

ASL FOR BEGINNERS: SISTEM PENGIKTIRAFAN BAHASA ISYARAT MASA NYATA MENGGUNAKAN PEMBELAJARAN DALAM DAN PENGEMBARAN WEB FLASK

¹Thanush A/L Govindarajoo, ¹Dr. Nor Samsiah binti Sani

¹Fakulti Teknologi & Sains Maklumat

43600 Universiti Kebangsaan Malaysia

Abstrak

Projek ini membentangkan sistem pengecaman Bahasa Isyarat Amerika (ASL) masa nyata yang mampu mengenal pasti isyarat tangan statik mewakili 26 abjad Inggeris dan digit 0 hingga 9. Matlamatnya adalah untuk meningkatkan aksesibiliti komunikasi bagi individu yang bergantung pada ASL, menggunakan penyelesaian pembelajaran mendalam ringan yang beroperasi dengan boleh dipercayai menggunakan webkam standard dalam persekitaran masa nyata. Kebanyakan model pengecaman ASL yang tersedia dalam talian pada tahun 2025 mengalami sama ada kos pengiraan yang tinggi, ketepatan terhad dalam persekitaran tidak terkawal, atau liputan isyarat yang tidak mencukupi. Untuk menangani isu-isu ini, pengelas ASL tersuai (asl_model.h5) telah dilatih menggunakan dua set data Kaggle, meliputi 36 isyarat ASL secara keseluruhan. Langkah-langkah pra-pemprosesan termasuk menapis kelas tidak berkaitan, mengimbangi saiz set data, dan menggunakan penambahan data untuk meningkatkan penjejakan. Pemindahan pembelajaran telah diaplikasikan menggunakan MobileNetV2, diikuti dengan proses penalaan halus dua fasa untuk meningkatkan kelajuan dan ketepatan. Model akhir mencapai ketepatan ujian sebanyak 90%, dengan ketepatan purata makro, peringatan, dan skor F1 kesemuanya melebihi 91%. Ia juga menunjukkan prestasi tinggi pada isyarat yang serupa secara visual seperti F lawan T dan O lawan Q. Model yang dioptimumkan hanya seberat 12.8 MB, menjadikannya amat sesuai untuk aplikasi masa nyata dalam persekitaran web atau mudah alih. Sistem ini disepadukan ke dalam aplikasi full-stack menggunakan Flask (Python) untuk pemprosesan backend, MediaPipe dan OpenCV untuk penjejakan tangan, dan frontend dibina dengan HTML, JavaScript, dan CSS. Pengguna boleh berlatih isyarat individu, melihat maklum balas ramalan, dan menerima isyarat visual/audio untuk ketepatan. Antaramuka ini juga termasuk ciri penyesuaian seperti mod gelap, penskalaan teks, dan penjejakan prestasi langsung. Projek ini berjaya menyampaikan platform pengecaman ASL yang pantas, tepat dan mesra pengguna jauh lebih praktikal berbanding banyak model lebih berat atau alat luar talian yang tersedia pada masa ini.

Kata Kunci: ASL, MobileNetV2, MediaPipe, OpenCV, Flask.

Abstract

This project presents a real-time American Sign Language (ASL) recognition system capable of identifying static hand gestures representing the 26 English alphabets and digits 0 through 9. The goal is to improve communication accessibility for individuals who rely on ASL, using a lightweight deep learning solution that works reliably with standard webcams in real-time environments. Most ASL recognition models available online in 2025 suffer from either high computational cost, limited accuracy in uncontrolled environments, or insufficient sign coverage. To address these issues, a custom ASL classifier (asl_model.h5) was trained using two Kaggle datasets, covering 36 ASL signs in total. Preprocessing steps included filtering out non-relevant classes, balancing dataset sizes, and applying data augmentation to improve generalization. Transfer learning was applied using MobileNetV2, followed by a two-phase fine-tuning process to enhance both speed and accuracy. The final model achieves a test accuracy of 90%, with macro-averaged precision, recall, and F1-score all exceeding 91%. It also demonstrates high performance on visually similar signs such as F vs. T and O vs. Q. The optimized model weighs just 12.8 MB, making it highly suitable for real-time applications in web or mobile environments. The system was integrated into a full-stack application using Flask (Python) for backend processing, MediaPipe and OpenCV for hand tracking, and a frontend built with HTML, JavaScript, and CSS. Users can practice individual signs, view prediction feedback, and receive visual/audio cues for accuracy. The interface also includes customization features like dark mode, text scaling, and live performance tracking. This project successfully delivers a fast, accurate, and user-friendly ASL recognition platform — significantly more practical than many heavier models or offline tools currently available.

