarat menjadi bentuk alami bahasa adalah salah satu area terbaru dari domain pembelajaran mesin. Berbagai penelitian berfokus pada mengkategorikan bahasa isyarat ke dalam pengenalan isyarat atau wajah (Amin et al., 2021).

Tujuan penelitian ini adalah untuk mengetahui faktor akurasi yang mempengaruhi tingkat akurasi penerapan objek deteksi dan klasifikasi gambar maupun video secara *realtime* untuk BISINDO dengan menggunakan metode LSTM dan CNN *Deep Learning*.

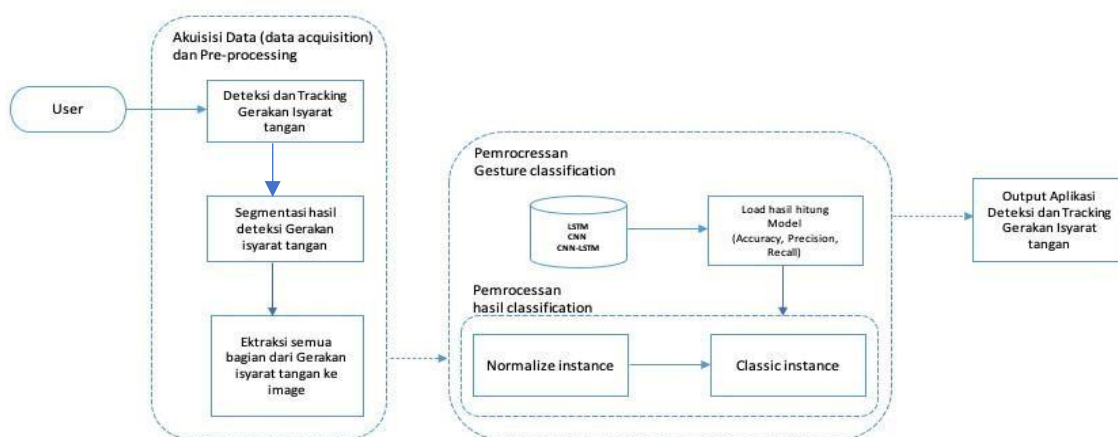
METODE

Peneliti melakukan studi dan membaca buku, jurnal, makalah, dan contoh-contoh tulisan yang berkaitan dengan topik kajian. Selanjutnya mengumpulkan informasi dan data dengan melakukan tugas-tugas antara lain mendokumentasikan, mengamati, dan mencatat untuk memenuhi kebutuhan data. Selanjutnya, proses pendefinisian kebutuhan dilakukan dengan mengidentifikasi data yang dibutuhkan, memantau kemajuan proses, menganalisis sistem saat berjalan, dan membuat hasil evaluasi sistem. *Preprocessing* adalah langkah selanjutnya, dan banyak data yang telah dikumpulkan sedang diproses berupa objek seperti pelatihan data, posisi objek, pose, pencahayaan, dan latar belakang objek. Preprocessing berguna untuk mengekstrak informasi dari ulasan, karena dapat membantu data yg tidak terstruktur menjadi bentuk standar. Disisi lain, *preprocessing* sangat penting dan kritis dalam data mining untuk image. Setelah dilakukan prepoessing tahapan yang dilakukan adalah proses dengan menggunakan metode LSTM dan CNN *deep learning* untuk mencari tingkat nilai keakuratan dari gestur masukan sehingga dapat diketahui arti dari gestur pada BISINDO. Setelah itu dilakukan evaluasi model, evaluasi model dilakukan perhitungan, tingkat akurasi, presisi, dan *recall*. Tempat pelaksanaan penelitian dilaksanakan di Universitas Muhammadiyah Bangka Belitung dan akan dilaksanakan untuk pengumpulan data sesuai pada jadwal tabel penelitian. Pengembangan deteksi *real-time* BISINDO menggunakan metode LSTM dan CNN *Deep learning*. Rincian pelaksanaan penelitian pengembangan deteksi ini dapat dilihat pada Gambar 1 berikut ini:



Gambar 1. Alur Penelitian

Untuk alur prototipe pengembangan deteksi realtime untuk BISINDO (Bahasa Isyarat Indonesia) dengan menggunakan metode LSTM dan CNN *Deep Learning* dapat dilihat pada Gambar 2 berikut ini:



Gambar 2. Alur untuk *prototipe* hasil pengembangan

HASIL DAN PEMBAHASAN

Sebelum diproses, data diperoleh menggunakan kamera setinggi tiga kaki dengan resolusi HD 720p. Dengan menggunakan metode *preprocessing*, kamera merekam objek dan mengarahkannya untuk menyimpan objek sebagai dataset (*background subtraction* dan *gaussian blur*). Setiap data gambar berukuran 50 x 50 piksel. Setiap item ditugaskan 1000 untuk pelatihan dan 100 untuk validasi yang dapat dilihat pada Gambar 3 berikut.

```
#load_dataset function to load the data and resize the images to 50x50
def load_dataset(directory):
    images = []
    labels = []
    for idx, label in enumerate(uniq_labels):
        for file in os.listdir(directory + '/' + label):
            filepath = directory + '/' + label + "/" + file
            img = cv2.resize(cv2.imread(filepath), (50,50))
            images.append(img)
            labels.append(idx)
    images = np.asarray(images)
    labels = np.asarray(labels)
    return images, labels
```

Gambar 3. Dataset yang telah dimasukkan ke python untuk dikelola

Dari sekumpulan data ditangkap dan dikonversi menjadi foto skala abu-abu dengan mengurangi latar belakang. Namun, preprocessing menggunakan bobot 0,5 dan gaussian blur 7. Tahap pertama adalah *background subtraction* untuk *preprocessing*. Kumpulan data yang telah diubah menjadi objek abu-abu (objek tangan). Untuk melakukan ini, skala gambar diubah menjadi 50x50 piksel dan semua kumpulan data kemudian dinormalisasi sama seperti sebelum dataset diolah. Dataset pelatihan terdiri dari 1.000 foto untuk setiap objek. Dan validasi data terdiri dari 100 foto per objek yang tersaji pada Gambar 4 berikut.

```
def display_images(x_data,y_data, title, display_label = True):  
    x, y = x_data,y_data  
    fig, axes = plt.subplots(5, 8, figsize = (18, 5))  
    fig.subplots_adjust(hspace = 0.5, wspace = 0.5)  
    fig.suptitle(title, fontsize = 18)  
    for i, ax in enumerate(axes.flat):  
        ax.imshow(cv2.cvtColor(x[i], cv2.COLOR_BGR2RGB))  
        if display_label:  
            ax.set_xlabel(uniq_labels[y[i]])  
            ax.set_xticks([])  
            ax.set_yticks([])  
    plt.show()
```

Gambar 4. Proses Preprocessing

Selanjutnya dibuat arsitektur dari model dengan menggunakan 3 lapis *convolution* dan 2 *hidden layer* dengan sebanyak 128 yang tersaji pada Gambar 5 berikut.

```
# building our model  
model = tf.keras.Sequential([  
    tf.keras.layers.Conv2D(16, (3,3), activation='relu', input_shape=(50,50,3)),  
    tf.keras.layers.Conv2D(16, (3,3), activation='relu'),  
    tf.keras.layers.Conv2D(16, (3,3), activation='relu'),  
    tf.keras.layers.MaxPool2D((2,2)),  
    tf.keras.layers.Conv2D(32, (3,3), activation='relu'),  
    tf.keras.layers.Conv2D(32, (3,3), activation='relu'),  
    tf.keras.layers.Conv2D(32, (3,3), activation='relu'),  
    tf.keras.layers.MaxPool2D((2,2)),  
    tf.keras.layers.Conv2D(64, (3,3), activation='relu'),  
    tf.keras.layers.Conv2D(64, (3,3), activation='relu'),  
    tf.keras.layers.Conv2D(64, (3,3), activation='relu'),  
    tf.keras.layers.Flatten(),  
    tf.keras.layers.Dense(128, activation='relu'),  
    tf.keras.layers.Dense(25, activation='softmax')  
)  
  
model.summary()
```

Gambar 5. Model *Deep Learning* untuk Deteksi *Realtime BISINDO*

Selanjutnya dilakukan pengetesan model *deep learning* untuk BISINDO sebelum diintegrasikan ke Android yang tersaji pada Gambar 6.

