IDENTIFIKASI KOSAKATA BAHASA ISYARAT INDONESIA (BISINDO) MENGGUNAKAN METODE VISION TRANSFORMER (ViT) DAN JARINGAN SARAF TIRUAN BERBASIS MODEL SKELETON

PROPOSAL PENELITIAN

Marsia Yohana Apriani Loblar

PROGRAM DOKTOR TEKNOLOGI INFORMASI

UNIVERSITAS GUNADARMA

JAKARTA 2024

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I

PENDAHULUAN

1.1 Latar Belakang

Organisasi Kekayaan Intelektual Dunia atau yang dikenal dengan *World Intellectual Property Organization* (WIPO) menyatakan bahwa kecerdasan artifisial adalah suatu cabang ilmu komputer yang bertujuan menghasilkan mesin dan sistem yang mampu menjalankan tugas-tugas yang memerlukan tingkat kecerdasan manusia (WIPO, 2019). Kecerdasan artifisial adalah kumpulan berbagai alat bantu untuk membuat komputer dapat bekerja secara cerdas (Sooai, A.G., Mandira, W., Magdalena, N., Mamualak, R., Nani, P.A., 2021). Pengembangan kecerdasan artifisial telah mencapai kemampuan untuk melakukan penalaran secara mandiri. Penggunaan kecerdasan artifisial telah meningkat secara signifikan dalam dekade terakhir, Indonesia mendapati banyak kesempatan untuk memanfaatkan kecerdasan artifisial, yang bisa meningkatkan produktivitas usaha, efisiensi dalam investasi sumber daya manusia, serta menghasilkan inovasi di beragam bidang seperti keuangan, kesehatan, pendidikan, pertanian, keamanan, transportasi, dan kelautan. Kecerdasan artifisial menawarkan solusi untuk mengurangi biaya infrasturktur, meningkatkan efektivitas layanan kesehatan dan sosial, merancang sumber daya pendidikan berkualitas, mendukung pemerintah dalam pengambilan kebijakan yang akurat, memperluas pasar digital, dan memperbaiki kualitas layanan publik (Wisjnu, Sri Saraswati, Wardhani, Ismunandar, Purwoadi, Michael A., Nugroho, Anto S., 2020).

Dalam pengembangan kecerdasan artifisial, terdapat beberapa pendekatan yang digunakan untuk menciptakan sistem yang cerdas. Pembelajaran Mesin merupakan istilah yang sudah tidak asing dan dipakai dalam berbagai aplikasi kecerdasan artifisial yang digunakan untuk memecahkan berbagai masalah (Roihan, A., Sunarya, P.A., Rafika, A.S., 2020). Pembelajaran mesin merupakan salah satu cabang atau bagian dari kecerdasan artifisial. Tugas dari pembelajaran mesin adalah melatih pembelajaran mesin agar dapat mempelajari data historis

untuk menemukan tren jaringan (Putri, V.A., Sotyawardani, K.C.A., Rafael, R.A., 2023). Pendekatan lainnya dalam kecerdasan artifisial adalah pembelajaran mendalam yang merupakan sub-bidang dari pembelajaran mesin yang algoritmanya terinspirasi oleh struktur otak manusia. Pembelajaran mendalam memberikan landasan komprehensif yang kuat dengan menggunakan jaringan saraf tiruan berlapis-lapis, sehingga dimungkinkan model membangun pemahaman tentang data secara bertahap dalam konsep yang paling sederhana ke yang paling kompleks (Heaton, 2018).

Bahasa isyarat yang digunakan di Indonesia dalam penerapannya di masyarakat ada dua bentuk, yaitu Bahasa Isyarat Indonesia atau yang dikenal dengan nama BISINDO dan Sistem Isyarat Bahasa Indonesia atau yang dikenal dengan SIBI. Bahasa isyarat di Indonesia lahir dari pendekatan budaya lokal dan pengaruh kejadian sehari-hari di lingkungan (Antara, 2014), (Sugianto & Samopa, 2015). Menurut (Kautsar, I., Borman, R.I., Sulistyawati, Ari, 2015), bahasa isyarat Indonesia atau BISINDO dikembangkan oleh orang-orang tunarungu sendiri melalui Gerkatin (Gerakan Kesejahteraan Tuna Rungu Indonesia) sebuah organisasi yang memperjuangkan BISINDO menjadi bahasa nasional dengan landasan bahasa ibu yang sudah dipraktekkan secara alamiah sejak kecil dalam keluarga tunarungu. Jumlah penduduk berumur 5 tahun keatas menurut kelompok umur, daerah perkotaan/pedesaan, jenis kelamin dan tingkat kesulitan mendengar di Indonesia pada tahun 2022 tercatat ada 253.679.348 jiwa (Badan Pusat Statistik, 2022). Terdapat kesenjangan komunikasi bagi mereka yang mengalami gangguan pendengaran atau bicara ketika berinteraksi dengan individu yang memiliki kemampuan komunikasi normal. Pemahaman makna dan persepsi antara dua kelompok ini memerlukan pemahaman yang kompleks, terutama bagi individu yang tidak pernah mempelajari bahasa isyarat.

Kosakata merupakan salah satu materi pembelajaran bahasa Indonesia di sekolah yang menempati peran sangat penting sebagai dasar siswa untuk menguasai materi mata pelajaran bahasa Indonesia dan penguasaan mata pelajaran lainnya (Kasno, 2004 dalam Pramesti, 2015). Penguasaan kosakata memengaruhi cara berpikir dan kreativitas siswa dalam proses pembelajaran bahasa sehingga penguasaan kosakata dapat menentukan kualitas seorang siswa dalam berbahasa

(Kasno, 2004 dalam Pramesti, 2015). Basa-basi atau *phatic communion* didefinisikan sebagai ungkapan atau tuturan yang dipergunakan hanya untuk sopan santun dan tidak untuk menyampaikan informasi (Crystal, 1991:257). Dalam bahasa Inggris ada ahli yang menyebutkan istilah *phatic communion* atau komunikasi fatis atau basa-basi yang memiliki arti sebagai pertuturan ungkapan baku, seperti halo, apa kabar, dan lain-lain yang tidak mempunyai makna, dalam arti untuk menyampaikan informasi melainkan dipergunakan untuk mengadakn kontak sosial di antara pembicara atau untuk menghindari kesenyapan yang menimbulkan rasa kikuk (Crystal, 1991:257). Pada umumnya, kategori fatis digunakan dalam ragam lisan, baik standar maupun nonstandar. Kata maaf, tolong, terimakasih dan sama-sama dapat dikaitkan dengan perilaku budaya basa-basi (Setyadi, Ari, 2021). Keberadaan kata tolong, maaf, terima kasih bersifat fungsional dan strategis demi kepentingan tertentu dalam sebuah tuturan dan termasuk dalam 3 kata ajaib yang mampu menciptakan cerminan perilaku budaya berbahasa dalam berkomunikasi (Setyadi, Ari, 2021).

Implementasi pemanfaatan kecerdasan artifisial dalam bidang teknologi informasi dapat membantu proses pengolahan data, analisis prediktif, keamanan informasi, analisis sentimen, pengenalan wajah dan suara, pemrosesan bahasa alami, pengembangan perangkat lunak dan pengujian, pengelolaan sumber daya IT, sistem rekomendasi, dan memberikan manfaat besar dalam meningkatkan efisiensi, produktivitas, dan kemampuan pengambilan keputusan dalam konteks teknologi informasi. Salah satu pemanfaatan kecerdasan artifisial dalam bidang teknologi informasi bahasa isyarat adalah untuk mendeteksi dan mengklasifikasi 26 huruf dan 10 angka BISINDO menggunakan model CNN yang disederhanakan dengan membandingkan AlexNet dan VGG-16 yang dilakukan oleh Dwijayanti, (Dwijayanti, S., Hermawati, Taqiyyah, Sahirah Inas, Hikmarika, Hera, Suprapto, Bhakti Yudho 2021), mengusulkan pendekatan *deep learning* yaitu dengan membuat model CNN baru yang diberi nama model C (Dwijayanti, Suci et al., 2021) untuk mengenali BISINDO yang terdiri dari 26 huruf dan 10 angka. Penelitian ini memiliki tujuan untuk membandingkan kinerja pengenalan BISINDO dari model CNN yang disederhanakan dengan AlexNet dan VGG-16. Model CNN C yang diusulkan bekerja dengan baik dalam memprediksi gerakan tangan dengan

nilai akurasi 98,3%. (Ahmad, Nizhamuddin, Wijaya, Eko Saputra, Tjoaquinn, Calvin, Lucky, Henry, Iswanto, Irene Anindaputri 2023) melakukan penelitian menggunakan model CNN (*convolutional neural netwrok*) untuk pengenalan BISINDO dengan tujuan untuk mengembangkan model pengenalan bahasa isyarat yang akurat dan efisien berdasarkan dataset BISINDO. CNN dipilih karena kemampuannya mengesktraksi fitur spasial dari data video bahasa isyarat. Subjek dataset yang digunakan adalah 26 gestur tangan dari huruf A sampai Z dalam bahasa isyarat Indonesia (BISINDO) dengan total 936 citra gambar.

Penelitian lain dengan metode berbeda (Basri, Syartina Elfarika, Indra, Dolly, Darwis, Herdianti, Mufila, A. Widya, Ilmawan, Lutfi Budi, Purwanto, Bobby, 2021) melakukan penelitian menggunakan metode *Fourier Descriptor* yang digunakan untuk mengekstraksi fitur citra Bahasa Isyarat Indonesia (BISINDO) untuk pengenalan huruf abjad. Berdasarkan hasil pengujian, *Fourier Descriptor* dapat digunakan untuk mengekstraksi citra gambar huruf BISINDO dan semakin tinggi koefisiennya maka semakin akurat hasil pengenalannya. Hal ini dibuktikan dengan nilai akurasi terbaik diperoleh pada koefiesien 25 dan 50 dengan persamaan akurasi 96,92%. Sementara itu, hasil kombinasi dari *Fourier Descriptor* dan *Euclidean Distance* masih dinilai cukup untuk mengenali citra *standard* dengan nilai akurasi 74,15% dan citra *scale* dengan nilai akurasi 72,30%, sedangkan untuk citra *rotation* mendapat nilai akurasi 57,43% dan citra *translation* mendapatkan nilai akurasi terendah sebesar 34,36%. (Kharat, A., Patil, Y., Jagtap, O., Sonawale, R., 2022) mengembangkan metode *real time* dengan menerapkan *convolutional neural network* (CNN) untuk *American Sign Language* (ASL) 26 huruf alfabet. Penelitian ini menggunakan pendekatan berbasis penglihatan (*vision based approach*) dan mengumpulkan dataset dengan menangkap 800 citra pada tiap simbol ASL untuk data latih dan 200 citra pada setiap simbol untuk data uji. Pada penelitian ini juga menerapkan *Gaussian Blur* yang digunakan pada saat *input citra*. Nilai akurasi akhir yang didapat pada penelitian ini sebesar 98,0% dengan melakukan peningkatan prediksi. Penelitian ini dapat memverifikasi dan memprediksi simbol yang memiliki kemiripan yang hampir sama, sehingga kelebihan dari penelitian ini adalah dapat mendeteksi hampir semua simbol huruf dengan catatan huruf tersebut ditampilkan dengan posisi yang benar, tidak ada *noise* pada *background*, dan

pencahayaan yang memadai. (Enri, U., Rozikin, C., Ilhamsyah, M., Irawan, Agung Susilo Yuda, Garno, Solihin, Indra Permana, Jayanta, 2023) Penelitian ini mengimplementasikan LSTM (*Long Short Term Memory*) dan *Mediapipe* untuk mengidentifikasi gerakan bahasa isyarat BISINDO dengan tujuan mengembangkan model *deep learning* (Enri, Ultach et.al., 2023). Penelitian ini menggunakan metode *deep learning* dengan arsitektur LSTM dan data sekuensial. *Mediapipe* digunakan untuk ekstraksi fitur pada setiap citra gerak isyarat. Model arsitektur menggunakan 6 *layer* yang terdiri dari 3 *layer* LSTM dan 3 *layer dense*. Tahapan penelitian dimulai dari pengumpulan data, dimana data dikumpulkan dari video demonstrasi gerakan bahasa isyarat yang terdiri dari 5 kelas kata ganti orang, yaitu: "Saya", "Kamu", "Dia", "Kami", dan "Mereka". Setiap kata diucapkan sebanyak 550 kali dengan total gerakan 2.750, data divalidasi oleh 4 laki-laki dan 2 perempuan penutur asli dan diukur intensitas cahaya saat pengambilan citra. Skenario pertama model memiliki kinerja luar biasa dengan akurasi 99% dan 89% untuk tes dan data aktual, sedangkan skor ROC-AUC adalah 99,995% dan 98,390%.

Penelitian dengan menggunakan metode *Vision Transformer* (ViT) dilakukan oleh (Agrawal, Agrima, Sreemathy, R., Turuk, Mousami, Jagdale, Jayashree, Kumar, Vishal, 2023) yang mengembangkan sebuah model pengenalan bahasa isyarat India (*Indian Sign Language*) yang efektif menggunakan teknologi *Skin Segmentation* dan *Vision Transformer* dengan menggunakan dataset primer yang berisi 72 kata dalam bahasa isyarat India dan melibatkan konversi gamabr ke *YcbCr*, segmentasi dengan operasi morfologi, penggunaan *Vision Transformer* dengan dua lapis transformer yang telah dilatih untuk mengenali dan memproses citra gambar. Model yang diusulkan berhasil mencapai akurasi pengujian sebesar 99,56% dan menunjukkan peningkatan performa. Selain penelitian yang dilakukan oleh (Agarwal, et al., 2023) ada pula penelitian yang dilakukan untuk meningkatkan pengenalan gestur tangan dengan model *Vision Transformer* (ViT) dengan subjek penelitian 3 dataset gestur tangan yaitu *American Sign Language* (ASL) *dataset*, ASL *with Digits dataset*, dan *National University of Singapore* (NUS) *hand gesture dataset*. Model diberi nama HGR-ViT dan mencapai akurasi yang sangat tinggi pada ketiga dataset yaitu: 99,98% untuk ASL, 99,36% untuk ASL *with Digits*, dan 99,85% untuk NUS *dataset*. Penelitian ini menggunakan *encoder transformer*

standar yang digunakan untuk mendapatkan representasi gestur tangan, dan kepala *preceptron multilayer* yang ditambahkan untuk klasifikasi (Tan, Chun Keat, Lim, Kian Ming, Chang, Roy Kwang Yang, Lee, Chin Poo, Alqahtani, Ali, 2023).