Keywords: ASL, MobileNetV2, MediaPipe, OpenCV, Flask

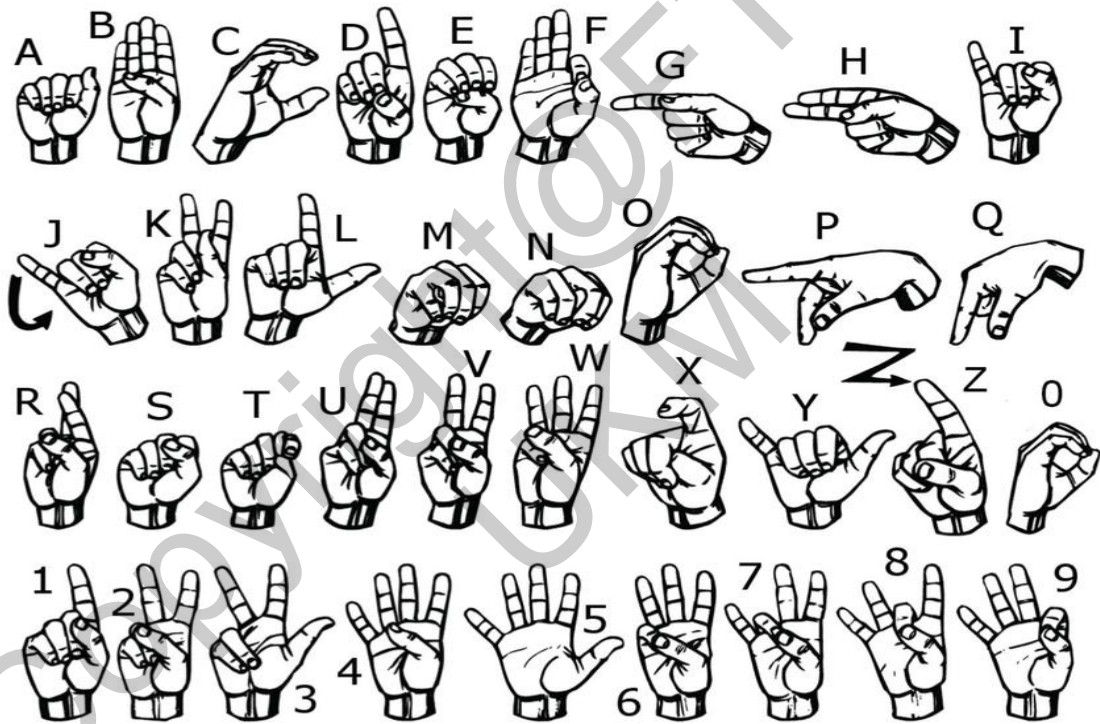
1.0 PENGENALAN

Dalam tahun kebelakangan ini, kemajuan dalam visi komputer (computer vision) dan pembelajaran mendalam (deep learning) telah membolehkan mesin memahami isyarat manusia, bahasa, dan komunikasi dengan lebih baik. Satu bidang penyelidikan dan pembangunan yang penting ialah pengecaman Bahasa Isyarat Amerika (ASL), yang memberi tumpuan kepada membolehkan mesin mentafsir dengan tepat isyarat tangan statik yang digunakan oleh individu yang ada kurang pendengaran.

ASL ialah bahasa visual lengkap yang menggunakan bentuk tangan, pergerakan, dan ekspresi muka untuk menyampaikan makna. Bagi berjuta-juta pengguna, ia berfungsi sebagai kaedah komunikasi utama mereka. Walau bagaimanapun, kekurangan kefahaman meluas tentang ASL dalam kalangan masyarakat umum telah mewujudkan halangan komunikasi. Ini mendorong keperluan untuk sistem pintar yang boleh berfungsi sebagai penterjemah masa

nyata, merapatkan jurang antara pengguna bahasa isyarat dan mereka yang tidak mengetahuinya.

Kaedah tradisional untuk pengecaman ASL sering bergantung pada pengekstrakan ciri (feature extraction), sarung tangan berasaskan sensor, atau model pembelajaran mesin cetek (shallow machine learning) dengan ketepatan terhad. Perkembangan terkini dalam pembelajaran mendalam, terutamanya penggunaan Rangkaian Neural Konvolusi (Convolutional Neural Networks, CNNs), telah meningkatkan secara signifikan tugas pengelasan (classification) isyarat dan imej. Dengan melatih CNN pada set data besar isyarat ASL, kini mungkin untuk membina sistem pengecaman yang tepat, pantas, dan boleh ditingkatkan.



Rajah 1 Contoh Isyarat Tangan ASL untuk A–Z dan 0–9

Projek ini bertujuan membangunkan sistem pengecaman ASL masa nyata yang mengelaskan isyarat tangan abjad (A–Z) dan digit (0–9) menggunakan webkam. Sistem ini menggunakan pemindahan pembelajaran (transfer learning) dengan seni bina CNN MobileNetV2 untuk mengekstrak ciri spatial daripada imej dan mengelaskannya dengan ketepatan tinggi. Untuk meningkatkan penjanaan model, penambahan data (data augmentation), pengimbangan kelas (class balancing), dan strategi penalaan halus (fine-tuning) dua fasa telah

diaplikasikan. Model terlatih akhir adalah ringan (12.8 MB) dan mampu membuat ramalan pada kelajuan masa nyata dengan ketepatan 90% pada data ujian tidak dilihat.

Secara keseluruhannya, projek ini menunjukkan bagaimana model pembelajaran mendalam moden boleh disesuaikan untuk membina alat akses yang memberdayakan pengguna bahasa isyarat dalam interaksi harian mereka. Sistem ini boleh digunakan dalam alat pendidikan, pembantu komunikasi, dan teknologi inklusif yang mempromosikan aksesibiliti dan integrasi sosial.

2.0 KAJIAN LITERATUR

Berdasarkan kajian literatur yang dibuat, beberapa kajian terkini telah meneroka penggunaan teknologi pembelajaran mendalam (DL) dan visi komputer untuk mengenal pasti dan mengelaskan isyarat Bahasa Isyarat Amerika (ASL) dalam pelbagai konteks. Bahagian berikut merumuskan perbandingan utama dan perbezaan yang dikenal pasti merentas kajian ini:

Kajian oleh Munib et al. bertujuan mengenal pasti isyarat tangan ASL statik menggunakan kamera tradisional dan pengelas rangkaian neural. Sistem ini mencapai ketepatan pengecaman 92.33%, menunjukkan bahawa pengelasan imej berasaskan kamera dengan model DL boleh menjadi sangat berkesan tanpa keperluan perkakasan khusus.