Deep learning merupakan bagian dari *machine learning* yang menggunakan *Deep Neural Networks* untuk menyelesaikan permasalahan pada domain *machine learning* (Alzubaidi et al., 2021; Papastratis et al., 2021) yang menirukan cara berpikir manusia. *Deep learning* merupakan metode yang memanfaatkan *Artificial Neural Networks* yang berlapis-lapis (*multi layer*). *Artificial Neural Networks* ini dibuat mirip dengan otak manusia, *neuron-neuron* terkoneksi satu sama lain sehingga membentuk sebuah jaringan *neuron* yang sangat rumit dengan memanfaatkan *multiple nonlinear transformation*.

Setelah model diluncurkan, diperoleh hasil, dan dipahami bahwa setiap lapisan kompleksitas pada model menggunakan filter 3x3. Aturan dalam arsitektur CNN dan LSTM untuk dataset BISINDO telah disesuaikan agar sesuai dengan *rasio hit* yang diinginkan antara filter dan input. Contohnya adalah pada jumlah saluran. Pada input diketahui bahwa input gambar memiliki 3 channel, maka setiap filter harus memiliki 3 channel juga agar sesuai

dengan aturan arsitektur CNN dan LSTM. Jika dicetak semua nilai bobot, kemungkinan nilai bobot bisa bernilai positif ataupun negatif yang dicantumkan pada 0,0. Untuk itu, perlu mengkonversi nilai-nilai tersebut menjadi range 0-1 dengan menggunakan min-max standardization.

```
Output exceeds the size limit. Open the full output data in a text editor
Model: "sequential"

```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 16)	448
conv2d_1 (Conv2D)	(None, 46, 46, 16)	2320
conv2d_2 (Conv2D)	(None, 44, 44, 16)	2320
max_pooling2d (MaxPooling2D)	(None, 22, 22, 16)	0
conv2d_3 (Conv2D)	(None, 20, 20, 32)	4640
conv2d_4 (Conv2D)	(None, 18, 18, 32)	9248
conv2d_5 (Conv2D)	(None, 16, 16, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 8, 8, 32)	0
conv2d_6 (Conv2D)	(None, 6, 6, 64)	18496
conv2d_7 (Conv2D)	(None, 4, 4, 64)	36928
conv2d_8 (Conv2D)	(None, 2, 2, 64)	36928
...		
Total params:		156,697
Trainable params:		156,697
Non-trainable params:		0

Gambar 6. Output hitungan

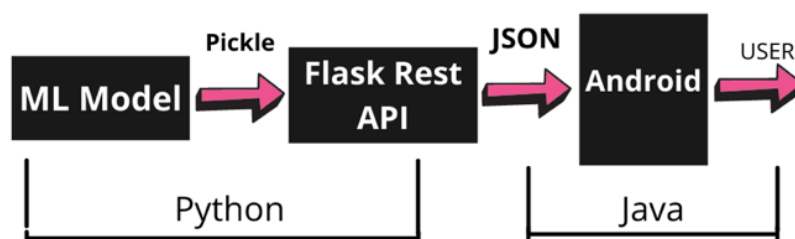
Deep learning memungkinkan untuk menjelajahi representasi data pada berbagai tingkat abstraksi menggunakan model komputer yang terdiri dari beberapa lapisan pemrosesan. Metode ini telah sangat meningkatkan kecanggihan dalam pengenalan ucapan, pengenalan objek visual, pengenalan objek, dan banyak bidang lainnya seperti penemuan obat dan genomik (Kurniawan & Mustikasari, 2021). *Deep learning* menggunakan algoritma *backpropagation* untuk memanipulasi kumpulan data besar dengan menentukan bagaimana mesin memodifikasi parameter internal yang digunakan untuk menghitung representasi setiap lapisan dari representasi lapisan sebelumnya. Jaringan *deep convolutional* menyajikan variasi yang berbeda dalam memproses gambar, video, ucapan, dan audio, dan jaringan iteratif menyajikan data berurutan seperti teks dan ucapan. (Chevtchenko et al., 2018; Lecun et al., 2015).