Meskipun tingkat akurasi yang tinggi telah dicapai oleh sebagian model yang diusulkan, namun dataset masih dalam bentuk huruf dan angka dalam sistem deteksi dan pengklasifikasian. Model yang dikembangkan oleh (Dwijayanti, Suci et al., 2021) belum mencakup pendeteksian kosakata masih dalam bentuk 26 huruf dan 10 angka BISINDO. Selain itu, penelitian yang dilakukan oleh (Ahmad et al., 2023) walaupun sudah tidak melakukan pendeteksian huruf dan angka namun penelitian baru menggunakan 5 kelas kata ganti orang , yaitu: “Saya”, “Kamu”, “Dia”, “Kami”, dan “Mereka”. Penelitian yang dilakukan (Agarwal, et al., 2023) sudah menggunakan *Vision Transformer,* namun untuk subjek penelitian adalah bahasa isyarat India (*Indian Sign Language*). Begitu pula penelitian yang dilakukan oleh (Tan et al., 2023) subjek penelitian menggunakan subjek ASL (*American Sign Language*) dengan nilai akurasi yang sudah tinggi.

1.2 Rumusan Masalah

Berdasarkan topik penelitian yang diajukan dan latar belakang yang telah dijelaskan, rumusan masalah dari penelitian ini adalah:

1. Bagaimana mensegmentasi atau memisahkan setiap bentuk gerakan tangan pada kosakata BISINDO dari informasi lain pada setiap *frame video*?
2. Bagaimana mendeteksi dan mengekstraksi *skeleton* gerakan tangan kosakata BISINDO?
3. Bagaimana mengenali setiap gerakan tangan kosakata BISINDO?
4. Bagaimana membangun sebuah perangkat aplikasi sebagai alat bantu yang dapat digunakan oleh penyandang tunarungu atau pengguna lainnya sehingga mempermudah dan mempercepat deteksi kosakata bahasa isyarat Indonesia (BISINDO)?

1.3 Tujuan Penelitian

Sehubungan dari rumusan masalah yang telah dipaparkan, tujuan dari penelitian ini sebagai berikut:

1. Menghasilkan algoritma segmentasi yang dapat menghasilkan gerak kosakata bahasa isyarat Indonesia (BISINDO) dari gerakan tangan seorang ahli bahasa isyarat dalam setiap citra/*frame video*.
2. Menghasilkan algoritma *skeletonisasi* bersifat *vision transformer* (ViT) yang dapat mengekstraksi *skeleton* dari gerak kosakata bahasa isyarat Indonesia (BISINDO)
3. Menghasilkan algoritma identifikasi kosakata bahasa isyarat Indonesia (BISINDO) berdasarkan pada ciri *vision tranformer* (ViT) menggunakan Jaringan Saraf Tiruan.
4. Tujuan akhir dari penelitian ini adalah menghasilkan sebuah *prototype* perangkat lunak untuk identifikasi kosakata bahasa isyarat Indonesia (BISINDO). *Prototype* ini merupakan hasil implementasi metode dan algoritma yang telah dikembangkan dalam penelitian disertasi.

1.4 Batasan Masalah

Penelitian deteksi bahasa isyarat Indonesia ini memiliki beberapa batasan masalah diantaranya:

1. Penelitian ini mendeteksi kosakata bahasa isyarat Indonesia untuk kata sehari-hari yang dikelompokkan menjadi enam kelas yaitu: “halo”, “apa kabar”, “maaf”, “terima kasih”, “tolong”, dan “sama-sama”.
2. Dataset yang digunakan adalah dataset yang bersifat primer karena data diambil sendiri oleh peneliti yang dibantu oleh tenaga ahli bahasa dari SLB B/C, Cempaka Putih, Jakarta Pusat.

1.5 Kontribusi dan Manfaat Penelitian

Hasil penelitian ini dapat memberikan kontribusi dalam bidang keilmuan berupa model dan algoritma akuisisi video identifikasi bahasa isyarat Indonesia (BISINDO) secara *real time*, algoritma skeletonisasi bersifat *vision transformer* (ViT), algoritma ekstraksi dan menghasilkan model untuk identifikasi gerakan tangan untuk kosakata BISINDO.

Dalam bidang teknologi, penelitian ini berkontribusi berupa prototype perangkat lunak sistem identifikasi gerakan tangan kosakata BISINDO dan terintegrasi sebagai perangkat TI bidang ilmu pengetahuan bahasa yang dapat difungsikan

untuk membantu para ahli bahasa dalam mengembangakan media pembelajaran serta dapat dijadikan sebagai media untuk memberikan informasi mengenai bahasa isyarat Indonesia (BISINDO). Penggunaan *vision transformer* (ViT) dan jaringan saraf tiruan berbasis model skeleton memungkinkan pendekatan yang lebih akurat dalam meidentifikasi gestur dan pola dalam bahasa isyarat, dibandingkan dengan metode tradisional. Integrasi ViT dalam pengenalan bahasa isyarat mendorong inovasi dalam pengolahan citra dan pemahaman visual, khususnya dalam konteks pengolahan data berbentuk non-tekstual seperti bahasa isyarat.

Dalam bidang sosial, memberikan aksesbilitas bagi komunitas tuli untuk mendapatkan manfaat dari komunikasi yang lebih lancar dan efektif dengan masyarakat luas, termasuk dalam layanan pubilk dan sosial. Selain itu, memberikan pendidikan dan pelatihan untuk pengembangan materi pendidikan dan pelatihan yang lebih baik untuk pembelajaran bahasa isyarat, tidak hanya untuk komunitas tuli tetapi juga untuk penerjemah bahasa isyarat dan masyarakat umum. Ada pula integrasi teknologi dalam layanan publik karena adanya aplikasi praktis dari penelitian yang mencakup pengembangan sistem otomatis. Terakhir dapat memberikan kesadaran dan inklusi sosial dalam kemajuan teknologi pengenalan bahasa isyarat terhadap kebutuhan dan keterampilan komunitas tuli.

BAB 2

TELAAH PUSTAKA

2.1 Citra Digital

Citra adalah representasi atau gambaran yang menyerupai sebuah objek. Sebagai hasil dari sistem pencatatan data optik, citra digital bisa berbentuk foto, atau analog dalam bentuk sinyal video yang tampak pada layar televisi, atau juga dalam format digital yang bisa langsung disimpan ke dalam perangkat penyimpanan (Sutoyo, T., Mulyanto, E., Suhartono, V., Nurhayanti, O.D., dan Wijananrto, 2009). Gonzalez dan Woods dalam buku mereka "Digital Image Processing" menjelaskan bahwa citra bisa dianggap sebagai proyeksi dari tiga dimensi ke dua dimensi, yang memberikan informasi visual dari scene yang diamati (Gonzalez & Woods, 2018). Citra digital dapat diperoleh dari berbagai sumber, seperti kamera digital, *scanner*, atau hasil simulasi komputer (E. Woods & C. Gonzalez, 2008). Sementara itu, citra analog memiliki sifat kontinu, seperti gambar di layar televisi, foto, lukisan, atau hasil *CT Scan* yang disimpan pada media penyimpanan. Citra analog tidak bisa langsung diproses oleh komputer dan memerlukan konversi terlebih dahulu. Sebaliknya, citra digital sudah siap untuk diproses langsung oleh komputer. Tujuan utama dari pengolahan citra digital adalah memproses dan memanipulasi citra digital untuk berbagai keperluan. Pengolahan citra ini melibatkan serangkaian operasi yang bertujuan untuk meningkatkan kualitas citra, mengekstraksi informasi penting, dan membuat citra lebih mudah dipahami atau digunakan dalam berbagai aplikasi (Svoboda, T., Kybic, J., & Hlavas, V., 2007).

2.1.1 Citra Biner (Monokrom)

Merupakan jenis citra yang memiliki nilai 0 dan nilai 1, dimana nilai 0 menyatakan warna hitam dan 1 menyatakan warna putih. Jenis citra biner banyak digunakan untuk pemrosesan citra khususnya untuk memperoleh tepi bentuk suatu objek. Pada citra biner dibutuhkan 1 bit memori untuk menyimpan kedua warna tersebut (Kadir, 2013, p.23).

2.1.2 Citra Keabuan (Grayscale)

(Kadir, 2013, p.23) menjelaskan bahwa sesuai dengan nama yang melekat, citra jenis ini menangani gradasi warna hitam dan putih, yang tentu saja menghasilkan efek warna abu-abu. Pada jenis gambar ini, warna dinyatakan dengan intensitas. Dalam hal ini, intensitas berkisar antara 0 sampai dengan 255. Nilai 0 menyatakan hitam dan nilai 255 menyatakan putih.

Agus (2013,p.309) menjelaskan bahwa Citra yang memiliki warna grayscale cenderung kurang menarik utuk dilihat dibandingkan dengan citra berwarna, karena kamera pada jaman dahulu hanya mampu menghasilkan citra dengan format warna greyscale, sehingga hasil citra tersebut menjadi kurang menarik untuk dilihat. Padahal, banyak citra zaman dahulu memiliki nilai sejarah yang cukup tinggi yang semestinya disampaikan dari generasi kegenerasi.

2.1.3 Citra Warna RGB (True Color)

Gonzales yang diterjemahkan oleh Arifin (2009,p.17) menguraikan bahwa system yang dipakai untuk mewakili warna yaitu sistem RGB (Red, Green,Blue). Sistem RGB adalah system penggabungan antara warna - warna primer (additiveprimary colours) yaitu merah (Red), hijau (Green) dan biru (Blue) untuk memperolehwarna tertentu. Misalnya warna putih diperoleh dari hasil gabungan warna merah =255, hijau = 255, dan biru = 255. Dalam system RGB, warna putih cerah dinyatakan dengan RGB (255, 255, 255). Range nilai dari setiap warna primer adalah 0 sampai 255. Sehingga kemungkinan warna yang dapat terbentuk dengan sistem RGB adalah 256 x 256 x 256 yakni kurang lebih 16.7 juta warna.

2.2 Segmentasi Citra

Segmentasi citra merupakan proses yang ditujukan untuk mendapatkan objek-objek yang terkandung di dalam citra atau membagi citra ke dalam bentuk daerah dengan setiap objek atau daerah memiliki kemiripan atribut. Segmentasi citra bertujuan untuk membagi wilayah-wilayah yang homogen, pada citra yang hanya mengandung sebuah objek, maka objek dibedakan dari latar belakang (*background*). Pembagian dalam proses segmentasi bergantung pada masalah yang akan diselesaikan. Pada proses segmentasi merupakan tahapan atau metode yang sangat penting digunakan untuk mengubah citra *input* ke dalam citra *output* berdasarkan atribut yang diambil dari citra tersebut. Proses segmentasi harus dihentikan apabila masing-masing objek telah terisolasi atau terlihat dengan jelas. Tingkat keakurasian segmentasi tergantung pada tingkat keberhasilan prosedur analisis yang dilakukan, dan diharapkan proses segmentasi memiliki tingkat akurasi yang tinggi. (Sutoyo, T., et al., 2009). Algoritma dalam proses segmentasi dibagi menjadi 2 macam, yaitu:

1) Diskontinuitas, merupakan pembagian citra berdasarkan perbedaan dalam intensitasnya, contohnya: titik, garis dan *edge* (tepi).
2) Similaritas, pembagian citra berdasarkan kesamaan-kesamaan kriteria yang dimilikinya, misalnya: *thresholding*, *region growing*, *region splitting*, dan *region merging*.

2.3 *Cropping*

Cropping image atau pemotongan citra bertujuan untuk membuat area of interest, untuk mempertegas fenomena geospasial dan pembahasan pada daerah kajian. Hal ini dilakukan untuk menghindari adanya analisis di luar daerah kajian. Selain itu, hal ini dilakukan untuk lebih memudahkan perencana melakukan analisis citra dari daerah kajian (Rina 2011). Pemotongan juga mengakibatkan ukuran obyek menjadi lebih besar, sehingga konten yang ada (informasi berupa warna) terlihat lebih jelas. Cropping citra merupakan salah satu langkah yang dilakukan setelah koreksi geometrik dan koreksi radiometrik.

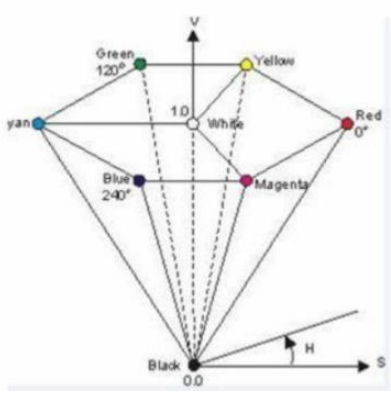
2.4 *Skeleton*

Skeleton dalam konteks ekstraksi fitur adalah sebuah representasi yang disederhanakan dari bentuk objek yang menunjukkan struktur dasar atau garis tengah (midline) dari bentuk tersebut. Proses untuk mendapatkan skeleton ini sering disebut sebagai skeletonization, yang bertujuan untuk mengurangi objek menjadi bentuk paling sederhana tanpa kehilangan informasi struktural utama. Metode ini sangat berguna dalam analisis bentuk, pengenalan pola, dan pemrosesan citra. (Gonzalez, Rafael C., and Richard E. Woods. "Digital Image Processing." Pearson Education, Inc., 2008).

Skeletonization membantu dalam mengidentifikasi dan menganalisis karakteristik bentuk dengan cara yang lebih efisien karena mengurangi jumlah data yang harus diproses. Ini sering digunakan dalam bidang seperti pengolahan citra medis, robotika, dan pengenalan tulisan tangan untuk mendeteksi struktur dasar dan fitur penting dari bentuk atau gambar. (Sonka, Milan, Vaclav Hlavac, and Roger Boyle. "Image Processing, Analysis, and Machine Vision." Thomson, 2008).

2.5 HSV (*Hue, Saturation, Value*)

HSV (*Hue Saturation Value*) merupakan salah satu ruang warna yang digunakan manusia dalam memilih warna cat atau tinta. Sistem ini dipandang lebih dekat dibandingakan dengan RGB dalam mendeskripsikan sensasi warna oleh mata manusia. Dalam terminologi para seniman HSV berkaitan dengan tint, shade, dan tone (Awaludin, Muryan, 2016).



Gambar 2. 1 Representasi Ruang Warna HSV

(Awaludin, Muryan, 2016).

Dari Gambar 2.1 perhatikan apabila R, G, dan B bernilai sama, maka warna menjadi keabuan yang membentuk intensitas putih. Warna tersebut hanya warna putih, akan memiliki nilai saturation nol. Sebaliknya, kika nilai-nilai RGB berbeda, maka warna yang dihasilkan nilai saturation yang tinggi. Dapat kita amati bahwa jika salah satu dari nilai-nilai RGB bernilai nol, maka saturation bernilai 1.