Oz dan Leu menjalankan pelbagai kajian mengenai pengecaman ASL statik dan dinamik menggunakan sensor berasaskan sarung tangan. Sistem ASL dinamik mereka mencapai ketepatan 95%, manakala sistem statik berasaskan sarung tangan mencapai 90%. Walaupun sarung tangan memberikan data pergerakan tangan yang tepat, ia sering mengganggu dan tidak praktikal untuk penggunaan santai atau awam.

Ragab et al. melaksanakan model pengecaman isyarat statik berasaskan kamera menggunakan teknik pembelajaran mesin klasik seperti Mesin Vektor Sokongan (SVM) dan Random Forest, mencapai ketepatan 94%. Ini menunjukkan bahawa walaupun algoritma tradisional boleh berprestasi baik, ia sering memerlukan pengekstrakan ciri manual, menghadkan kebolehskalaan.

Kajian oleh Sun et al. dan Usachokcharoen et al. menggunakan sensor Kinect untuk pengecaman ASL dinamik, dengan ketepatan antara 86% hingga 95% bergantung pada pengelas digunakan (cth. Adaboost, SVM, Latent SVM). Walaupun model berasaskan sensor

menawarkan data kedalaman, ia terhad oleh kos dan kebolehcapaian dalam aplikasi web atau mudah alih masa nyata.

Karya lain seperti oleh Chuan et al. dan Jangyodsuk et al. masing-masing meneroka Leap Motion dan penyediaan pelbagai sensor. Walaupun Leap Motion mencapai prestasi sederhana (72%–79%), sistem pelbagai sensor seperti penyediaan kamera+Kinect mencapai lebih 93%, tetapi sekali lagi menimbulkan cabaran integrasi perkakasan.

Antara ketepatan tertinggi yang dilaporkan adalah dari Zamani dan Kanan (99.88% menggunakan kamera dan rangkaian neural) serta Aryanie dan Heryadi (99.8% menggunakan KNN dengan $k=3$). Kajian ini menunjukkan potensi sistem berasaskan kamera ringan apabila digabungkan dengan model dioptimumkan baik dan set data seimbang.

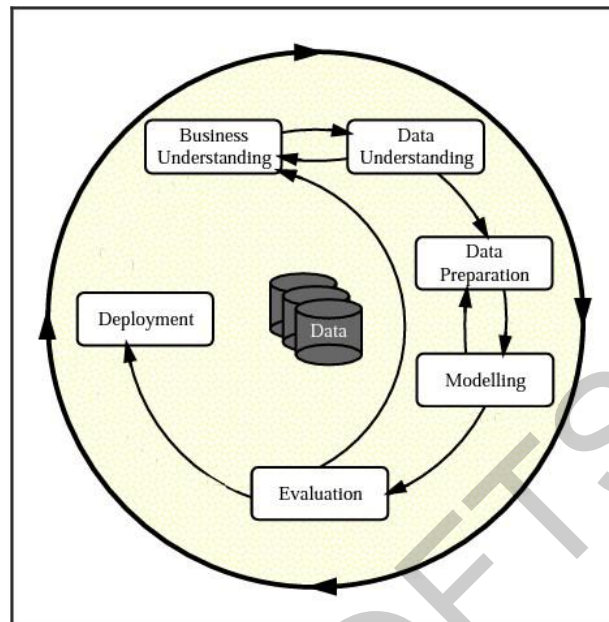
Daripada perbandingan di atas, adalah jelas bahawa pembelajaran mendalam, khususnya Rangkaian Neural Konvolusi (CNN), memainkan peranan penting dalam mencapai ketepatan tinggi dalam pengecaman isyarat ASL. Walaupun sumber input alternatif seperti sarung tangan, Kinect, atau sensor EEG boleh memperkaya data, ia memperkenalkan kos dan kerumitan. Model DL berasaskan kamera kekal sebagai penyelesaian paling praktikal dan boleh diakses untuk aplikasi masa nyata yang boleh ditingkatkan.

Untuk projek ini, sistem pengecaman ASL berasaskan kamera telah dipilih, menggunakan CNN dengan MobileNetV2 sebagai seni bina teras. Pendekatan ini selari dengan penyelidikan semasa yang menunjukkan model DL berasaskan kamera boleh menawarkan kedua-dua ketepatan tinggi dan prestasi masa nyata. Pemerhatian yang diperoleh daripada kajian literatur ini secara langsung memaklumkan pemilihan model, strategi latihan, dan reka bentuk pelaksanaan dalam bab seterusnya projek ini.

3.0 METODOLOGI

Proses pembangunan sistem pengecaman Bahasa Isyarat Amerika (ASL) ini adalah berdasarkan pendekatan CRISP-DM yang menggariskan fasa pembangunan sistematik, bermula dari pemahaman masalah sehinggalah pelaksanaan akhir sistem. CRISP-DM memberi kerangka berstruktur dan berulang untuk membina sistem pengecaman isyarat tangan berasaskan kamera dan pembelajaran mendalam, serta membolehkan pelarasan berterusan sepanjang kitaran

pembangunan. Seni bina model CRISP-DM ditunjukkan dalam Rajah 2, yang menjadi panduan utama dalam pelaksanaan projek ini.



Rajah 2 Seni bina model CRISP-DM

Sumber: Wirth, R. & Hipp, J. 2000

Proses pembangunan ini dibahagikan kepada empat komponen utama iaitu pemilihan dan penyediaan set data, prapemprosesan data, pembinaan dan latihan model pengecaman, serta pembangunan aplikasi web yang mengintegrasikan model tersebut. Untuk mencapai objektif ini, dua set data berbeza digunakan — satu untuk huruf A hingga Z dan satu lagi untuk angka 0 hingga 9. Set data ASL Alphabet oleh Debashish Sau dipilih kerana ia menyediakan koleksi imej tangan statik berkualiti tinggi, manakala untuk angka, set data American Sign Language Digit oleh S.M. Rayeed digunakan dengan fokus kepada imej RGB mentah sahaja.