Sistem pengenalan bahasa isyarat memfasilitasi komunikasi antara komunitas tunarungu dan mayoritas pendengar. Penelitian ini mengusulkan model baru *convolutional neural network* (CNN), SignNet, untuk mengenali tanda-tanda gerakan tangan dengan memasukkan teori ruang skala ke dalam kerangka pembelajaran yang mendalam. CNN

membuktikan bahwa pendekatan melalui pendeteksian tangan dan klasifikasi gerakan memberikan arsitektur yang efisien dan kuat (Mohammed et al., 2019). Model yang diusulkan adalah ansambel rata-rata tertimbang CNN - jaringan resolusi rendah (LRN), jaringan resolusi menengah (IRN) dan jaringan resolusi tinggi (HRN). Versi tambahan dari VGG-16 digunakan sebagai LRN, IRN dan HRN. Ansambel bekerja pada resolusi spasial yang berbeda dan pada kedalaman CNN yang berbeda-beda. Model SignNet dinilai dengan tanda statis Bahasa Isyarat Amerika – huruf dan angka. Karena tidak ada kumpulan data tanda untuk pembelajaran mendalam, kinerja ansambel dievaluasi pada kumpulan data sintetik yang telah dikumpulkan untuk tugas ini. Penilaian kumpulan data sintetik oleh SignNet melaporkan akurasi yang mengesankan lebih dari 92%, terutama lebih unggul dari model lain yang ada (Aloysius & Geetha, 2020).

Model *encoder-decoder* berbasis perhatian dengan multi-channel convolutional neural network (CNN) untuk mewujudkan SLR kontinu yang akurat, terukur, dan *end-to-end* tanpa segmentasi tanda. Telah diterapkan DeepSLR pada smartphone dan mengevaluasi keefektifannya melalui evaluasi ekstensif. Rata-rata tingkat kesalahan kata dari pengenalan kalimat terus menerus adalah 10,8%, dan dibutuhkan waktu kurang dari 1,1 detik untuk mendeteksi sinyal dan mengenali kalimat dengan 4 kata tanda, memvalidasi efisiensi pengenalan dan kemampuan DeepSLR *real-time* dalam skenario dunia nyata (Wang et al., 2022). SLR mampu mengklasifikasikan secara spesifik untuk membantu komunikasi melalui kerangka dasar yang telah disajikan dalam *deep learning* melalui kombinasi fitur video maupun skeletal (Kenstantinidis et al., 2018; Tarimo et al., 2020).

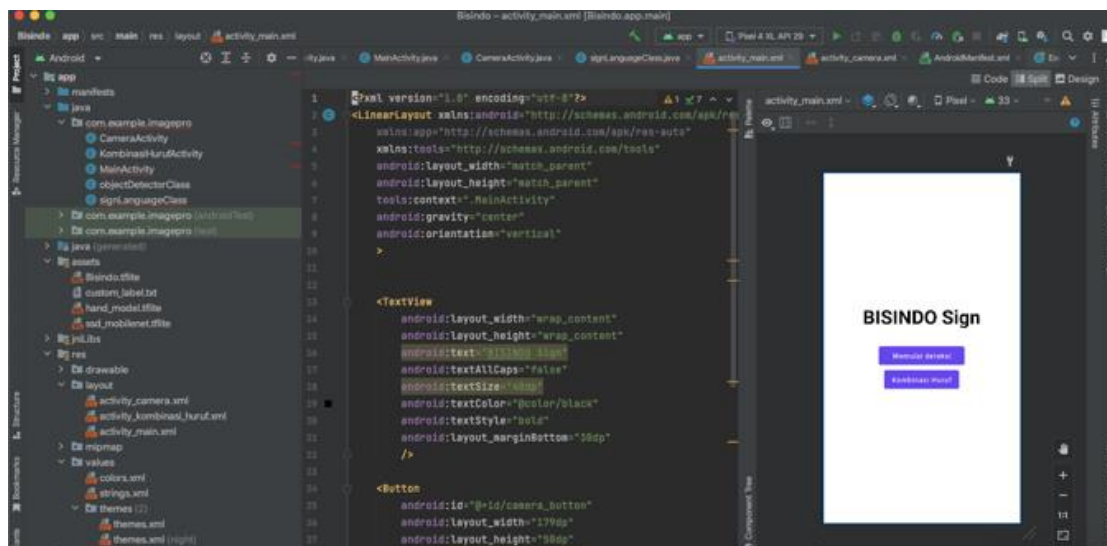
Setelah model BISINDO ditraining dengan menggunakan metode CNN dan LSTM *Deep Learning*, langkah selanjutnya memproses model tersebut ke dalam sebuah file *Tflite*, kemudian dikembangkan pada program *android*. Studi untuk bisa dijadikan sebuah program *prototype* awal yang bisa digunakan untuk aplikasi BISINDO untuk Tuna Wicara. Berikut proses *deploy* dari AI model ke sebuah aplikasi *android* yang tersaji pada Gambar 7-8 berikut:



Gambar 7. Proses AI Model ke Android

Android studio dikembangkan dari IntelliJ IDEA, sebuah IDE buatan JetBrains yang juga memproduksi aplikasi seperti PHPStorm dan PyCharm. Sebagai IDE, tentu saja

Android Studio memiliki komponen yang lengkap meliputi source code editor, compiler dan debugger. Android Studio diluncurkan pada tanggal 16 Mei 2013 pada acara Google I/O 2013. Sebelumnya, Google memberikan dukungan terhadap *Eclipse IDE*. *Plugin Eclipse Android Developer Tool* (ADT) dibuat agar *Eclipse* dapat melakukan kompilasi dan uji coba menjalankan proyek Android.



Gambar 8. Proses Pengembangan ke Android

Sebuah sistem *real-time* adalah dimana fungsi dari sistem bergantung pada hasil yang dihasilkan oleh sistem dan waktu (Mujahid et al., 2021). Suatu sistem yang *real-time* akan memungkinkan komputer untuk mengikuti sebuah proses yang berubah secara terus menerus. Pada sistem *real-time*, waktu merupakan hal yang dianggap paling penting tetapi bukan berarti apabila sistem tersebut tidak bisa berjalan dengan cepat maka sistem tersebut gagal. Sebuah sistem dapat dikatakan *real-time* jika sistem tersebut mampu memenuhi batasan respon yang dibutuhkan. Tetapi batasan waktu pada sistem tersebut harus tegas agar tidak terjadi penyimpangan. Keakuratan dari hasil segmentasi penggunaan *CIELab* dan *Deep Learning* pada pengenalan huruf alfabet (a-z) BISINDO dengan mendeteksi pergerakan tangan secara *real-time* menggunakan webcam sebagai perangkat yang tertanam pada laptop dan mengeluarkan hasil *recognition* kedalam sebuah *text* menghasilkan keberhasilan sebesar 60% atau lebih tepatnya 59.54% (Damatraseta et al., 2021).

Pentingnya penelitian ini karena hasilnya dapat dijadikan dasar untuk mempercepat pengembangan lebih lanjut aplikasi *sign language recognition* khusus untuk BISINDO yang bisa dimanfaatkan oleh penyandang disabilitas maupun masyarakat agar komunikasi dua arah lebih mudah dilakukan dimasa depan secara *realtime* (Yang et al., 2017). Penelitian ini memberikan kemudahan bagi penyandang disabilitas untuk berkomunikasi dalam

sekelompok masyarakat melalui beberapa bahasa isyarat dengan variabilitas berbeda misalnya bentuk tangan, profil gerak, posisi tangan, wajah, dan bagian tubuh yang berkontribusi pada setiap tanda (Rastgoo et al., 2021; Schmidhuber, 2014; Vo et al., 2019)