2.6 Vision Transformer (ViT)

Vision Transformer (ViT) adalah model berbasis transformator yang awalnya dikembangkan untuk tugas-tugas pemrosesan bahasa alami, tetapi kemudian diadaptasi untuk analisis gambar. Dalam konteks ekstraksi fitur, ViT memperlakukan gambar sebagai sekumpulan patch dan menggunakan mekanisme self-attention untuk menangkap hubungan antar-patch dalam gambar, yang sangat bermanfaat untuk memahami konteks visual yang kompleks. Secara lebih spesifik, Vision Transformer memulai dengan membagi gambar menjadi patch kecil, seringkali berukuran 16x16 atau 32x32 piksel (Dosvitskiy, A., et al., 2021). Setiap patch ini kemudian di-flatten dan diproses melalui serangkaian blok transformator, yang setiap bloknya terdiri dari mekanisme self-attention dan feed-forward neural network. Output dari blok ini digunakan sebagai fitur tingkat tinggi yang mencerminkan informasi visual yang penting dari gambar (Khan, S., et al., 2021).

Keunggulan utama dari ViT dalam ekstraksi fitur adalah kemampuannya untuk mengelola hubungan spasial jarak jauh antar patch dalam gambar. Ini memungkinkan ViT untuk mengidentifikasi pola dan objek dengan lebih efektif dibandingkan dengan arsitektur CNN tradisional yang lebih fokus pada informasi lokal melalui operasi konvolusi (Han, K., et al., 2022). Vision Transformer telah menunjukkan kinerja yang sangat baik dalam berbagai tugas

pengenalan gambar dan telah menjadi pilihan populer untuk banyak aplikasi di bidang visi komputer. Hal ini sebagian besar karena skalabilitasnya, kemampuan untuk melatih dengan dataset besar, dan efisiensi dalam memanfaatkan informasi yang ada pada gambar besar atau kompleks (Touvron, H., et al., 2021).

2.7 Kecerdasan Artifisial

Kecerdasan artifisial adalah teknologi komputer yang memungkinkan untuk melakukan tugas-tugas manusia yang membutuhkan intelegensi (Healey, 2020). Kedalaman intelegensi sampai saat ini masih perlu dikembangkan agar memiliki kemampuan untuk memproses pemahaman yang mendalam. Intelegensi pada kecerdasan artifisial yang sangat handal adalah menyimpan data atau "mengingat pengetahuan" dalam jumlah yang besar dan kemampuan untuk mentransfer kemanapun dan kapanpun dengan kecepatan yang tinggi. Kehandalan lainnya adalah pada hal-hal yang terkait pada nalar kepakaran berbasis matematika dan statistika dalam berapapun jumlah data yang diberikan kepadanya. Namun, kecerdasan manusia yang sulit diraih oleh kecerdasan artifisial adalah adaptasi (Ertel, 2018) dan kreatifitas (Simplilearn, 2020). Kecerdasan artifisial difungsikan sebagai perangkat yang membantu kegiatan manusia dalam segala aktifitasnya. Kecerdasan ini bukan dijadikan sebagai pengganti peranan manusia secara absolut karena kecerdasan artifisial sebagai pemegang peranan sentral yang dikelola atau tetap membutuhkan kendali manusia dan hanya berperan sebagai pendukung aktifitas manusia dalam perkara tertentu (Zebua, Rony Sandra Yofa, Khairunnisa, Hartatik, Pariyadi, Wahyuningtyas, Dessy Putri, Thantawi, Ahmad M., Sudipa, I Gede Iwan, Sumakul, Grace Christien, Sepriano, Kharisma, Lalu Puji Indra, 2023).

Eksplorasi aktif terkait kecerdasan artifisial sudah dimulai sejak tahun 1940an dan 1950an yaitu ketika terdapat ilmuwan yang melakukan eksplorasi terhadap kapabilitas komputer untuk memberikan solusi terhadap pekerjaan mereka (Simplilearn, 2020). John McCarthy mengusulkan kecerdasan artifisial pada saat seminar musim panas di Dartmouth pada tahun 1956 (Gheorge Tecuci, 2012). Tujuan dari kecerdasan artifisial adalah untuk menciptakan perangkat lunak dan perangkat keras komputer yang memiliki kemampuan berpikir seperti manusia (Stephen Lucci, Danny Kopec, 2016). Kecerdasan artifisial mencakup pengembangan dan implementasi algoritma untuk memproses data, belajar, dan menganalisis, selain itu dilibatkan pula berbagai aspek termasuk statistik dan pembelajaran mesin, pengenalan pola, pengelompokan, metode berbasis kesamaan, logika, dan teori probabilitas. Selain itu, pendekatan yang terinspirasi dari aspek fisiologis seperti jaringan saraf dan pemodelan *fuzzy* juga merupakan bagian dari bidang kecerdasan artifisial (Rahman & Saputra, 2023).

2.8 Pembelajaran Mesin

Pendekatan pembelajaran mesin merupakan pendekatan yang digunakan untuk membangun sebuah sistem kecerdasan artifisial yang mampu untuk mempelajari pola dan hubungan dalam data, dan menggunakan pola tersebut untuk membuat prediksi (Arankalle et al., 2020; Darapureddy et al., 2021; Pulipaka, 2021). Proses pembelajaran dilakukan dari contoh dan pengalaman, dimana contoh dan pengalaman tersebut tanpa diprogram secara eksplisit. Contoh yang menggunakan pendekatan ini adalah sistem rekomendasi, pengenalan wajah, pengenalan suara, deteksi objek, pengenalan karakter tulisan tangan, kendaraan otonom, prediksi harga saham, deteksi penipuan, diagnosis penyakit.

Pembelajaran mesin dan kecerdasan artifisial adalah dua konsep yang saling terkait, tetapi memiliki perbedaan yang penting. Baik pembelajaran mesin maupun kecerdasan artifisial bertujuan untuk mengembangkan sistem yang mampu melakukan tugas yang biasanya membutuhkan kecerdasan manusia dan memanfaatkan data sebagai sumber informasi untuk mempelajari pola dan membuat keputusan. Tujuan pembelajaran mesin yaitu memprediksi masa depan (*Unobserved Event*) atau memperoleh ilmu pengetahuan (Knowledge Discovery / Discovering Unknown Structure) (Liu et al., 2018).

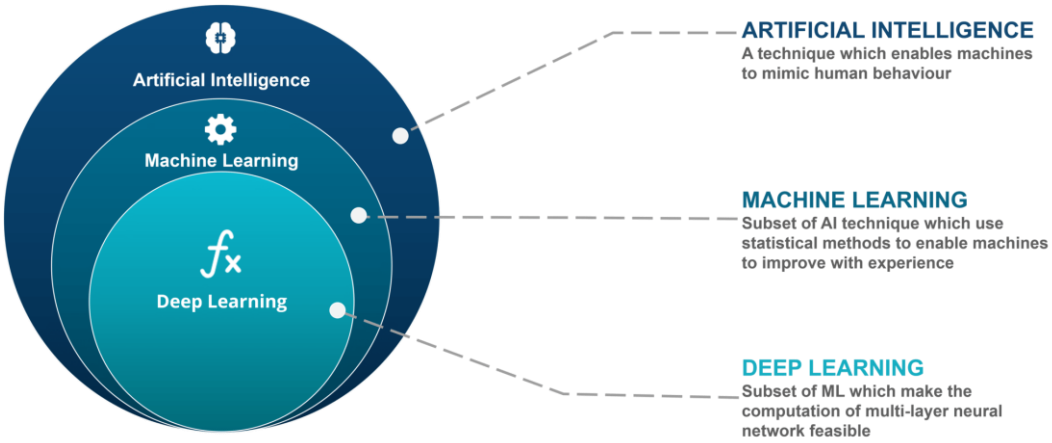
Gambar 2. 2 Hubungan Pembelajaran Mesin
(irishtechnews.ie, 2023)

Inti dari pembelajaran mesin adalah bagaimana membuat komputer dapat menyelesaikan berbagai persoalan dan dapat belajar sendiri seperti manusia belajar sesuatu. Jika digambarkan secara diagram, wilayah kajian kecerdasan artifisial akan jauh lebih besar dibandingkan pembelajaran mesin. Dapat disimpulkan bahwa pembelajaran mesin adalah bagian dari kecerdasan artifisial. Semua hal yang terkait dengan pembelajaran mesin praktis akan terkait juga dengan kecerdasan artifisial. Secara umum algoritma pembelajaran mesin dapat dikelompokan menjadi 3 kategori, yaitu *Supervised Learning* (prediksi dengan menggunakan

bantuan training dataset), *Unsupervised Learning* (untuk menemukan implicit relationship dan unlabeled dataset) dan *Reinforcement Learning* (mempelajari suatu policy kemudian komputer melakukan self-discovery) (Puspha Annabel et al., 2019).

2.9 Pembelajaran Mendalam

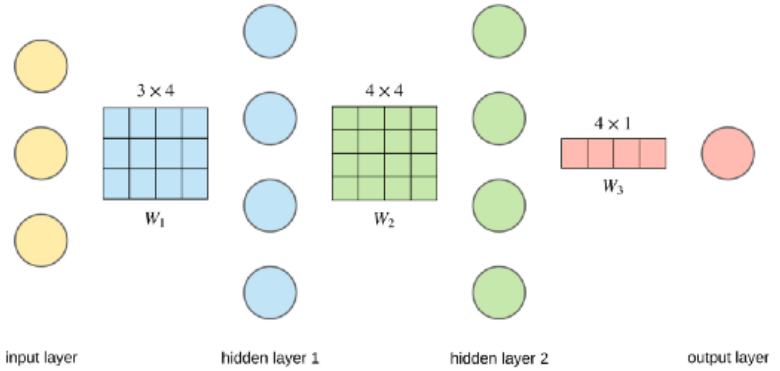
Sejarah Deep Learning dimulai pada tahun 2006, yaitu setelah Geoffrey Hinton mempublikasikan paper yang memperkenalkan salah satu varian neural network yang disebut deep belief nets. Paper ini merupakan awal kemunculan istilah Deep Learning, untuk membedakan arsitektur neural network konvensional (single layer) dengan arsitektur neural network multi layer atau banyak layer. Deep Learning adalah salah satu cabang machine learning yang menggunakan Deep Neural Network untuk menyelesaikan permasalahan pada domain machine learning. Sayangnya ide Deep Learning sangat kompleks, sehingga membutuhkan komputer dengan spesifikasi tinggi yang belum dapat dipenuhi saat ini (Roy et al., 2020).



Gambar 2. 3 Arsitektur AI, ML dan Deep Learning

(Prajwal Shrestha, 2021)

Pada Gambar 2.3 adalah gambar arsitektur dari *Artificial Intelligence*, *Machine Learning*, dan *Deep Learning*. Pada tahun 2009, Andrew memperkenalkan penggunaan GPU untuk *Deep Learning* melalui *paper* yang berjudul *Large-scale Deep Unsupervised Learning using Graphics Processors*. Dengan menggunakan GPU, algoritma *deep learning* dapat dijalankan lebih cepat dibandingkan dengan tanpa GPU (hanya menggunakan CPU). Perkembangan *deep learning* maju pesat berkat keberadaan *hardware* yang memadai. Dan saat ini, *Deep Learning* sudah banyak diaplikasikan di berbagai bidang area, seperti pengenalan wajah, *self-driving car*, pengenalan suara, dan lain sebagainya (Prajwal Shrestha, 2021).



Gambar 2. 4 Diagram Network Model Deep Learning

Bentuk diagram *network* model *deep learning* dapat dilihat seperti Gambar 2.4 diatas. Perhatikan bahwa *hidden layer* hanya digambarkan dua lapis saja. Padahal kenyataannya bisa berjumlah sangat banyak. *Deep Learning* sudah dikembangkan ke berbagai model atau arsitektur yang berbeda-beda.

2.10 Bahasa Isyarat Indonesia (BISINDO)

Bahasa isyarat adalah bahasa yang universal karena bahasa isyarat dipakai oleh penyandang tuna rungu ketika mereka berbicara atau melakukan komunikasi dengan orang normal atau dengan sesamanya. Menurut Wedayanti (2019: 144) BISINDO merupakan isyarat alamiah yang diciptakan dan digunakan oleh tuna rungu sesuai dengan persepsi mereka terhadap segala sesuatu disekitar mereka, bukan bahasa isyarat rumahan (*home sign*) atau gestur.

Melalui bahasa isyarat tuna rungu dapat berkomunikasi dengan lingkungan sekitarnya. Menurut Yuwono Imam, Dewi Ratih R, Evian Damastuti (2020: 15) BISINDO adalah bahasa isyarat yang berpedoman pada ekspresi, gerakan, posisi tubuh, kontak mata yang dikembangkan oleh individu tuna rungu. Karakteristik BISINDO menurut Wedayanti Ni Putu Luhur (2019: 144) menyatakan bahwa karakteristik BISINDO selain memiliki isyarat ikonis, ketika berisyarat diikuti berbagai ekspresi wajah maupun mulut untuk melengkapi makna dari isyarat atau hal yang ingin diutarakan

2.11 Kajian Penelitian

Beberapa penelitian sebelumnya telah melakukan sejumlah penelitian yang berkaitan dengan deteksi dan klasifikasi bahasa isyarat. Berikut adalah ringkasan dari penelitian tersebut yang berkaitan dengan deteksi dan klasifikasi bahasa isyarat yang dijelaskan sebagai berikut.

2.11.1 Tinjauan 1

(Ojha, Ankit, Pandey, Ayush, Maurya, Shubham, Thakur, Abhishek, P., Dayananda, 2020) *Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Netrwork*.

Penelitian ini merupakan demonstrasi sederhana tentang bagaimana CNN bisa dikembangkan untuk pemecahan masalah dengan tingkat akurasi yang tinggi (Ojha, Ankit et al., 2020). Penelitian ini menghasilkan aplikasi desktop dengan menggunakan *webcam* komputer dengan model CNN yang digunakan untuk mendeteksi gestur tangan agar dapat menerjemahkan ASL (*American Sign Language*) ke dalam teks yang diubah menjadi audio dengan secara *real time* menggunakan *library pyttsx3*. Tahapan yang dilakukan dalam penelitian ini dimulai dengan akuisisi citra gestur tangan yang diambil melalui kamera *web* dengan menggunakan *video sreeam OpenCV* dengan dimensi 50x50 piksel. Setelah itu, dilakukan segmentasi dan deteksi bagian tangan untuk mendapatkan prediksi. Selanjutnya, dilakukan pengenalan postur tangan dimana citra yang telah diproses sebelumnya dimasukkan ke model keras CNN. Model dilatih lalu menghasilkan label prediksi. Semua label gestur diberi probabilitas dan label dengan probabilitas tertinggi dianggap sebagai label prediksi. Setelah itu, dihasilkan tampilan dalam bentuk teks dan audio dimana model mengakumulasi gestur yang dikenali menjadi kata-kata lalu dikonversi menggunakan *library pyttsx3*. Selanjutnya, dilakukan pelatihan model CNN, dimana CNN dilatih dengan dataset citra tangan ASL sebanyak 144 kelas/gestur. Model CNN dalam penelitian ini terdiri dari 11 *layer* dimana terdapat 3 *layer convolutional*. Nilai akurasi yang diperoleh dari penelitian ini sebesar 95% sehingga dapat dikatakan bahwa penelitian ini mampu memecahkan bagian dari masalah terjemahan Bahasa Isyarat.