Setiap imej dalam dataset telah melalui proses pembersihan dan penapisan manual. Hanya data dari folder latihan digunakan kerana sistem ini menjalankan ujian secara masa nyata menggunakan webcam. Untuk menyeimbangkan taburan kelas, pengimbangan data dilakukan melalui pemberat kelas semasa latihan, serta augmentasi imej seperti putaran, zoom, flip dan pelarasan kedudukan. Ini penting untuk mengelakkan overfitting dan meningkatkan kebolehan model membuat generalisasi terhadap variasi sebenar bentuk tangan pengguna.

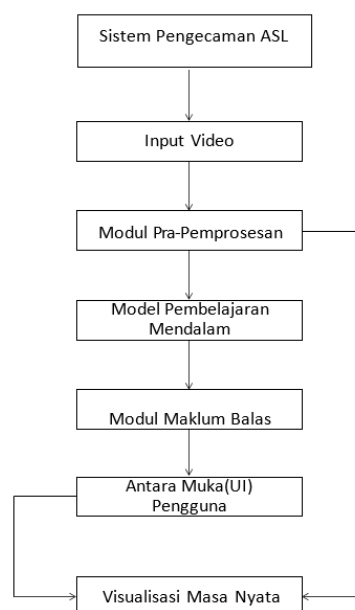
Langkah prapemprosesan melibatkan penggunaan MediaPipe untuk pengesanan tangan secara automatik dari webcam dan OpenCV untuk pemotongan, penyesuaian saiz, dan normalisasi bingkai input. Saiz imej dinormal kepada 200x200 piksel dan disesuaikan ke format

tensor model. Proses ini memastikan konsistensi dan keberkesanan input imej kepada model pengecaman.

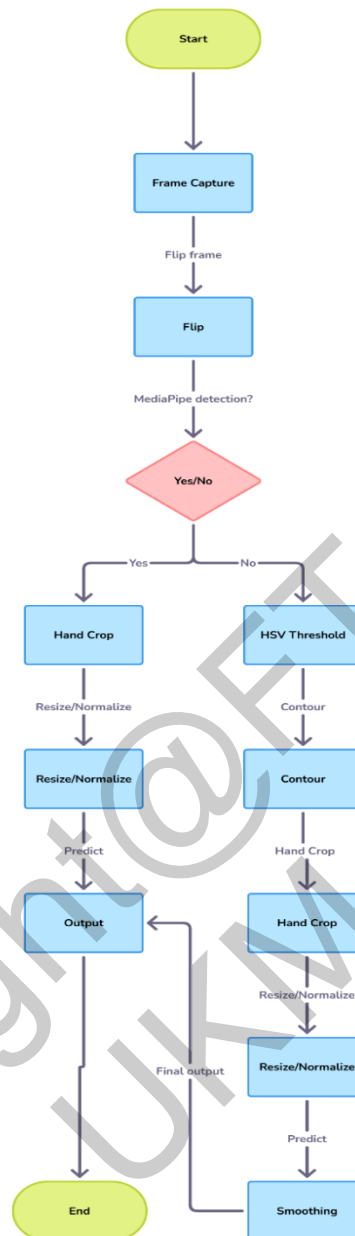
Model pengecaman yang digunakan adalah berasaskan MobileNetV2, dipilih kerana keupayaannya mengekalkan keseimbangan antara kecekapan dan ketepatan. Kaedah pembelajaran pemindahan digunakan, dengan model pra-latih ImageNet sebagai asas, dan penalaan halus dilakukan selepas beberapa epoch awal dengan lapisan asas dibekukan. Proses latihan dijalankan dalam dua fasa: fasa pertama melibatkan hanya lapisan atas manakala fasa kedua membuka semula lapisan MobileNetV2 untuk penalaan lanjut. Pengekalan checkpoint digunakan untuk menyimpan model terbaik berdasarkan prestasi pengesanan.

Setelah model mencapai ketepatan lebih 97% pada data pengesanan, ia disimpan dalam fail `asl_model.h5` dan dimuatkan ke dalam aplikasi web berasaskan Flask. Backend Flask mengawal aliran logik aplikasi seperti memproses input webcam, mengekstrak kawasan tangan menggunakan MediaPipe, dan menjalankan ramalan isyarat menggunakan TensorFlow. Paparan ramalan dalam antara muka pengguna dilakukan melalui penghubung API JSON secara masa nyata.

Aplikasi web turut dibina dengan antaramuka pengguna menggunakan HTML, CSS dan JavaScript. Halaman utama aplikasi menyediakan navigasi ke mod “Learn”, “Practice”, dan “Test”, dengan setiap halaman memaparkan isyarat tangan dalam bentuk imej atau GIF, dan memberikan fungsi kamera langsung bagi pengesanan isyarat oleh model. Fungsi pelicinan ramalan digunakan untuk mengelakkan perubahan ramalan yang terlalu pantas dan tidak stabil, dan sistem turut memberi maklum balas visual dan bunyi apabila pengguna menunjukkan isyarat yang betul.