KESIMPULAN DAN REKOMENDASI

Pesatnya kecerdasan tiruan dan kemampuan metode *deep learning* mampu memberikan manfaat untuk mengembangkan pendekatan pembelajaran mendalam untuk prediksi tangan bahasa isyarat. Prediksi Bahasa isyarat memanfaatkan metode CNN + LSTM dan fokus pada filtering, layer, dan objek BISINDO. Menurut model pelatihan, validasi, dan pengujian, CNN+LSTM adalah model yang optimal untuk item ini. Dengan model tersebut, akurasi mencapai 90%, kehilangan 19%, dan pengujian 80%. Untuk penelitian lanjutan dimasa depan direkomendasikan untuk menambah masukan dari deteksi ekspresi wajah, deteksi gerakan tubuh, dan deteksi mulut.

REFERENSI

- Aljabar, A., & Suharjito. (2020). BISINDO (Bahasa isyarat indonesia) sign language recognition using CNN and LSTM. *Advances in Science, Technology and Engineering Systems*, 5(5), 282–287. <https://doi.org/10.25046/AJ050535>
- Aloysius, N., & Geetha, M. (2020). A scale space model of weighted average CNN ensemble for ASL fingerspelling recognition. *International Journal of Computational Science and Engineering*, 22(1), 154–161. <https://doi.org/10.1504/IJCSE.2020.107268>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1), 1–74. <https://doi.org/10.1186/s40537-021-00444-8>
- Amin, M., Hefny, H., & Mohammed, A. (2021). Sign Language Gloss Translation using Deep Learning Models. *International Journal of Advanced Computer Science and Applications*, 12(11), 686–692. <https://doi.org/10.14569/IJACSA.2021.0121178>
- Chevtchenko, S. F., Vale, R. F., Macario, V., & Cordeiro, F. R. (2018). A convolutional neural network with feature fusion for real-time hand posture recognition. *Applied Soft Computing Journal*, 73, 748–766. <https://doi.org/10.1016/j.asoc.2018.09.010>
- Damatraseta, F., Novariany, R., & Ridhani, M. A. (2021). Real-time BISINDO Hand Gesture Detection and Recognition with Deep Learning CNN. *Jurnal Informatika Kesatuan*, 1(1), 71–80. <https://doi.org/10.37641/jikes.v1i1.774>

- Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, 32(4), 582–596. <https://doi.org/10.1007/s10278-019-00227-x>
- Ikram, K., Khairunizam, W., Aziz, A. A., Bakar, S. A., Razlan, Z. M., Zunaidi, I., & Desa, H. (2018). Adaptive gesture recognition based on human physical characteristic. *Proceedings - 2018 IEEE 14th International Colloquium on Signal Processing and Its Application, CSPA 2018, March*, 129–134. <https://doi.org/10.1109/CSPA.2018.8368699>
- Irwanto, Kasim, E. R., Fransiska, A., Lusli, M., & Siradj, O. (2010). Analisis Situasi Penyandang Disabilitas di Indonesia: Sebuah Desk Review. In *Pusat Kajian Disabilitas Fakultas Ilmu Sosial dan Politik UI* (Vol. 1, Issue S2). <https://doi.org/10.5694/j.1326-5377.1981.tb135719.x>
- Kenstantinidis, D., Dimitropoulos, K., & Daras, P. (2018). A Deep Learning Approach for Analyzing Video and Skeletal Features in Sign Language Recognition. *IEEE*, 1–6.
- Kurniawan, A. A., & Mustikasari, M. (2021). Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia. *Jurnal Informatika Universitas Pamulang*, 5(4), 544. <https://doi.org/10.32493/informatika.v5i4.6760>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Mohammed, A. A. Q., Lv, J., & Islam, M. D. S. (2019). A deep learning-based end-to-end composite system for hand detection and gesture recognition. *Sensors (Switzerland)*, 19(23), 1–23. <https://doi.org/10.3390/s19235282>
- Mujahid, A., Awan, M. J., Yasin, A., Mohammed, M. A., Damaševičius, R., Maskeliūnas, R., & Abdulkareem, K. H. (2021). Real-time hand gesture recognition based on deep learning YOLOv3 model. *Applied Sciences (Switzerland)*, 11(9), 1–15. <https://doi.org/10.3390/app11094164>
- Papastratis, I., Chatzikonstantinou, C., Konstantinidis, D., Dimitropoulos, K., & Daras, P. (2021). Artificial intelligence technologies for sign language. *Sensors*, 21(17), 1–25. <https://doi.org/10.3390/s21175843>
- Rakun, E., Andriani, M., Wiprayoga, I. W., Danniswara, K., & Tjandra, A. (2013). Combining depth image and skeleton data from Kinect for recognizing words in the sign system for Indonesian language (SIBI [Sistem Isyarat Bahasa Indonesia]). *2013 International Conference on Advanced Computer Science and Information Systems*,