2.11.2 Tinjauan 2

(Kembuan, Olivia, Rorimpandey, Gladly Caren, Tengker, Soenandar Milian Tompunu, 2020) *Convolutional Neural Network (CNN) for Image Classification of Indonesia Sign Language Using Tensorflow*.

Penelitian ini mengusulkan penggunaan arsitektur *Convolutional Neural Network* (CNN) dan *Library Tensorflow* untuk membangun model klasifikasi sistem pengenalan gambar Bahasa Isyarat Indonesia (BISINDO) berdasarkan pada gambar statis dengan nilai akurasi yang tinggi (Kembuan, Olivia et al., 2020). Tahapan dari penelitian ini dimulai dengan pengumpulan data, dimana data diperoleh secara sekunder melalui Kaggle dengan ukuran data 2,83 GB dalam gambar RGB standar. Pada data yang diperoleh digunakan 80% (2.113 data) sebagai dataset pelatihan, dan 20% (546 data) sebagai dataset validasi. Dataset yang digunakan adalah Bahasa

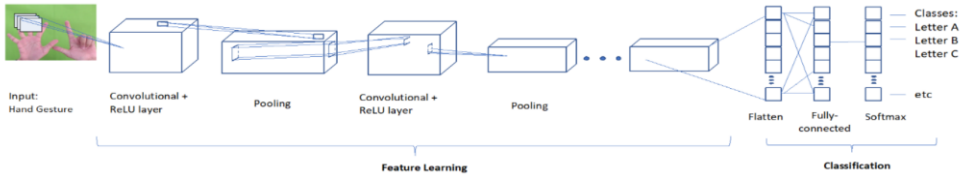
Isyarat Indonesia (BISINDO) yang berisi 2.659 gambar dari 26 huruf abjad. Tahapan selanjutnya adalah persiapan data dengan melakukan ekstraksi gambar-gambar dari dataset dan membuat folder pelatihan dan validasi untuk menyimpan gambar-gambar secara terpisah, mengubah skala nilai tensor gambar menjadi 0-1, dan mengubah ukuran gambar menjadi 150x150 piksel. Setelah itu, dilakukan tahapan pembuatan dan pelatihan model dengan melakukan perancangan arsitektur model CNN dengan 4 blok konvolusi dan *max pooling*, pelatihan model selama 5 *epoch*, dengan *batch size* 10. Tahapan yang dilakukan selanjutnya adalah evaluasi model dengan melakukan perhitungan akurasi dan *loss* untuk dataset pelatihan dan validasi, serta perhitungan akurasi klasifikasi gambar uji. Hasil dari penelitian ini adanya sebuah sistem *Convolutional Neural Network* untuk klasifikasi gambar Bahasa Isyarat Indonesia (BISINDO) menggunakan *Tensorflow* yang berhasil diimplementasikan dengan nilai akurasi untuk dataset pelatihan sebesar 96,67%, nilai akurasi untuk dataset validasi sebesar 100% dan nilai akurasi klasifikasi gambar uji untuk setiap huruf BISINDO yang diuji juga mencapai 100%, sehingga dapat disimpulkan penggunaan CNN dan *Tensorflow* efektif untuk pengenalan BISINDO.

2.11.3 Tinjauan 3

(Dwijayanti, Suci, Hermawati, Taqiyyah, Sahirah Inas, Hikmarika, Hera, Suprapto, Bhakti Yudho, 2021) *Indonesia Sign Language Recognition using Convolutional Neural Network*.

Penelitian ini mengusulkan pendekatan *deep learning* yaitu dengan membuat model CNN baru yang diberi nama model C (Dwijayanti, Suci et al., 2021) untuk mengenali BISINDO yang terdiri dari 26 huruf dan 10 angka. Penelitian ini memiliki tujuan untuk membandingkan kinerja pengenalan BISINDO dari model CNN yang disederhanakan dengan AlexNet dan VGG-16. Tahapan penelitian dimulai dari pengumpulan *image dataset* yang diperoleh dari 10 responden dalam dua kondisi pencahayaan yaitu kondisi redup dan terang. Setiap responden melakukan 37 gerakan tangan yang terdiri dari 26 huruf dan 11 angka (0-10). Data direkam dalam format video (.mp4) selanjutnya data yang sudah diperoleh diubah menjadi gambar dengan format .jpg. Selanjutnya adalah tahapan *data pre-processing* yang dilakukan dengan mengubah ukuran gambar dan menskalakan fitur menjadi 60x60 piksel dan *scaling* nilai piksel menjadi rentang 0-1. Setelah itu, tahapan *data split* dengan menggunakan data yang telah diperoleh sebelumnya sebanyak 39.455 data dan kemudian dibagi menjadi tiga bagian yaitu: data latih sebanyak 60%, data validasi sebanyak 20%, dan data uji sebanyak 20%. Tahapan selanjutnya adalah *training process*, arsitektur model CNN dibagi menjadi 3 bagian yaitu model A yang merupakan versi

modifikasi dari AlexNet, model B yang merupakan versi modifikasi dari VGG-16, dan model C yang merupakan arsitektur baru yang diusulkan dalam penelitian ini. Model CNN yang diusulkan terdiri dari 4 lapisan *convolutional*, 3 lapisan *pooling*, dan 3 lapisan *fully-connected*. Model diuji menggunakan data uji dan dievaluasi performanya berdasarkan *accuracy*, *precision*, *recall*, dan *F1-Score*. Model arsitektur CNN yang digunakan untuk mengolah citra terdapat pada gambar 2.5.



Gambar 2. 5 Arsitektur CNN

(Dwijayanti, Suci et al., 2021)

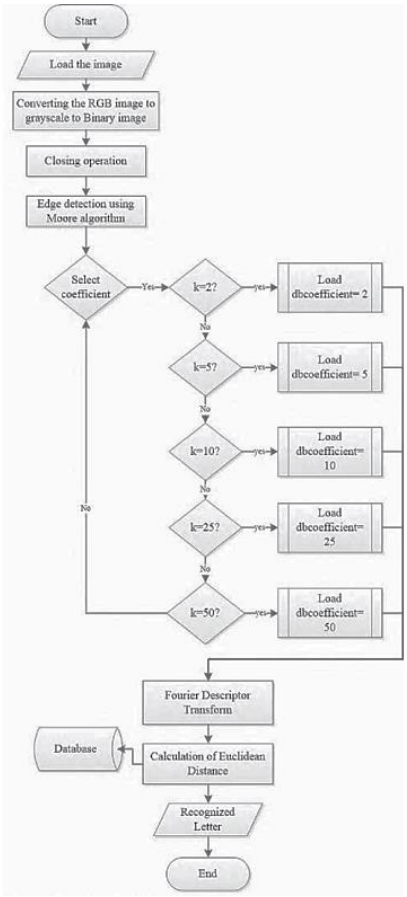
Hasil pengujian untuk pengujian kondisi pencahayaan terang untuk model A mendapat nilai akurasi 0,985, lalu untuk model B mendapat nilai akurasi 0,038 dan model C mendapat nilai akurasi 0,979. Sementara untuk kondisi pencahayaan redup untuk model A mendapat nilai akurasi 0,987, model B mendapat nilai akurasi 0,038, dan model C mendapat nilai akurasi 0,987. Hasil pengujian untuk perspektif orang pertama untuk model A mendapat nilai akurasi sebesar 0,984, lalu untuk model B mendapat nilai akurasi sebesar 0,031, dan model C mendapat nilai akurasi sebesar 0,978. Sementara dari perspektif orang kedua untuk model A mendapat nilai akurasi sebesar 0,987, untuk model B mendapat nilai akurasi 0,043, dan model C 0,987. Model CNN C yang diusulkan bekerja dengan baik dalam memprediksi gerakan tangan dengan nilai akurasi 98,3%.

2.11.4 Tinjauan 4

(Basri, Syartina Elfarika, Indra, Dolly, Darwis, Herdianti, Mufila, A. Widya, Ilmawan, Lutfi Budi, Purwanto, Bobby, 2021) *Recognition of Indonesian Sign Language Alphabets Using Fourier Descriptor Method*.

Penelitian ini menggunakan metode *Fourier Descriptor* yang digunakan untuk mengekstraksi fitur citra Bahasa Isyarat Indonesia (BISINDO) untuk pengenalan huruf abjad (Basri, Syartina Elfarika et al., 2021). Penelitian ini menerapkan empat langkah utama, yang dimulai dengan proses pra-pemrosesan dengan mengonversi citra RGB menjadi *grayscale* dan *binary*, serta melakukan operasi *closing* untuk pemulusan citra. Tahapan selanjutnya adalah tahap deteksi kontur menggunakan *Moore's Algorithm*, lalu setelahnya adalah tahapan ekstraksi fitur

Fourier Descriptor dengan menggunakan 5 koefisien yaitu 2, 5, 10, 25, dan 50 untuk mewakili fitur dari citra. Terakhir adalah proses pengenalan fitur dengan menghitung kemiripan citra menggunakan *Euclidean Distance*. Citra yang digunakan dalam penelitian ini sebanyak 1.820 citra yang terbagi menjadi citra *standard*, citra *scale*, citra *rotation* dan citra *translation* yang diuji dengan 130 citra data pelatihan. Proses pengenalan huruf BISINDO yang dipakai pada penelitian ini dapat dilihat pada Gambar 2.6.



Gambar 2. 6 Proses Pengenalan Huruf BISINDO

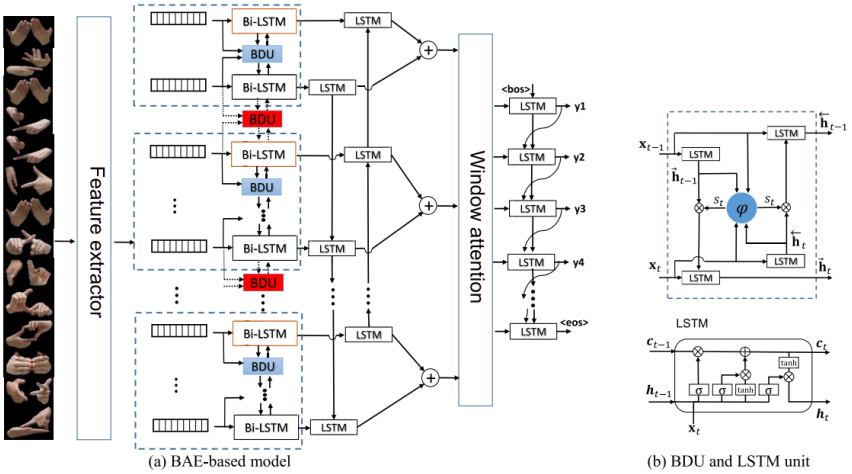
(Basri, Syartina Elfarika et al., 2021)

Berdasarkan hasil pengujian, *Fourier Descriptor* dapat digunakan untuk mengekstraksi citra gambar huruf BISINDO dan semakin tinggi koefisiennya maka semakin akurat hasil pengenalannya. Hal ini dibuktikan dengan nilai akurasi terbaik diperoleh pada koefiesien 25 dan 50 dengan persamaan akurasi 96,92%. Sementara itu, hasil kombinasi dari *Fourier Descriptor* dan *Euclidean Distance* masih dinilai cukup untuk mengenali citra *standard* dengan nilai akurasi 74,15% dan citra *scale* dengan nilai akurasi 72,30%, sedangkan untuk citra *rotation* mendapat nilai akurasi 57,43% dan citra *translation* mendapatkan nilai akurasi terendah sebesar 34,36%.

2.11.5 Tinjauan 5

(Huang, Shiliang, Ye, Zhongfu, 2021) *Boundary Adaptive Encoder With Attention Method for Chinese Sign Language Recognition*.

Penelitian ini mengembangkan metode *Sign Language Recognition* (SLR) dengan mengusulkan metode untuk pengenalan bahasa isyarat Tiongkok berbasis *Boundary Adaptive Encoder* (BAE) yang menggabungkan *window attention* model untuk meningkatkan efisiensi dan mampu mengenali baik kata terisolasi maupun kalimat kontinu (Huang, Shiliang, Ye, Zhongfu, 2021). Penelitian ini mengusulkan *Boundary Adaptive Encoder* (BAE) hirarkis dengan dua lapisan *Bidirectional LSTM* (*Long Short Term Memory*) yang dapat mempelajari informasi batas temporal. Kerangka keseluruhan untuk pemodelan yang diusulkan pada penelitian ini dapat dilihat pada Gambar 2.7.



Gambar 2. 7 Kerangka Keseluruhan Model SLR yang diusulkan (a), dan (b) menunjukkan hubungan BDU (*Boundary Detection Unit*) dan LSTM dua arah dalam kotak bertitik biru. BDU Merah mewakili sinyal deteksi batas $s_t = 0$

(Huang, Shiliang, Ye, Zhongfu, 2021)