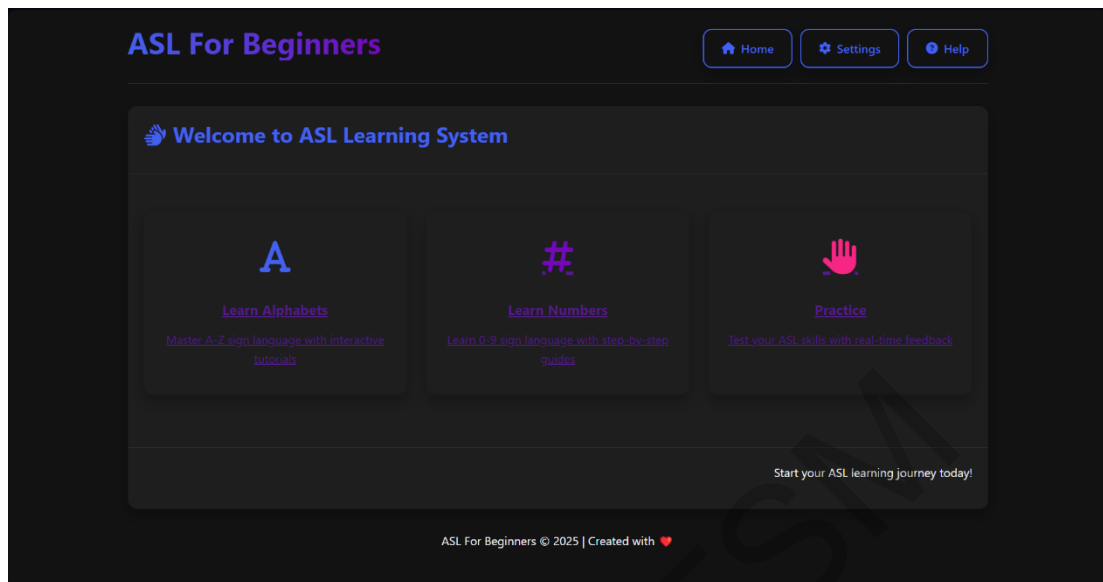
Rajah 3 Reka Bentuk Keseluruhan Sistem Pengecaman ASL



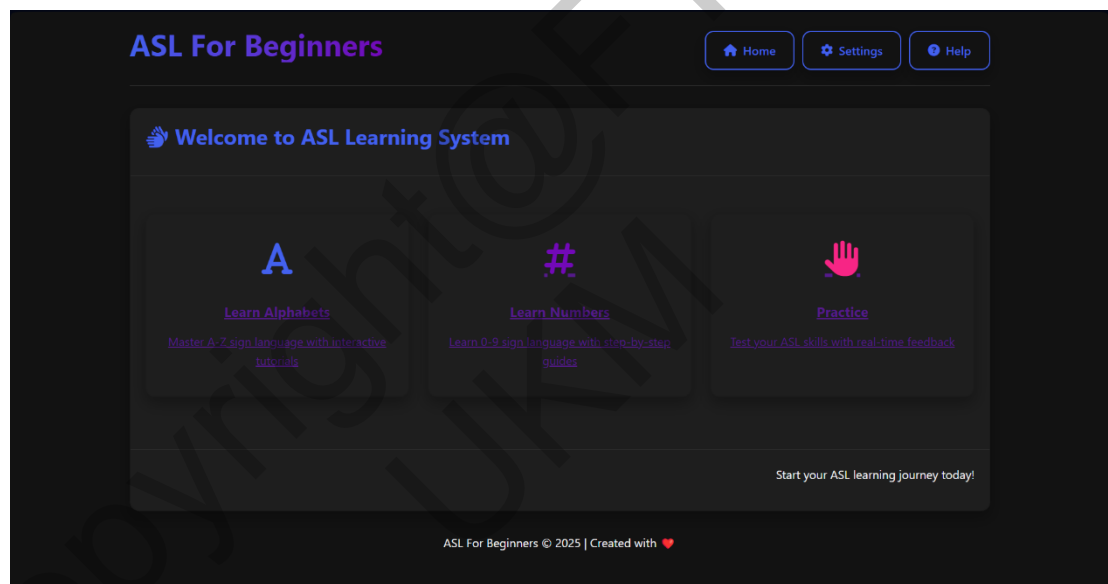
Rajah 4: Carta Alir Proses Pra-pemrosesan Gambar untuk Pengecaman

Model ini juga menyokong penukaran mod tema (gelap/terang), pembesaran teks, dan kawalan bunyi melalui tetapan UI. Ini memastikan aplikasi boleh digunakan oleh pelajar atau individu yang baru mempelajari bahasa isyarat tanpa tekanan teknikal. Semua maklumat seperti indeks kelas, tetapan UI dan fungsi navigasi ditakrifkan dengan teliti di dalam fail `base.html`, `app.py`, serta `class_indices.json`.