ICACISIS 2013, 387–392. <https://doi.org/10.1109/ICACISIS.2013.6761606>

- Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign Language Recognition: A Deep Survey. *Expert Systems with Applications*, 164(February 2020), 113794. <https://doi.org/10.1016/j.eswa.2020.113794>
- Schmidhuber, J. (2014). Deep Learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Shahriar, S., Siddiquee, A., Islam, T., Ghosh, A., Chakraborty, R., Khan, A. I., Shahnaz, C., & Fattah, S. A. (2019). Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2018-Octob(October), 1168–1171. <https://doi.org/10.1109/TENCON.2018.8650524>
- Sharma, S., & Singh, S. (2021). Vision-based hand gesture recognition using deep learning for the interpretation of sign language. *Expert Systems with Applications*, 182(July), 1–12. <https://doi.org/10.1016/j.eswa.2021.115657>
- Subburaj, S., & Murugavalli, S. (2022). Survey on sign language recognition in context of vision-based and deep learning. *Measurement: Sensors*, 23(May), 1–11. <https://doi.org/10.1016/j.measen.2022.100385>
- Suharjito, Gunawan, H., Thiracitta, N., & Nugroho, A. (2019). Sign Language Recognition Using Modified Convolutional Neural Network Model. *1st 2018 Indonesian Association for Pattern Recognition International Conference, INAPR 2018 - Proceedings*, 1–5. <https://doi.org/10.1109/INAPR.2018.8627014>
- Suharjito, Thiracitta, N., & Gunawan, H. (2021). SIBI Sign Language Recognition Using Convolutional Neural Network Combined with Transfer Learning and non-trainable Parameters. *Procedia Computer Science*, 179(2019), 72–80. <https://doi.org/10.1016/j.procs.2020.12.011>
- Supria, S., Herumurti, D., & Khotimah, W. N. (2016). Pengenalan Sistem Isyarat Bahasa Indonesia Menggunakan Kombinasi Fitur Statis Dan Fitur Dinamis Lmc Berbasis L-Gcnn. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 14(2), 217. <https://doi.org/10.12962/j24068535.v14i2.a574>
- Tarimo, W., Sabra, M. M., & Hendre, S. (2020). Real-Time Deep Learning-Based Object Detection Framework. *IEEE Symposium Series on Computational Intelligence (SSCI)*, 1829–1836. <https://medium.com/@arifwicaksanaa/pengertian-use-case-a7e576e1b6bf>
- Vo, A. H., Pham, V. H., & Nguyen, B. T. (2019). Deep learning for Vietnamese Sign

Language recognition in video sequence. *International Journal of Machine Learning and Computing*, 9(4), 440–445. <https://doi.org/10.18178/ijmlc.2019.9.4.823>

Wang, Z., Zhao, T., Ma, J., Chen, H., Liu, K., Shao, H., Wang, Q., & Ren, J. (2022). Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System. *IEEE Transactions on Mobile Computing*, 21(7), 2398–2410. <https://doi.org/10.1109/TMC.2020.3038303>

Yang, J., Zhu, C., & Yuan, J. (2017). Real time hand gesture recognition via finger-emphasized multi-scale description. *Proceedings - IEEE International Conference on Multimedia and Expo*, July, 631–636. <https://doi.org/10.1109/ICME.2017.8019348>