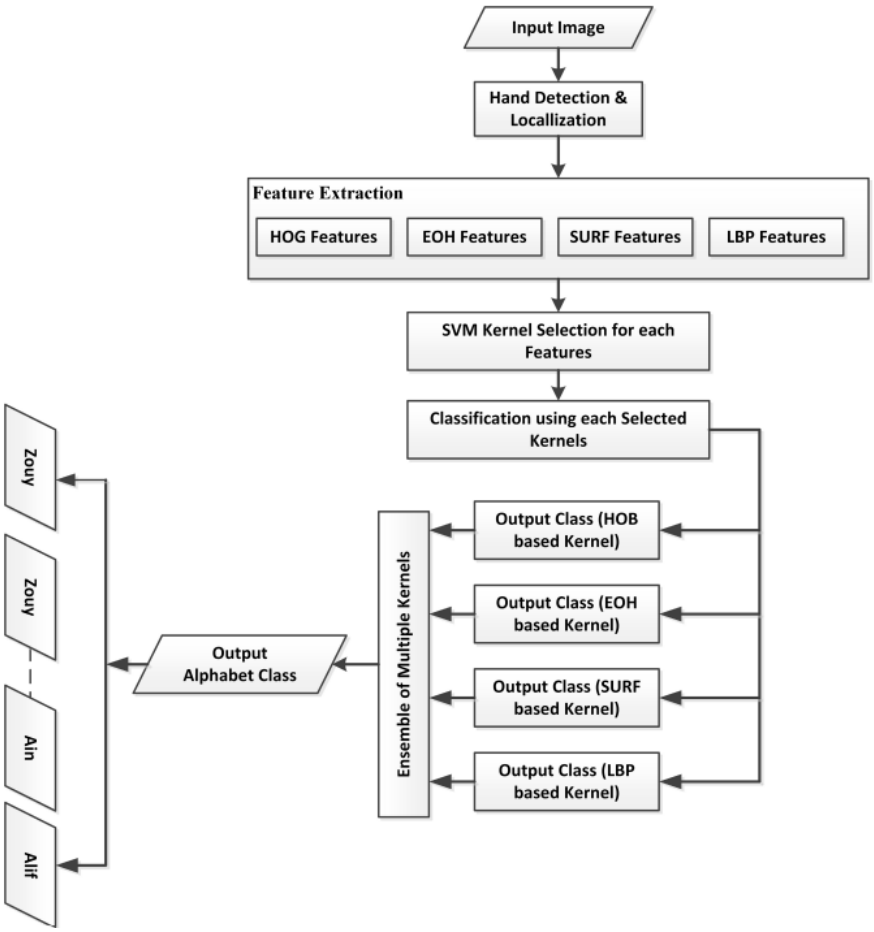
Dalam penelitian ini menggunakan BDU (*Boundary Detection Unit*) yang merupakan bagian integral dari proses BAE (*Boundary Adaptive Encoder*) yang memiliki dua layer *Bidirectional LSTM*. Pada layer encoder pertama, diantara setiap Bi-LSTM terdapat BDU yang bertugas mempelajari batas-batas waktu (*boundary*) dari sinyal inputan dan mendeteksi apakah terjadi perubahan besar pada sinyal pada titik waktu tertentu. Apabila terdeteksi perubahan besar, BDU akan melakukan *reset state LSTM* agar sinyal dapat dibagi menjadi beberapa bagian *boundary*. *Output* dari *layer encoder* pertama yang telah dibagi per bagian ini akan menjadi *input* pada layer kedua. Tahapan proses yang dilakukan dalam penelitian ini dimulai dengan dilakukannya pengumpulan data berupa *video* dengan berbagai kata dan kalimat yang terbagi menjadi 2 kelompok yaitu: *isolated dataset* (ID1, ID2) dan *continuous dataset* (CD). Setelah

dikumpulkan, tahap selanjutnya adalah ekstraksi fitur dengan menggunakan 2D CNN dan 3D CNN untuk ekstraksi fitur *spasial temporal* dari *video* dan fitur-fitur ini akan dijadikan *input* model. Tahap selanjutnya adalah tahap pelabelan sub-kata bahasa isyarat dengan membagi menjadi unit sub-kata yang lebih kecil berdasarkan makna. Sub-kata ini digunakan sebagai unit dasar pengenalan. Setelah palabelan, tahapan selanjutnya adalah pelatihan model. Pada tahap pelatihan model dibangun model *encoder-decoder* hierarkis menggunakan BAE dan Bi-LSTM. Tahapan selanjutnya adalah tahap uji coba dengan melakukan eksperimen pengenalan kata dan kalimat dan hasilnya dibandingkan dengan metode lain seperti s2vt, HRNE, HRF-S dan lainnya. Penelitian ini tidak secara spesifik menyebutkan nilai akurasi yang diperoleh, namun berdasarkan tabel hasil yang ditampilkan dapat ditarik kesimpulan bahwasannya pada dataset ID1 akurasi terbaik adalah 96,1% menggunakan metode yang diusulkan dengan fitur 3D CNN dan dataset ID2-split1 akurasi terbaik 94,6% yang sama, sementara dataset ID2-split2 akurasi terbaik adalah 91,7% dengan metode yang sama, sementara untuk dataset CD metode yang diusulkan mencapai *word error rate* terbaik sebesar 15,1%. Secara keseluruhan, metode yang diusulkan dalam penelitian ini berhasil mencapai peningkatan akurasi yang signifikan dibandingkan metode sebelumnya untuk pengenalan bahasa isyarat Tiongkok baik kata maupun kalimat.

2.11.6 Tinjauan 6

(Shah, Farman, Shah, Muhammad Saqlain, Akram, Waseem, Manzoor, Awais, Mahmoud, Rasha Orban, Abdelminaam, Diaa Salama, 2021) *Sign Language Recognition Using Multiple Kernel Learning*: *A Case Study of Pakistan Sign Language*.

Penelitian ini mengusulkan teknik untuk pengenalan 36 huruf statis dari *Pakistan Sign Language* (PSL) menggunakan pembelajaran *Multiple Kernel Learning* pada (SVM) *Support Vector Machine* (Shah, Farman et al., 2021). Penelitian ini menganalisis dan mengekstraksi empat jenis fitur dari *grayscale images* yang disertakan *histogram of oriented gradients* (HOG), *edge orientation histogram* (EOH), *local binary patterns* (LBP), dan *speeded up robust features* (SURF). Fitur-fitur diekstraksi secara terpisah kemudian diklasifikasikan menggunakan *multiple kernel learning* pada SVM. Pengklasifikasian menggunakan 3 fungsi kernel (*Gaussian*, *Linier*, dan *Polynomial*). *Kernel* dengan akurasi tertinggi dipilih sebagai *kernel* optimal untuk fitur tersebut dan hasil klasifikasi dari ke-4 fitur di-*ensemble* untuk mendapatkan kelas akhir. Dataset PSL terdiri dari total 6.633 citra statis alfabet PSL. Pembagian kelas data terdiri dari 70% untuk data pelatihan, 15% untuk validasi dan 15% lainnya untuk pengujian data. Alur pengerjaan untuk penelitian ini dapat dilihat pada Gambar 2.8.



Gambar 2. 8 Alur Pengerjaan Yang Diusulkan

(Shah, Farman et al., 2021)

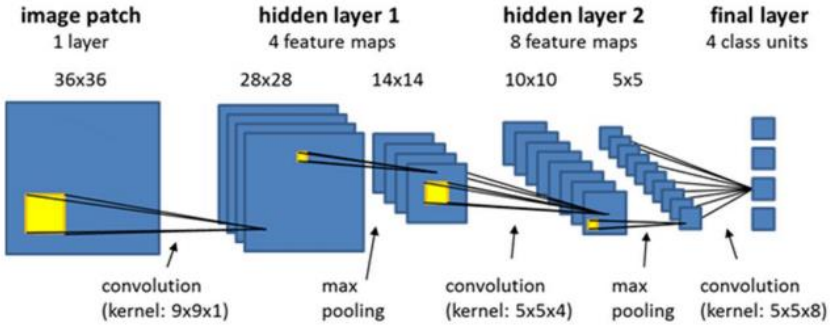
Pemodelan untuk arsitektur yang digunakan pada penelitian ini dimulai dari pemrosesan citra, ekstraksi fitur, klasifikasi SVM dengan menggunakan *multiple kernel learning*, dan terakhir *ensemble* hasil klasifikasi. Penelitian ini melakukan evaluasi dengan menggunakan akurasi, presisi, *recall* dan *f-score*. Secara keseluruhan, metode yang diusulkan memperoleh akurasi cukup tinggi sebesar 91,93% untuk pengenalan 36 huruf citra statis dari PSL, sementara untuk presisi 89,2%, selanjutnya untuk *recall* mendapatkan nilai sebesar 90,1%, dan *f-score* berada di rata-rata 89,6%.

2.11.7 Tinjauan 7

(Kharat, Aditya, Patil, Yash, Jagtap, Omkar, Sonawale, Rajashri, 2022) *Sign Language to Text Conversion*.

Penelitian ini mengembangkan metode *real time* dengan menerapkan *convolutional neural network* (CNN) untuk *American Sign Language* (ASL) 26 huruf alfabet (Kharat, Aditya et.al., 2022). Penelitian ini menggunakan pendekatan berbasis penglihatan (*vision based approach*) dan mengumpulkan dataset dengan menangkap 800 citra pada tiap simbol ASL untuk data latih dan 200 citra pada setiap simbol untuk data uji. Pada penelitian ini juga menerapkan *Gaussian*

Blur yang digunakan pada saat *input citra*. Penelitian ini menggunakan 2 lapisan algoritma untuk melakukan klasifikasi, lapisan pertama bertugas untuk memprediksi karakter menggunakan CNN dan lapisan kedua bertugas untuk mengklasifikasi kembali simbol-simbol yang memiliki kemiripan. Model CNN dilatih dengan menggunakan pengoptimalan *cross entropy* menggunakan *Adam Optimizer*. CNN digunakan sebagai model utama klasifikasi gestur dan menggunakan *classifier* tambahan untuk mengklasifikasikan kembali simbol-simbol yang mirip. Tahapan proses yang ada dalam penelitian ini dimulai dari pengumpulan dataset, kemudian pra-pemrosesan citra dengan *Gaussian Blur*, ekstraksi fitur, lalu klasifikasi gestur dengan CNN, selanjutnya klasifikasi ulang kemiripan simbol, dan terakhir pembentukan kata dari gestur jari. Arsitektur CNN yang digunakan dapat dilihat pada Gambar 2.9.



Gambar 2. 9 CNN Architecture

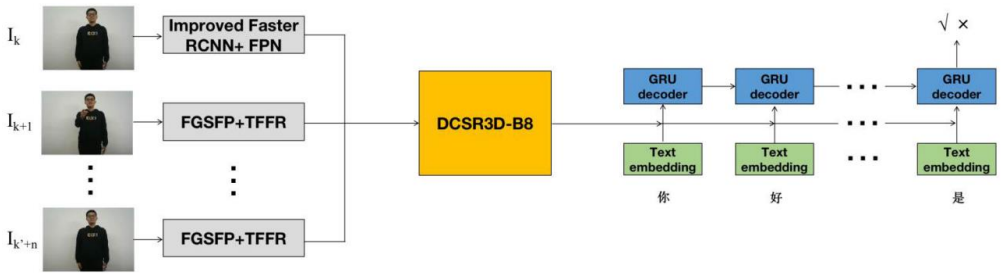
(Kharat, Aditya et.al., 2022)

Pada lapisan *convolutional* dipelajari filter untuk diaktifkan ketika ingin mendeteksi fitur visual, seperti tepi pada arah tertentu atau titik dengan warna tertentu. Sementara itu, pada lapisan *pooling* digunakan untuk mengurangi ukuran matriks aktivasi untuk mendapatkan parameter yang dapat dipelajari. Ada dua jenis *pooling* yang digunakan yaitu *max pooling* dan *average pooling*. Lapisan selanjutnya adalah *fully connected*, digunakan untuk menghubungkan semua *input* ke *neuron*. Lapisan terakhir adalah *final output*, dimana pada lapisan ini telah dihubungkan dengan lapisan *neuron* terakhir dan akan digunakan untuk memprediksi probabilitas setiap citra di tiap kelas yang berbeda. Nilai akurasi akhir yang didapat pada penelitian ini sebesar 98,0% dengan melakukan peningkatan prediksi. Penelitian ini dapat memverifikasi dan memprediksi simbol yang memiliki kemiripan yang hampir sama, sehingga kelebihan dari penelitian ini adalah dapat mendeteksi hampir semua simbol huruf dengan catatan huruf tersebut ditampilkan dengan posisi yang benar, tidak ada *noise* pada *background*, dan pencahayaan yang memadai.

2.11.8 Tinjauan 8

(Zhang, Menglin, Yang, Shuying, Zhao, Min, 2023) ***Deep Learning Based Standard Sign Language Discrimination*.**

Penelitian ini mengusulkan *sign language category* dan *strandardization correctness discrimination model* untuk edukasi pembelajaran bahasa isyarat. Model yang diusulkan diimplementasikan dengan menerapkan *hand detection* dan *standard sign language discrimination method*. *Hand detection* menggunakan usulan metode *utilizes flow guided features* dan membaca penelitian yang relevan yang juga menggunakan *stable and flow key frame detection*. Tujuan dari penelitian ini adalah mengembangkan model *deep learning* untuk diskriminasi akurasi kategori daan standarisasi bahasa isyarat secara komprehensif (Zhang, Menglin et.al., 2023). Struktur dari deteksi tangan dan model diskriminasi kebenaran bahasa isyarat berkelanjutan dapat dilihat pada Gambar 2.10.



Gambar 2. 10 Struktur Sign Language Correctness Discrimination (SLCD)

(Zhang, Menglin et.al., 2023)

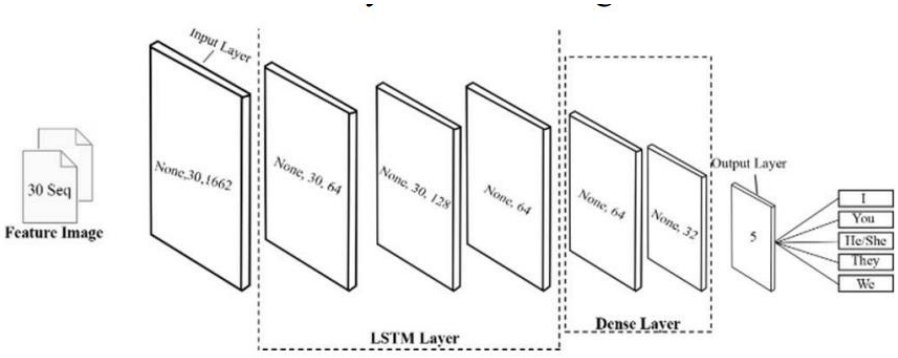
Deteksi yang dilakukan dalam penelitian ini dimulai dari dataset SLCD dikumpulkan dengan label kategori dan standarisasi bahasa isyarat. Selanjutnya, digunakan FGSFP+TFFR untuk mendeteksi tangan dengan *tubelet* dan *optical flow* agar satu atau dua tangan dalam setiap bingkai video dapat dideteksi. SLCD menggunakan *encoder-decoder* DCSR3D+GRU, dimana *encoder* DCSR3D menggabungkan hasil konvolusi 3D dan 2D *deformable* dengan struktur residual. Sementara *decoder* GRU menerima *concatenasi fitur encoding* dan *embedding text*. Dataset dikumpulkan dengan melibatkan 76 siswa yang merekam video bahasa isyarat, dimana setiap video memmiliki dua label, yaitu kategori bahasa isyarat dan kategori standarisasi. Tahapan penelitian dimulai dengan deteksi tangan dalam setiap bingkai video, kemudian dilakukan ekstraksi *patch* tangan, selanjutnya dilakukan *encoding fitur spatiotemporal patch* tangan oleh DCSR3D, selanjutntya *decoding* oleh GRU untuk mendapatkan hasil diskriminasi. Secara keseluruhan, model yang diusulkan mencapai akurasi rata-rata cukup tinggi dalam mendeteksi tangan yaitu 99,0% mAp50 dan untuk membedakan kebenaran bahasa isyarat

mencapai akurasi 81,64% dengan *patch* tangan. Hal ini menunjukkan kemampuan model yang baik dalam diskriminasi akurasi dan standarisasi bahasa isyarat.

2.11.9 Tinjauan 9

(Enri, Ultach, Rozikin, Chaerur, Ilhamsyah, M., Irawan, Agung Susilo Yuda, Garno, Solihin, Indra Permana, Jayanta, 2023) *Sign Language Detection Using Mediapipe and Long Short Term Memory Network*.