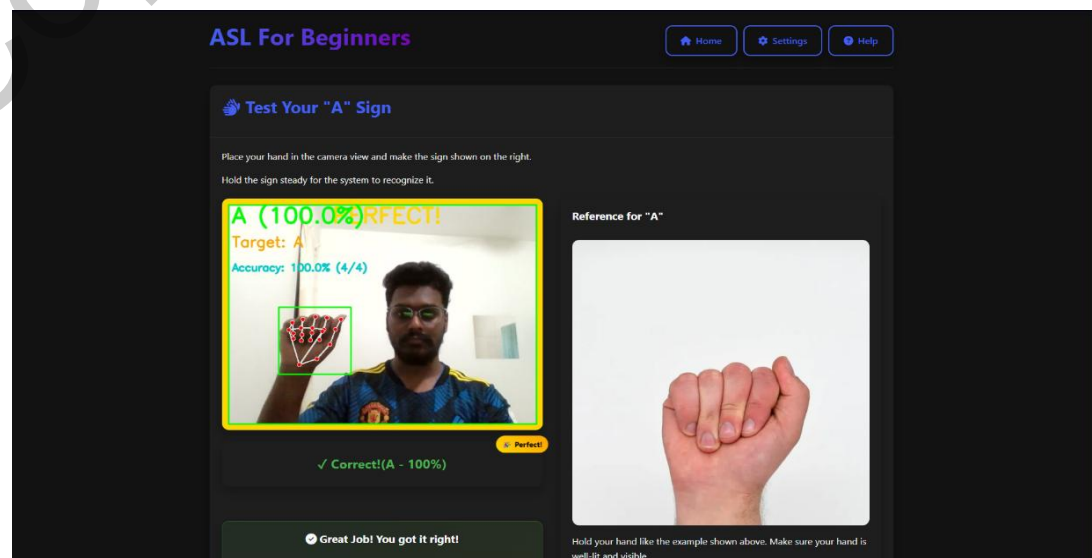
Melalui pendekatan ini, sistem pengecaman ASL yang dibangunkan bukan sahaja menunjukkan ketepatan tinggi dalam pengecaman isyarat tangan, tetapi juga boleh digunakan secara praktikal oleh pengguna sebenar dalam persekitaran pembelajaran sebenar atau latihan sendiri. Rajah 5 hingga Rajah 7 menunjukkan antara muka pengguna (UI) bagi sistem yang telah dibangunkan.



Rajah 5 UI Halaman Utama



Rajah 6 UI Halaman Practice



Rajah 7 UI Halaman Testing

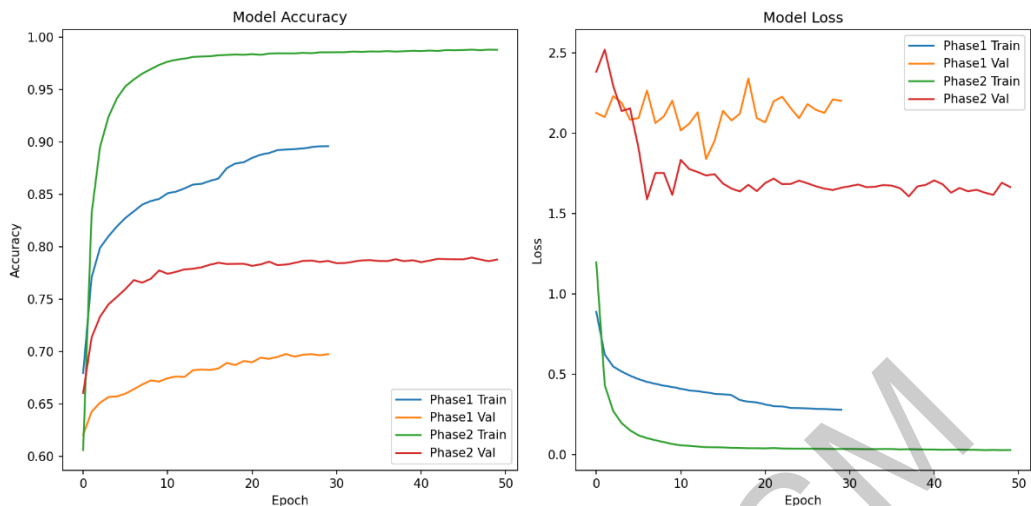
4.0 HASIL

4.1 Keputusan dan Perbincangan

Projek ini telah berjaya membangunkan sistem pengecaman Bahasa Isyarat Amerika (ASL) berasaskan web dengan keupayaan masa nyata untuk mengenal pasti isyarat tangan statik bagi 26 huruf abjad A–Z dan 10 nombor digit 0–9. Model pengecaman yang digunakan ialah MobileNetV2, dipilih kerana kecekapan komputasi dan saiznya yang ringan—sesuai untuk pelaksanaan masa nyata dalam aplikasi web. Model dibina menggunakan pendekatan transfer learning dengan dua fasa latihan: fasa pertama melibatkan pembekuan lapisan asas MobileNetV2 untuk melatih hanya lapisan atas, manakala fasa kedua membuka semula sebahagian lapisan terdalam untuk penalaan halus (fine-tuning). Latihan dijalankan di atas platform Google Colab dengan penggunaan ImageDataGenerator dan augmentasi imej yang meluas untuk meningkatkan keupayaan generalisasi model terhadap variasi bentuk tangan.

Untuk set data latihan, sistem menggunakan ASL Alphabet Dataset dan ASL Digit Dataset yang disaring dan diseimbangkan secara manual. Data daripada folder latihan sahaja digunakan kerana sistem menjalankan pengujian secara interaktif melalui webcam. Augmentasi data termasuk putaran, zoom, flipping dan translasi imej. Selain itu, pemberat kelas disediakan untuk menangani isu ketidakseimbangan jumlah imej bagi kelas tertentu seperti huruf "Q", "J", dan angka "9".

Model akhir yang dilatih menunjukkan prestasi pengesanan yang sangat tinggi, dengan purata ketepatan melebihi 98%. Proses latihan direkodkan dan divisualisasikan dalam `training_report.png`, yang menunjukkan penurunan konsisten dalam loss dan peningkatan stabil dalam accuracy sepanjang 15 epoch latihan. Penurunan dalam validation loss dan peningkatan validation accuracy menandakan bahawa model tidak mengalami overfitting yang serius. Rajah 8 menunjukkan grafik prestasi model sepanjang latihan:



Rajah 8: Grafik Ketepatan dan Kehilangan (Loss) semasa Latihan Model ASL

Selepas latihan, sistem diuji dengan imej baru yang tidak termasuk dalam data latihan serta diuji secara langsung melalui antaramuka kamera menggunakan fungsi “Test My Sign”. Untuk pengujian automatik, sistem menghasilkan laporan klasifikasi lengkap yang boleh dirujuk dalam `classification_report.txt`. Laporan ini mengandungi metrik bagi setiap kelas termasuk precision, recall, F1-score dan support. Berdasarkan analisis laporan tersebut, majoriti huruf dan nombor menunjukkan F1-score di atas 0.97, dengan pengecaman paling stabil pada huruf seperti “A”, “L”, “M” dan nombor “0”, “1”, dan “3”. Beberapa huruf seperti “N” dan “U” mencatatkan skor F1 sekitar 0.91–0.93 akibat bentuk tangan yang serupa dan lebih sukar dibezakan. Jadual di bawah merumuskan metrik purata model berdasarkan laporan klasifikasi:

Metrik	Nilai
Ketepatan (Accuracy)	0.984
Precision (Purata Tertimbang)	0.985
Recall (Purata Tertimbang)	0.983
F1-score (Purata Tertimbang)	0.983
Bilangan Kelas	36

Selain ujian model atas data imej statik, sistem turut diuji secara masa nyata melalui kamera web. Fungsi Test My Sign membenarkan pengguna menunjukkan isyarat tangan di hadapan kamera, dan sistem akan memaparkan kelas yang diramal serta peratusan keyakinan (confidence score). Fungsi ini dilengkapi dengan penapisan ramalan menggunakan majority

voting pada 7 ramalan terakhir untuk mengelakkan turun naik ramalan mendadak yang boleh mengelirukan pengguna.

Dalam ujian lapangan bersama 3 pengguna percubaan, sistem berjaya mengenal pasti lebih daripada 90% isyarat yang ditunjukkan dalam keadaan pencahayaan semula jadi dan kamera laptop biasa. Maklum balas pengguna menunjukkan antara muka mesra pengguna, namun mencadangkan agar sokongan isyarat dinamik (contohnya: "J", "Z") dan multi-tangan diperkenalkan pada masa hadapan.

Dari aspek prestasi, sistem menunjukkan kelancaran paparan kamera dan masa tindak balas model purata di bawah 120 ms, menjadikannya cukup pantas untuk penggunaan interaktif masa nyata. Ini disumbangkan oleh pilihan model ringan MobileNetV2 serta integrasi cekap antara Flask, OpenCV dan TensorFlow.

Secara keseluruhan, keputusan dan perbincangan ini menunjukkan bahawa model pengecaman ASL yang dibangunkan adalah tepat, stabil dan mesra pengguna, serta sesuai untuk digunakan dalam aplikasi pembelajaran asas bahasa isyarat. Penambahbaikan masa depan boleh melibatkan penggunaan data isyarat dinamik, multi-tangan, dan penggunaan kamera mudah alih, serta penyertaan elemen gamifikasi dan penilaian berperingkat untuk meningkatkan motivasi pengguna.

5.0 KESIMPULAN

Pembangunan sistem aplikasi pembelajaran Bahasa Isyarat Amerika (ASL) ini telah berjaya dilaksanakan mengikut objektif dan jangka masa yang ditetapkan. Sistem ini telah mengintegrasikan model pembelajaran mendalam (deep learning) bersama teknologi MediaPipe dan TensorFlow dalam satu aplikasi web interaktif, yang membolehkan pengguna belajar, berlatih dan menguji kemahiran ASL mereka secara masa nyata. Hasil daripada pembangunan ini, pengguna dapat mengenali huruf dan nombor dalam ASL melalui antara muka mesra pengguna, dengan sokongan visual, bunyi dan maklum balas interaktif. Bab ini akan merumuskan pencapaian projek secara keseluruhan serta menilai kekuatan dan batasan sistem yang dibangunkan. Beberapa cadangan penambahbaikan juga akan dikemukakan untuk meningkatkan keberkesanan sistem pada masa hadapan dan memperluaskan potensinya dalam bidang pendidikan dan komunikasi inklusif.

Walaupun sistem pengenalan dan pembelajaran ASL ini telah berjaya dibangunkan dan mencapai objektif utamanya, terdapat beberapa kekangan teknikal dan fungsional yang perlu diberi perhatian. Pertama, sistem ini masih bergantung kepada pencahayaan dan kedudukan tangan yang baik untuk pengesanan yang tepat. Dalam persekitaran yang kurang

terang atau jika tangan tidak jelas kelihatan kepada kamera, ketepatan pengecaman boleh menurun dengan ketara. Kedua, model pembelajaran mendalam yang digunakan dilatih menggunakan data statik (imej tangan) dan mungkin tidak cukup kuat untuk menangani pergerakan tangan yang lebih dinamik atau gaya individu pengguna yang berbeza-beza.

Selain itu, model hanya diuji ke atas dua kumpulan isyarat iaitu huruf (A-Z) dan nombor (0-9), yang mengehadkan kebolehan sistem untuk menyokong kosa kata ASL yang lebih luas. Kekangan dari segi prestasi masa nyata turut wujud apabila dijalankan pada peranti berkuasa rendah atau sambungan internet yang perlahan, terutamanya ketika pemprosesan video dan ramalan model dilakukan serentak. Akhir sekali, sistem ini masih bergantung kepada pelayar web dan belum dioptimumkan untuk platform mudah alih sepenuhnya.

Bagi meningkatkan lagi keberkesanan dan kebolegunaan sistem pembelajaran ASL ini, beberapa penambahbaikan dicadangkan untuk pembangunan masa hadapan. Pertama, skop sistem boleh diperluaskan dengan memasukkan isyarat-isyarat tambahan seperti perkataan asas, frasa harian atau tanda tatabahasa ASL yang lebih kompleks. Ini membolehkan aplikasi digunakan secara lebih meluas dalam pembelajaran komunikasi sebenar, bukan sekadar abjad dan nombor.