Penelitian ini mengimplementasikan LSTM (*Long Short Term Memory*) dan *Mediapipe* untuk mengidentifikasi gerakan bahasa isyarat BISINDO dengan tujuan mengembangkan model *deep learning* (Enri, Ultach et.al., 2023). Penelitian ini menggunakan metode *deep learning* dengan arsitektur LSTM dan data sekuensial. *Mediapipe* digunakan untuk ekstraksi fitur pada setiap citra gerak isyarat. Model arsitektur menggunakan 6 *layer* yang terdiri dari 3 *layer* LSTM dan 3 *layer dense*. Sementara untuk *input* berupa data sekuensial 30 *frame citra*. Model aristektur yang digunakan dapat dilihat pada Gambar 2.11.



Gambar 2. 11 Arsitektur Model

(Enri, Ultach et.al., 2023)

Tahapan penelitian dimulai dari pengumpulan data, dimana data dikumpulkan dari video demonstrasi gerakan bahasa isyarat yang terdiri dari 5 kelas kata ganti orang, yaitu: “Saya”, “Kamu”, “Dia”, “Kami”, dan “Mereka”. Setiap kata diucapkan sebanyak 550 kali dengan total gerakan 2.750, data divalidasi oleh 4 laki-laki dan 2 perempuan penutur asli dan diukur intensitas cahaya saat pengambilan citra. Tahapan kedua adalah eksplorasi data mentah dengan melihat atribut seperti *frame rate video* untuk menemukan perbedaan ukuran *frame* antar data. Tahap ketiga adalah pra-pemrosesan data yang dilakukan dengan penyesuaian rasio *frame* 1:1, mengubah data video menjadi citra (30 citra per gerakan), menghilangkan suara latar video, dan menghasilkan dataset 2.750 video masing-masing 30 *frame*. Tahapan keempat adalah ekstraksi fitur *Mediapipe* digunakan untuk ekstraksi fitur citra gerakan. Diekstraksi 21, 21, 33, dan 468 titik pada telapak tangan kanan, kiri, postur, dan wajah. Hasil ekstraksi disimpan dalam format *array NumPy*. Tahapan kelima adalah pemodelan LSTM, lalu untuk

tahapan keenam adalah pelatihan model yang dilakukan selama 1000 *epoch* dan dilakukan evaluasi menggunakn *callback*. Tahapan selanjutnya adalah pengujian model dan tahapan terakhir adalah evaluasi model. Model dirancang menggunakan *confusion matrix* dan ROC-AUC *score*. Skenario pertama model memiliki kinerja luar biasa dengan akurasi 99% dan 89% untuk tes dan data aktual, sedangkan skor ROC-AUC adalah 99,995% dan 98,390%.

2.11.10 Tinjauan 10

(Ahmad, Nizhamuddin, Wijaya, Eko Saputra, Tjoaquinn, Calvin, Lucky, Henry, Iswanto, Irene Anindaputri, 2023) *Transforming Sign Language using CNN Aproach based on BISINDO Dataset*.

Penelitian ini menggunakan model CNN (*convolutional neural netwrok*) untuk pengenalan BISINDO dengan tujuan untuk mengembangkan model pengenalan bahasa isyarat yang akurat dan efisien berdasarkan dataset BISINDO (Ahmad, Nizhamuddin et al., 2023). CNN dipilih karena kemampuannya mengesktraksi fitur spasial dari data video bahasa isyarat. Subjek dataset yang digunakan adalah 26 gestur tangan dari huruf A sampai Z dalam bahasa isyarat Indonesia (BISINDO) dengan total 936 citra gambar. Model arsitektur menggunakan CNN dengan 3 lapisan *convolutional* dan 1 lapisan *fully connected*. Aktivasi menggunakan ReLU dan fungsi *loss* yang digunakan adalah *categorical cross entropy*. Tahapan proses dalam penelitian ini dimulai dengan pengumpulan data dengan menggunakan dataset BISINDO yang berisi 26 kelas gestur tangan dari huruf A sampai dengan Z, setiap kelas berisi 36 sampel citra dan totalnya berjumlah 936 citra. Tahap kedua adalah tahap normalisasi data, dengan membuat nilai piksel yang dibagi menjadi 255 (nilai maksimum *grayscale* 8-bit) dan menstandarisasi nilai piksel antara 0-1. Tahapan ketiga adalah pemisahan dan palabelan data, dengan membagi data latih sebesar 70% dan data uji 30% dan pemberian nama kelas berdasarkan huruf alfabet. Tahapan keempat adalah pembuatan model CNN dengan menggunakan 3 lapisan konvolusional dengan 128 filter di setiap lapisan, menggunakan *max pooling* untuk mengurangi dimensi dan mengaktivasikan ReLU, 1 lapisan *fully connected* berisi 128 *neuron* dan fungsi aktivasi *output softmax* 26 *neuron*. Tahapan kelima adalah pelatihan model dan tahapan selanjutnya adalah pengujian model menggunakan data uji dataset BISINDO dan melakukan pengukuran dengan presisi, *recall*, dan *f1-score*. Model CNN terbaik memberikan akurasi sebesar 82,56%, presisi 84,76%, *recall* 82,56%, dan *f1-score* 82,30% dalam pengenalan 26 gestur tangan BISINDO. Secara keseluruhan, metode CNN memberikan akurasi yang cukup tinggi dalam pengenalan bahasa isyarat Indonesia berdasarkan dataset BISINDO.

2.11.11 Tinjauan 11

(Agrawal, Agrima, Sreemathy, R., Turuk, Mousami, Jagdale, Jayashree, Kumar, Vishal, 2023) ***Indian Sign Language Recognition using Skin Segmentation and Vision Transformer*****.**

Penelitian ini mengembangkan sebuah model pengenalan Bahasa Isyarat India (*Indian Sign Language*) yang efektif menggunakan teknologi skin Segmentation dan Vision Transformer untuk membantu komunikasi dengan orang yang memiliki keterbatasan dalam berbicara dan mendengar. Penelitian menggunakan dataset primer berisi 72 kata dalam Bahasa Isyarat India. Proses metodologi melibatkan konversi gambar ke YCbCr, segmentasi kulit dengan operasi morfologi, dan penggunaan *Vision Transformer* dengan dua lapis *transformer* yang telah dilatih untuk mengenali dan memproses gambar-gambar tersebut. Model yang diusulkan berhasil mencapai akurasi pengujian sebesar 99.56%. Model ini menunjukkan peningkatan performa dibandingkan dengan model-model sebelumnya dan telah diuji pada beberapa dataset publik dengan hasil yang superior. Subjek dalam penelitian ini melibatkan gambar tangan yang menunjukkan berbagai isyarat dalam Bahasa Isyarat India, yang direkam dalam kondisi yang dikontrol dan diproses untuk menghilangkan latar belakang dan hanya memfokuskan pada bagian tubuh yang relevan seperti tangan dan wajah. Model *Vision Transformer* yang diusulkan menunjukkan efektivitas yang sangat tinggi dalam mengenali Bahasa Isyarat India dengan akurasi yang sangat tinggi. Keberhasilan model ini membuka jalan bagi pengembangan lebih lanjut dalam aplikasi praktis untuk membantu komunikasi dengan komunitas tunarungu dan tunawicara. Tahapan dari penelitian ini adalah pertama dimulai dari pembuatan dataset dengan mengumpulkan dan memproses gambar untuk dataset Bahasa Isyarat India yang berisi 72 kata. Tahapan kedua adalah pra-pemrosesan gambar dengan konversi YCbCr dan segmentasi kulit menggunakan operasi morfologi. Tahapan ketiga adalah augmentasi data dengan memperbanyak data melalui teknik augmentasi seperti rotasi dan perubahan kecerahan. Lalu ada tahapan pelatihan model dengan menggunakan *vision transformer* untuk pelatihan dengan *dataset* yang telah diproses dan tahapan terakhir adalah evaluasi model untuk mengukur performa model dengan menguji pada data uji yang belum dilihat model sebelumnya.

2.11.12 Tinjauan 12

(Tan, Chun Keat, Lim, Kian Ming, Chang, Roy Kwang Yang, Lee, Chin Poo, Alqahtani, Ali, 2023). ***HGR-ViT: Hand Gesture Recognition with Vision Transformer.***

Penelitian ini menggunakan model *Vision Transformer* (ViT) dengan mekanisme perhatian untuk mengenali gestur tangan. Model ini mengolah gambar gestur tangan yang dibagi menjadi potongan tetap dan menggabungkan embedding posisional untuk merepresentasikan informasi

spasial. *Encoder Transformer* standar digunakan untuk mendapatkan representasi gestur tangan, dan kepala *perceptron multilayer* ditambahkan untuk klasifikasi. Tujuan dari penelitian ini adalah untuk meningkatkan pengenalan gestur tangan dengan menggunakan model *Vision Transformer (ViT)*, yang mengatasi kelemahan metode sebelumnya dalam mengkodekan orientasi dan posisi tangan dalam gambar. Penelitian ini menggunakan tiga *dataset* gestur tangan untuk evaluasi: *American Sign Language* (ASL) *dataset*, *ASL with Digits dataset*, dan *National University of Singapore* (NUS) *hand gesture dataset*. Model HGR-ViT mencapai akurasi yang sangat tinggi pada ketiga dataset: 99.98% untuk ASL, 99.36% untuk ASL *with Digits*, dan 99.85% untuk NUS *dataset*. Penelitian ini mengintegrasikan *Vision Transformer* dalam pengenalan gestur tangan statis, sebuah pendekatan baru yang memanfaatkan kekuatan model *Transformer* yang biasanya digunakan dalam pemrosesan bahasa alami untuk aplikasi penglihatan komputer. Tahapan penelitian ini terdiri dari 6 tahapan, yang pertama adalah tahapan pra-pemrosesan gambar dimana, gambar diubah ukuran dan dinormalisasi, kemudian tahapan kedua adalah pembagian gambar, dimana gambar dibagi menjadi potongan-potongan tetap. Tahap ketiga adalah proyeksi linear, dimana potongan gambar diproyeksikan ke ruang dimensi yang lebih rendah. Tahap keempat adalah *encoder transformer* yang digunakan untuk mengolah *embeddings*, kemudian tahapan kelima adalah klasifikasi dengan kepala *perceptron multilayer* yang digunakan untuk menentukan kelas gestur tangan. Tahapan terakhir adalah evaluasi model dengan menggunakan teknik validasi silang pada tiga *dataset*. Penelitian ini berhasil mengembangkan sebuah model yang sangat akurat untuk pengenalan gestur tangan menggunakan *Vision Transformer*, menunjukkan peningkatan signifikan dibandingkan metode yang ada dan membuka peluang untuk penggunaan lebih lanjut dari arsitektur *Transformer* dalam pengenalan gestur tangan. Model ini menunjukkan kinerja yang superior dalam eksperimen, mampu menangani variasi gestur dan kondisi pengambilan gambar dengan baik, memberikan basis yang kuat untuk pengembangan lebih lanjut dalam aplikasi pengenalan gestur secara *real-time*.

2.12 Perbandingan Tinjauan Pustaka

Penelitian yang sudah dilakukan peneliti terdahulu terangkum dalam Tabel 2.1 dibawah ini.

Tabel 2. 1 Perbandingan Tinjauan Pustaka

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
Ojha, Ankit, Pandey,	Deteksi *Real Time* 26 huruf & 10	CNN (*Convolutional*	• Nilai akurasi mencapai 95%	• Implementasi belum dapat mengenali

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
Ayush, Maurya, Shubham, Thakur, Abhishek, P., Dayananda, 2020	angka ASL (*American Sign Language*)	*Neural Network*)	untuk penerjemaahan ejaan jari ASL. • Mampu menerjemahkan bahasa isyarat secara *real time*. • Proses penerjemahan dilakukan secara langsung tanpa adanya penundaan yang signifikan. • Fleksibel sehingga dapat diekspansi ke bahasa isyarat lain dengan melakukan pengumpulan data dan pelatihan ulang model.	bahasa isyarat secara kontekstual karena masih diperlukan tingkat pemrosesan yang lebih tinggi. • Hanya dapat menerjemahkan ASL, belum dapat menerjemahkan bahasa isyarat lainnya. • Ketergantungan yang tinggi terhadap ketepatan postur isyarat yang dimasukkan, sehingga bentuk tangan yang tidak tepat masih berpotensi menghasilkan prediksi yang tidak tepat.
Kembuan, Olivia, Rorimpandey, Gladly Caren, Tengker, Soenandar Milian	Deteksi & Klasifikasi 26 huruf BISINDO	CNN (*Convolutional Neural Network*) & *Library Tensorflow*	• Menggunakan arsitektur CNN dan *library Tensorflow* terbukti sangat baik untuk deteksi dan klasifikasi	• Hanya mengklasifikaskan 26 huruf dari BISINDO, belum mencakup kosakata.

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
Tompunu, 2020			citra gambar dengan nilai akurasi sebesar 96,67%. • Dataset yang digunakan cukup besar yaitu 2.659 gambar dari 26 huruf BISINDO. • Implementasi model menggunakan *Google Colaboratory* sehingga proses pelatihan lebih cepat. • Dapat mengklasifikasikan gambar input dengan akurasi 100% untuk setiap karakter huruf BISINDO.	• Hanya menggunakan data gambar statis, belum mempertimbangkan gerakan dan isyarat visual. • Akurasi 100% hanya berdasarkan data validasi internal, belum diuji dengan data baru di luar dataset. • Performa model belum dievaluasi secara komprehensif dengan metrik lain seperti *precision* dan *recall*.
Dwijayanti, Suci, Hermawati, Taqiyyah, Sahirah Inas, Hikmarika, Hera,	Deteksi & Klasifikasi 26 huruf dan 10 angka BISINDO	CNN (*Convolutional Neural Network*) & dibandingkan dengan model	• Model yang diusulkan dapat mengenali gerakan tangan BISINDO dengan memperoleh hasil kinerja sebesar	• Perlu adanya penyempurnaan Model C untuk mengatasi faktor kinerja karena masih adanya kesalahan prediksi

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
Suprapto, Bhakti Yudho, 2021		AlexNet dan VGG-16.	98,3% dengan pencahayaan redup dan terang serta dari perspektif orang pertama dan orang kedua, sehingga adanya variasi data. • Mengusulkan model C yang merupakan arsitektur baru yang lebih sederhana dan sedikit parameternya dan dibandingkan dengan model modifikasi AlexNet dan VGG-16.	pada beberapa kelas gestur tangan yang bentuknya mirip. • Perlu adanya pertimbangan dalam proses pengambilan data dengan latar belakang yang berbeda-beda. • Belum diujicobakan secara *real time*. • Hanya mengklasifikaskan 26 huruf dari BISINDO, belum mencakup kosakata.
Basri, Syartina Elfarika, Indra, Dolly, Darwis, Herdianti, Mufila, A. Widya,	Deteksi & Klasifikasi 26 huruf BISINDO	*Fourier Descriptor*	• Akurasi pengenalan mencapai 96,92% untuk koefisien *Fourier* 25 dan 50 pada citra standar. • Dapat mengenali citra BISINDO	• Akurasi pengenalan menurun drastis untuk citra yang dirotasi dan ditranslasi, hanya mencapai sekitar 57,43% dan

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
Ilmawan, Lutfi Budi, Purwanto, Bobby, 2021			dengan baik tanpa terpengaruh oleh translasi dan penskalaan. • Penggunaan jarak *Euclidien* sebagai metrik kesamaan fitur.	34,36% secara berturut-turut. • Penelitian ini belum menangani variasi posisi tangan, ekspresi wajah, dan gerakan tubuh dalam bahasa isyarat yang lebih kompleks.
Huang, Shiliang, Ye, Zhongfu, 2021	Pengembangan SLR (*Sign Language Recognition*)	Jaringan *encoder-decoder* berbasis *Boundary Adaptive Encoder* (BAE)	• Kemampuan untuk secara otomatis belajar dan mengkodekan informasi batas sinyal bahasa isyarat, yang membantu dalam pengelolaan informasi jangka pendek dan jangka panjang secara efisien. • Penggunaan subunit subkata dalam bahasa isyarat yang memungkinkan model lebih tepat dalam merepresentasikan	• Kompleksitas tinggi dalam implementasi karena membutuhkan penanganan yang cermat terhadap struktur hirarkis dan adaptasi batas. • Model mungkin tidak ringkas atau optimal untuk implementasi real-time karena ukuran dan kebutuhan komputasi yang tinggi. • Dapat mengalami kesulitan dalam generalisasi ke varian bahasa

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
			nuansa bahasa isyarat. • Penggunaan *window attention* dalam *fase decoding* memperbaiki penanganan input yang berhubungan secara temporal, yang sangat penting untuk pemodelan urutan bahasa isyarat yang panjang.	isyarat yang berbeda tanpa penyesuaian yang signifikan.
Shah, Farman, Shah, Muhammad Saqlain, Akram, Waseem, Manzoor, Awais, Mahmoud, Rasha Orban, Abdelminaam, Diaa Salama, 2021	Deteksi & Klasifikasi 36 huruf statis PSL (*Pakistan Sign Language*)	*Multiple Kernel Learning* dan *Support Vector Machine*	• Nilai akurasi yang didapat adalah 91,93% dan menunjukkan efektivitas teknik MKL dan SVM. • Penelitian ini termasuk penggunaan tangan telanjang tanpa perlu peralatan tambahan yang mahal, seperti sarung tangan berwarna atau	• Ketergantungan kondisi cahaya yang baik dan *background* yang seragam untuk hasil optimal, yang dinilai tidak selalu praktis dalam aplikasi dunia nyata. • Membutuhkan *turning parameter* yang intensif untuk mencapai performa optimal pada set data yang berbeda.

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
			perangkat Kinect, yang membuat metode ini lebih mudah diakses dan murah. • Adanya keunggulan dalam penanganan variasi pencahayaan dan skala dalam gambar.	• Perlu adanya perluasan dan integrasi teknik *deep learning* untuk meningkatkan kemampuan generalisasi model ke variasi yang lebih besar dari gestur dan kondisi pencahayaan. • Pengembangan dataset yang lebih komprehensif yang mencakup lebih banyak variasi dalam gestur dan kondisi perekaman juga diperlukan untuk meningkatkan *robustness* sistem.
Kharat, Aditya, Patil, Yash, Jagtap, Omkar, Sonawale, Rajashri, 2022	Pengembangan *Real Time* 26 huruf ASL (*American Sign Language*)	*OpenCV, Gaussian Blur* Filter dan model CNN	• Akurasi yang tinggi 98% menunjukkan efektivitas model yang digunakan dalam mengenali berbagai simbol tangan.	• Sistem memerlukan kondisi pencahayaan yang baik dan minim gangguan visual di latar belakang untuk

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
			• Menggunakan peralatan yang mudah diakses seperti *webcam* laptop, membuatnya lebih praktis dan ekonomis.	hasil yang optimal. • Pembuatan *dataset* primer menunjukkan keterbatasan dalam ketersediaan *dataset* yang sudah ada, yang bisa mempengaruhi generalisasi sistem. • Pengurangan latar belakang untuk melihat peningkatan akurasi.
Zhang, Menglin, Yang, Shuying, Zhao, Min, 2023	Deteksi SLCD (*Standard Language Correctness Discrimination*)	Pengembangan SLCD menggunakan struktur *encoder-decoder*	• Model yang dikembangkan berhasil mencapai akurasi yang baik dalam mendeteksi dan membedakan bahasa isyarat yang benar dan tidak benar. • Penggunaan teknik deep	• Ketergantungan pada dataset yang secara spesifik dikumpulkan untuk bahasa isyarat Cina, yang mungkin tidak secara langsung dapat diaplikasikan untuk bahasa

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
			learning yang canggih untuk mendeteksi dan membedakan aksi bahasa isyarat yang tepat dan tidak tepat dengan tingkat keberhasilan yang tinggi.	isyarat dari budaya atau bahasa lain. • Pengembangan dataset yang lebih inklusif yang mencakup variasi bahasa isyarat yang lebih luas dari berbagai budaya.
Enri, Ultach, Rozikin, Chaerur, Ilhamsyah, M., Irawan, Agung Susilo Yuda, Garno, Solihin, Indra Permana, Jayanta, 2023	Deteksi & Klasifikasi 5 kelas kata ganti orang dalam BISINDO	Arsitektur LSTM (*Long Short Term Memory*)	• Akurasi pada data uji mencapai 99% dengan skor ROC-AUC sebesar 99.995%. • Pada data aktual, model mencapai akurasi 89% dengan skor ROC-AUC sebesar 98.390%. • Penggunaan Mediapipe untuk ekstraksi fitur meningkatkan kemampuan model dalam mendeteksi pergerakan	• Meskipun performa model sangat baik pada data terkontrol, akurasi menurun ketika diuji dengan data aktual yang diambil tanpa setelan cahaya dan latar belakang yang konsisten. • Model lebih sensitif terhadap variasi kondisi pencahayaan dan latar belakang.

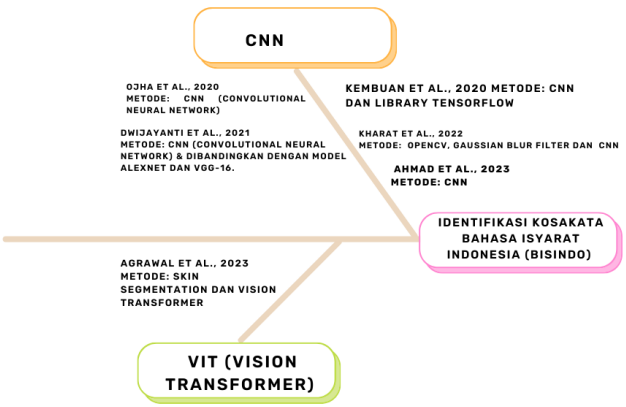
Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
			tangan dan ekspresi wajah.	• Perlu meningkatkan *robustness* model terhadap variasi kondisi pencahayaan dan latar belakang.
Ahmad, Nizhamuddin, Wijaya, Eko Saputra, Tjoaquinn, Calvin, Lucky, Henry, Iswanto, Irene Anindaputri, 2023	Deteksi & Klasifikasi 26 huruf BISINDO	Metode CNN	• Penelitian ini berhasil mengembangkan model pengenalan bahasa isyarat yang menghasilkan akurasi sebesar 82,56%. Model ini menggunakan arsitektur CNN dengan konfigurasi filter gambar sebanyak 128, epoch sebanyak 15, dan pembagian data 70% untuk pelatihan serta 30% untuk pengujian.	• Penelitian masih membutuhkan pengembangan lebih lanjut untuk meningkatkan kinerja dalam lingkungan yang bising dan dinamis. • Terbatasnya jumlah data bisa mempengaruhi generalisasi model. • Menggunakan pre-train model seperti VGG16 atau Inception V3 untuk melihat potensi hasil yang lebih baik.

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
			• Penggunaan CNN yang efektif dalam mengenali fitur spasial dari data gambar dan video, cocok untuk tugas pengenalan bahasa isyarat.	• Melakukan eksperimen dengan menambah jumlah data dan variasi gestur tangan untuk meningkatkan robustness model. • Menguji model dalam berbagai kondisi lingkungan untuk memastikan efektivitasnya dalam aplikasi dunia nyata.
Agrawal, Agrima, Sreemathy, R., Turuk, Mousami, Jagdale, Jayashree, Kumar, Vishal, 2023	Pengenalan 72 kata *Indian Sign Language*	*Skin Segmentation* dan *Vision Transformer*	• Akurasi yang sangat tinggi 99,56% menunjukkan keefektifan model. • Penggunaan *vision transformer* yang inovatif dalam pengenalan isyarat. • Penerapan ekstraksi fitur warna segmentasi	• Keterbatasan *dataset* sehingga tidak mencerminkan variasi penuh dari ISL, sehingga membatasi kemampuan generalisasi model ketika dihadapkan pada variasi isyarat yang lebih luas

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
			kulit untuk meningkatkan akurasi pengenalan isyarat.	atau dalam kondisi yang kurang ideal. • Ketergantungan pada pra-pemrosesan seperti segmentasi kulit, yang mungkin tidak selalu efektif dalam semua kondisi pencahayaan atau untuk semua warna kulit. • Diperlukan eksplorasi arsitektur model yang lebih kompleks.
Tan, Chun Keat, Lim, Kian Ming, Chang, Roy Kwang Yang, Lee, Chin Poo, Alqahtani, Ali, 2023	Pemodelan HGR-ViT untuk mengklasifikasikan 3 *dataset* ASL, ASL *with digits dataset*, NUS *hand gesture dataset*.	*Vision Transformer* (ViT)	• Model HGR-ViT berhasil mencapai akurasi yang sangat tinggi pada semua *dataset* yang digunakan (99.98% untuk ASL, 99.36% untuk ASL *with Digits*, dan 99.85% untuk NUS *hand gesture dataset*), menunjukkan	• Ketergantungan pada *dataset* besar. • Kompleksitas komputasi dibandingkan dengan model CNN standar. • Kemampuan menangkap detail halus untuk membedakan kemiripan dari gestur.

Peneliti, Tahun	Subjek Penelitian	Metode Penelitian	Kelebihan	Kekurangan
			efektivitas model dalam mengenali berbagai gestur tangan.	• Diperlukan pengembangan terhadap aplikasi secara *real time.* • Pengenalan gestur dinamis bukan hanya gestur statis. • Eksplorasi arsitektur hibrid menggabungkan kekuatan CNN dalam menangkap fitur lokal dan *vision transformer* dalam memproses konteks global.

2.13 *Roadmap* Penelitian



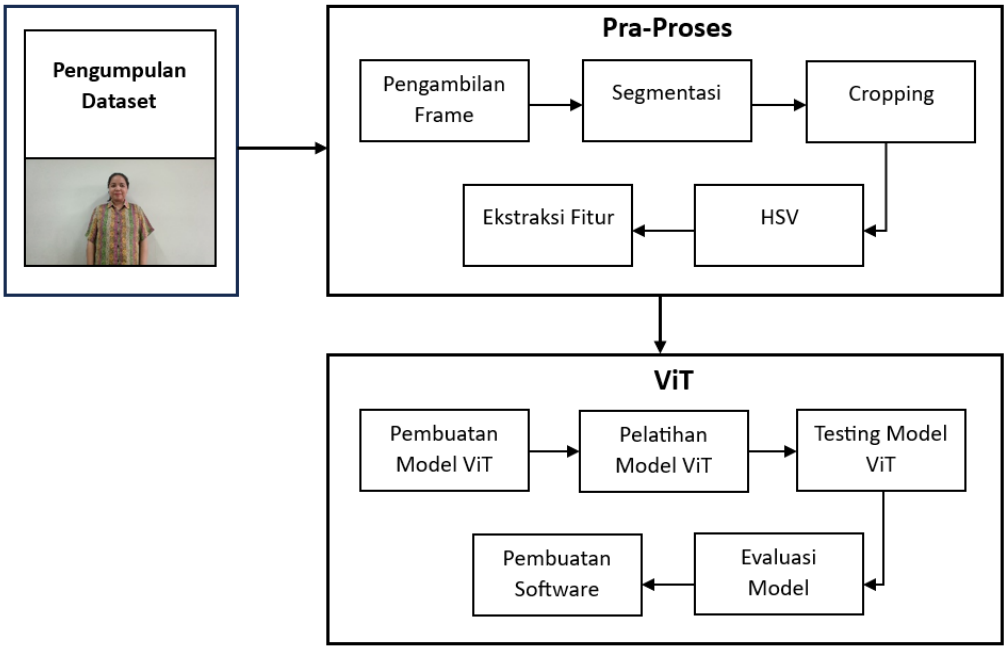
Gambar 2. 12 Fishbone Penelitian

BAB 3

METODE PENELITIAN

3.1 Tahapan Penelitian

Sesuai dengan topik, rumusan masalah dan tujuan penelitian yang ingin dicapai, maka disusun metode atau langkah-langkah penelitian seperti yang diperlihatkan oleh bagan pada gambar 3.1. Penelitian ini terdiri dari lima tahap.



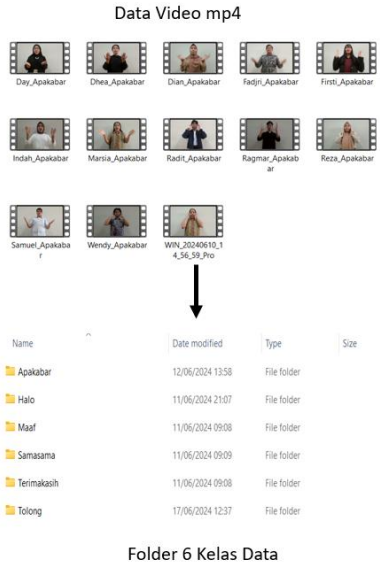
Gambar 3. 1 Bagan Tahapan Penelitian

Tahapan penelitian diatas menggambarkan tahapan yang akan dilakukan dalam penelitian. Tahapan penelitian akan dijelaskan secara mendetail sebagai berikut:

3.1.1 Pengumpulan Dataset

Pengumpulan data diambil di SLB B/C yang berlokasi di Jalan Pangkalan Asem, Gg. 1 No.1, Cempaka Putih Barat, Cempaka Putih, Jakarta Pusat. Pengumpulan data dilakukan secara langsung (*real time*). Penekanan konsep *real time* disini adalah kecepatan proses akuisisi dari objek yang diambil dan proses analisi yang dilakukan sama dengan proses gerak yang didasarkan oleh standar gerak tangan pada kosakata BISINDO. Spesifikasi ruangan yang digunakan secara khusus tidak diperlukan, namun dengan catatan ruangan yang digunakan harus memiliki cahaya yang cukup dan disesuaikan dengan kemampuan

kamera. Untuk menghasilkan citra atau gambar yang baik maka proses kalibrasi kamera dan ruangan tetap diperlukan. Data yang diambil dalam bentuk video mp4 yang disimpan dan diberi nama sesuai dengan nama kelas yang sudah ditentukan, yaitu: kelas Apakabar, kelas Halo, kelas Maaf, kelas Samasama, kelas Terimakasih, dan kelas Tolong. Video berdurasi 10-12 detik yang digunakan untuk menangkap gerakan bahasa isyarat Indonesia dari awal hingga akhir. Gambar 3.2 akan memperlihatkan data video mp4 yang sudah diberi nama dan dimasukkan kedalam folder kelas yang telah ditentukan.



Gambar 3. 2 Pengumpulan *Dataset*

Algoritma 3.1: Tahapan Pengumpulan Dataset
Input: Gerakan tangan BISINDO
1: Atur posisi antara jarak kamera, pencahayaan tambahan dan objek
2: Kumpulkan dataset sesuai dengan kelas yang telah ditentukan
3: Inisialisasi dataset
Output: Video mp4 yang sudah diinisialisasi

3.1.2 Pra-Proses

Tahapan pra proses dalam penelitian ini adalah proses pemisahan antara *background* terhadap *foreground* pada gerakan objek. Pra proses terdiri dari proses konversi *video* menjadi *frame*, konversi ruang warna menjadi citra keabuan, segmentasi dengan algoritam *threshold*, *cropping*, konversi citra ke *HSV*, dan ekstraksi fitur menggunakan model *skeleton*.

Algoritma 3.2: Tahapan Pra Proses

Input: Dataset mp4 yang sudah diinisialisasi

1: Konversi *video* ke *frame* sebanyak 30 fps

2: Konversi citra berwarna ke citra *grey-level*

3: Segmentasi frame ke dalam algoritma *threshold*

4: Proses *cropping* dan tentukan RoI (*Region of Interest*)

5: Konversi citra ke format *HSV* (*Hue, Saturation*, *Value*)

6: Ekstraksi fitur menggunakan *skeleton*

Output: Citra Skeleton

3.1.2.1 Pengambilan Frame

Video yang telah diambil sebelumnya pada tahapan pengumpulan *dataset*, kemudian dikonversi dengan *frame rate* 30 *frame/*detik selama 10 detik. Proses ekstraksi video dengan memisahkan *frame-frame* dalam sebuah video sehingga setiap *frame* menjadi 300 *frame* citra yang independen satu terhadap lainnya. Data video yang sudah diekstrak menjadi *frame* disimpan di folder Frame_Dataset. Alur proses pengambilan *frame* dapat dilihat seperti pada Gambar 3.3.

Gambar 3. 3 Proses Pengambilan *Frame*

3.1.2.2 Segmentasi

Pada tahapan segmentasi ini menggunakan algoritma *Threshold* sebagai teknik dasar dalam pemrosesan citra. Sebelum dilakukan segmentasi menggunakan algoritma *threshold*, dilakukan proses konversi citra berwarna ke citra keabuan. Semua hasil citra hasil proses ekstraksi *frame video* merupakan citra berwarna dalam ruang warna

RGB (*Red, Green, Blue*), lalu dikonversi menjadi citra *grey-level*, baru setelahnya dilakukan segmentasi dengan *threshold* dengan nilai tertentu. Tujuannya untuk memisahkan objek dari latar belakang dengan cara mengubah gambar menjadi biner berdasarkan intensitas piksel.

Algoritma 3.3: Tahapan Segmentasi

Input: Video yang sudah dikonversi menjadi *frame* berwarna

1: *Frame* berwarna untuk kelas kosakata BISINDO yang sudah diakuisisi

2: Konversi citra berwarna ke citra *grey-level*

3: Segmentasi menggunakan *threshold*

Output: citra keabuan yang sudah disegmentasi menggunakan threshold

3.1.2.3 *Cropping*

Setelah dilakukan segmentasi, lalu area yang mengandung objek di crop untuk mengisolasi objek dari bagian lain dari gambar. Proses *cropping* disebut uga dengan pengambilan area untuk mendapatkan RoI (*Region of Interest*). RoI yang dimaksud dalam penelitian ini adalah *region* yang meliputi luas area yang dibutuhkan untuk menampilkan semua pergerakan tangan kosakata bahasa isyarat Indonesia (BISINDO).

3.1.2.4 HSV (Hue, Saturation, Value)

Gambar yang telah di *crop* kemudian dikonversi dari *grayscale* ke format HSV. Ekstraksi fitur dilakukan dengan menganalisis distribusi nilai *hue, saturation,* dan *value* dalam gambar. Fitur dalam HSV memberikan informasi yang berguna tentang warna dan kecerahan untuk proses pengenalan pola.

3.1.2.5 Ekstraksi Fitur

Langkah selanjutnya adalah ekstraksi fitur dengan menggunakan model *skeleton*. Tujuan dari skeletonisasi adalah untuk mereduksi objek dalam citra ke bentuk rangka dasarnya (*skeleton*), yang merupakan representasi minimal dari struktur tersebut. Gambar 3.4 adalah *flowchart* dari tahapan skeletonisasi.

Algoritma 3.4: Tahap Pembentukan *Skeleton*

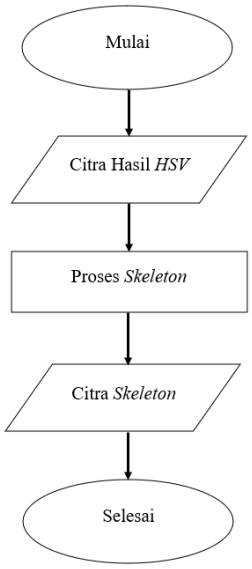
Input: citra hasil hsv

1: Baca citra hasil hsv

2: Proses skeletonisasi

3: Simpan hasil *skeleton*

Output: Citra *Skeleton*



Gambar 3. 4 Tahap Proses Skeletonisasi

3.1.3 Vision Transformer (ViT)

Tujuan dari tahapan ini adalah untuk membangun metode *Vision Transformer* (ViT) berbasis model *skeleton*. Tahapan ini terdiri dari pembuatan model ViT, pelatihan model ViT, *testing* model ViT, evaluasi model, dan terakhir adalah tahapan pembuatan *software*. Pada tahapan pembuatan model akan dilakukan inisialisasi arsitektur dari *vision transformer* yang akan digunakan, lalu untuk pelatihan akan melibatkan pelatihan model dengan dataset yang sudah diproses, model akan belajar untuk mengenali pola dari fitur yang diekstrak. Selanjutnya, pada tahapan *testing* akan dilakukan pengujian pada *dataset* baru atau bagian dari *dataset* yang tidak digunakan dalam pelatihan. Evaluasi model dilakukan untuk menganalisis hasil dari *testing*, seperti akurasi, *precision,* dan *recall*. Jika model sudah dirasa baik dalam mengidentifikasi gerakan tangan koskata BISINDO dan akurasi sudah tinggi, maka tahapan terakhir adalah proses pembuatan perangkat lunak (*software*).

3.2 Jadwal Penelitian

Tabel 3. 1 Jadwal Penelitian

No	Uraian	Tahun 1		Tahun 2		Tahun 3	
		Sem1	Sem2	Sem1	Sem2	Sem1	Sem2
1	Studi Literatur						
2	Perencanaan Penelitian						

No	Uraian	Tahun 1		Tahun 2		Tahun 3	
		Sem1	Sem2	Sem1	Sem2	Sem1	Sem2
3	Ujian Kualifikasi						
4	Pengolahan penelitian						
4	*Progress report*						
5	Publikasi						
6	Sidang Tertutup						
7	Sidang Terbuka						

Daftar Pustaka

Agarwal, A. (2023). Indian Sign Language Recognition using Skin Segmentation and Vision Transformer. *2023 IEEE 20th India Council International Conference (INDICON)*, 857–862. https://doi.org/10.1109/INDICON59947.2023.10440818

Ahmad, N., Wijaya, E. S., Tjoaquinn, C., Lucky, H., & Iswanto, I. A. (2023). Transforming Sign Language using CNN Approach based on BISINDO Dataset. *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, 543–548. https://doi.org/10.1109/icimcis60089.2023.10349011

Anwar, M. K. (2017). Pembelajaran Mendalam untuk Membentuk Karakter Siswa sebagai Pembelajar. *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah*, *2*(2), 97. https://doi.org/10.24042/tadris.v2i2.1559

Asmara, R. (2015). *BASA-BASI DALAM PERCAKAPAN KOLOKIAL BERBAHASA JAWA SEBAGAI PENANDA KARAKTER SANTUN BERBAHASA*. *11*(September), 80–95.

Bar, G., & Goldberg, Y. (2022). *Neural Network Methods for Natural Language Processing*. *December 2017*. https://doi.org/10.1162/COLI

Basri, S. E., Indra, D., Darwis, H., Mufila, A. W., Ilmawan, L. B., & Purwanto, B. (2021). Recognition of Indonesian Sign Language Alphabets Using Fourier Descriptor MethodBasri, Syartina Elfarika Indra, Dolly Darwis, Herdianti Mufila, A. Widya Ilmawan, Lutfi Budi Purwanto, Bobby. *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*, 405–409. https://doi.org/10.1109/EIConCIT50028.2021.9431883

Cholissodin, I., & Soebroto, A. A. (2021). *AI , MACHINE LEARNING & DEEP LEARNING (Teori & Implementasi). July 2019*.

Crystal, David. (1991). *A Dictionary of Linguistics and Phonetics*. Basil Blackwell.

Dwijayanti, S., Hermawati, Taqiyyah, S. I., Hikmarika, H., & Suprapto, B. Y. (2021). Indonesia Sign Language Recognition using Convolutional Neural Network. *International Journal of Advanced Computer Science and Applications*, *12*(10), 415–422. https://doi.org/10.14569/IJACSA.2021.0121046

Enri, U., Rozikin, C., Ilhamsyah, M., Irawan, A. S. Y., Garno, Solihin, I. P., & Jayanta. (2023). Sign Language Detection Using Mediapipe and Long-Short Term Memory Network. *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, 617–622. https://doi.org/10.1109/icimcis60089.2023.10349016

Eriana, E. S., & Zein, D. A. (2023). *Artificial Intelligence (Ai) Penerbit Cv. Eureka Media Aksara*. 24–32.

Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. *Genetic Programming and Evolvable Machines*, *19*(1–2), 305–307. https://doi.org/10.1007/s10710-017-9314-z

Huang, S., & Ye, Z. (2021). Boundary-Adaptive Encoder with Attention Method for Chinese Sign Language Recognition. *IEEE Access*, *9*, 70948–70960. https://doi.org/10.1109/ACCESS.2021.3078638

Indra, D., Madenda, S., & Wibowo, E. P. (2017). Recognition of Bisindo alphabets based on chain code contour and similarity of Euclidean distance. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(5), 1644–1652. https://doi.org/10.18517/ijaseit.7.5.2746

Indra, D., Purnawansyah, Madenda, S., & Wibowo, E. P. (2019). Indonesian sign language recognition based on shape of hand gesture. *Procedia Computer Science*, *161*(September), 74–81. https://doi.org/10.1016/j.procs.2019.11.101

Kautsar, I., Indra Borman, R., Sulistyawati, A., Informatika STMIK TEKNOKRAT Bandar Lampung Jl Zainal Abidin Pagaralam No, T. H., & Ratu Bandar Lampung, L. (2015). Aplikasi Pembelajaran Bahasa Isyarat Bagi Penyandang Tuna Rungu Berbasis Android Dengan Metode Bisindo. *Semnasteknomedia Online*, *3*(1), 4-4–69. https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/832

Kembuan, O., Rorimpandey, G. C., & Tengker, S. M. T. (2020). Convolutional Neural Network (CNN) for Image Classification of Indonesia Sign Language Using Tensorflow. *2020 2nd International Conference on Cybernetics and Intelligent System, ICORIS 2020*, *26*. https://doi.org/10.1109/ICORIS50180.2020.9320810

Kharat, A., Patil, Y., Jagtap, O., & Sonawale, R. (2022). *Sign Language to Text Conversion*. *2*(1).

Mamulak, N. M. R., Nani, P. A., & Sooai, A. G. (2021). *DETEKSI ANOMALI & PEMBELAJARAN MESIN*. PENERBIT KBM INDONESIA.

Ojha, A., Pandey, A., Maurya, S., Thakur, A., & P., D. (2020). Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network. *International Journal of Research Publication and Reviews*, *8*(2), 9–17.

Putri, V. A., Carissa, K., Sotyawardani, A., & Rafael, R. A. (2023). Peran Artificial Intelligence dalam Proses Pembelajaran Mahasiswa di Universitas Negeri Surabaya. *Prosiding Seminar Nasional*, 615–630.

Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *5*(1), 75–82. https://doi.org/10.31294/ijcit.v5i1.7951

Sandra, R., Zebua, Y., Khairunnisa, M. P., Pd, S., Hartatik, M. C., Si, S., Pariyadi, M. S., Kom Dessy, M., Wahyuningtyas, P., Pd, M., Ahmad, M., & Thantawi, S. T. (2023). *Fenomena Artificial Intelligence (Ai)* (Issue June). www.researchgate.net

Setyadi, A. (2021). *Budaya kesantunan penggunaan kata: maaf, tolong, terima kasih dalam berkomunikasi*. *5*(1), 87–93.

Shah, F., Shah, M. S., Akram, W., Manzoor, A., Mahmoud, R. O., & Abdelminaam, D. S. (2021). Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language. *IEEE Access*, *9*, 67548–67558. https://doi.org/10.1109/ACCESS.2021.3077386

Sooai, A. G., Katolik, U., Mandira, W., Magdalena, N., Mamulak, R., Katolik, U., Mandira, W., Nani, P. A., Katolik, U., & Mandira, W. (2021). *Deteksi Anomali & Pembelajaran Mesin* (Issue April).

Svoboda, T., Kybic, J., & Hlavas, V. (2007). *Image Processing, Analysis & Machine Vision a MATLAB companion.* Thomson Learning.

Tan, C. K., Lim, K. M., Kwang, R., Chang, Y., & Lee, C. P. (2023). *HGR-ViT : Hand Gesture Recognition with Vision Transformer*. 1–20.

Towards, A. I., & Vision, I. (n.d.). *S T RAT E GI N AS ION AL KE C E RD A S A N A RT I F I S I A L I ND ON E S IA TAHUN*.

WIPO. (2019). *Artificial Intelligence*. World Intellectual Property Organization.

Zhang, M., Yang, S., & Zhao, M. (2023). Deep Learning-Based Standard Sign Language Discrimination. *IEEE Access*, *11*(October), 125822–125834. https://doi.org/10.1109/ACCESS.2023.3330863