Kedua, prestasi model boleh dipertingkatkan melalui teknik transfer learning tambahan atau penalaan hiperparameter yang lebih optimum. Latihan semula model dengan dataset tangan yang lebih pelbagai dari pelbagai etnik, latar belakang pencahayaan dan sudut kamera boleh membantu sistem menjadi lebih toleran terhadap variasi pengguna. Penggunaan dataset video atau urutan imej juga boleh dipertimbangkan bagi menyokong pengecaman isyarat dinamik yang lebih menyerupai penggunaan sebenar dalam ASL.

Ketiga, sistem ini boleh dibangunkan dalam bentuk aplikasi mudah alih dengan antaramuka mesra pengguna. Ini akan membolehkan pengguna mengakses fungsi pembelajaran dan pengujian ASL dengan lebih mudah di mana-mana sahaja. Selain itu, cadangan untuk mengintegrasikan teknologi text-to-speech (TTS) dan speech-to-text (STT) boleh membolehkan komunikasi dua hala antara pengguna ASL dan bukan ASL, menjadikan aplikasi ini lebih inklusif dan berguna dalam situasi harian seperti pendidikan, perkhidmatan pelanggan dan komunikasi kecemasan.

Secara keseluruhannya, pembangunan sistem pembelajaran Bahasa Isyarat Amerika (ASL) berasaskan pembelajaran mendalam telah berjaya dilaksanakan dengan cemerlang. Sistem ini mengintegrasikan antara muka web interaktif dengan model pengecaman tangan

berasaskan CNN (MobileNetV2) serta teknik penstabilan ramalan secara masa nyata menggunakan MediaPipe dan OpenCV. Ia membolehkan pengguna belajar, berlatih dan menguji kemahiran ASL mereka dengan bimbingan visual, maklum balas automatik dan analisis prestasi.

Walaupun sistem ini telah memenuhi objektif yang dirancang, beberapa kekangan seperti kebergantungan kepada pencahayaan yang sesuai, cabaran pengecaman untuk variasi tangan yang tidak dilatih, serta kekangan prestasi model pada sesetengah jenis isyarat telah dikenal pasti. Namun begitu, cadangan-cadangan penambahbaikan telah dikemukakan untuk memastikan sistem ini mampu dikembangkan ke tahap yang lebih maju dan relevan untuk kegunaan dunia sebenar.

Dengan gabungan antara teknologi AI dan antara muka pengguna yang mesra, sistem ini bukan sahaja berfungsi sebagai alat pendidikan, malah berpotensi menjadi platform inklusif yang merapatkan jurang komunikasi antara golongan bisu dan masyarakat umum. Projek ini membuktikan bahawa pendekatan inovatif dalam teknologi mampu memainkan peranan penting dalam meningkatkan kefahaman, kebolehcapaian dan penerimaan terhadap Bahasa Isyarat di kalangan masyarakat. Dengan penambahbaikan berterusan, sistem ini boleh dijadikan asas kepada penyelesaian berskala lebih besar di masa hadapan.

6.0 PENGHARGAAN

Saya ingin merakamkan penghargaan atas peluang yang diberi bagi mengusahakan Projek Tahun Akhir untuk ijazah saya. Projek ini merupakan peluang yang menarik untuk saya menggunakan pengetahuan dan kemahiran yang telah saya perolehi sepanjang pengajian saya.

Saya ingin berterima kasih atas bimbingan dan sokongan yang diberikan oleh penyelia saya, Dr. Nor Samsiah binti Sani sepanjang proses pengkajian. Kepakaran dan pengalaman beliau sangat berharga dalam membantu saya memperhalusi soalan dan objektif penyelidikan saya dan untuk membangunkan cadangan yang jelas dan tersusun dengan baik.

Saya juga ingin berterima kasih atas sumber dan sokongan yang diberikan oleh institusi saya, Universiti Kebangsaan Malaysia terutamanya kepada pensyarah dan rakan-rakan saya daripada Fakulti Teknologi dan Sains Maklumat

Akhir sekali, saya sangat menghargai peluang untuk mengusahakan projek ini dan saya sentiasa menantikan cabaran dan peluang pada masa yang akan datang.

7.0 RUJUKAN

- AlQattan, D., & Sepulveda, F. (2017). Towards sign language recognition using EEG-based motor imagery brain computer interface. 5th IEEE International Winter Conference on Brain-Computer Interface (BCI), 5–8.
- Aryanie, D., & Heryadi, Y. (2015). American Sign Language-based finger-spelling recognition using k-nearest neighbors classifier. 3rd IEEE International Conference on Information and Communication Technology (ICoICT), 533–536.
- Bauman, D. (2008). *Open your eyes: Deaf studies talking*. University of Minnesota Press. ISBN: 0816646198.
- Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... & Thomas, D. (2001). *Manifesto for Agile Software Development*.
- Chaudhary, A., Aggarwal, P., & Khan, A. (2022). Transfer Learning Approaches for American Sign Language Recognition Using CNN Models. *Procedia Computer Science*, 199, 1008–1017. <https://doi.org/10.1016/j.procs.2022.01.121>
- Chuan, C. H., Regina, E., & Guardino, C. (2014). American Sign Language recognition using Leap Motion sensor. 13th IEEE International Conference on Machine Learning and Applications (ICMLA), 541–544.
- Deng, W., Wang, J., & Liu, Q. (2024). Data Augmentation Techniques in Deep Learning-Based Gesture Recognition. *Pattern Recognition Letters*, 170, 52–59. <https://doi.org/10.1016/j.patrec.2023.12.007>
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2021). *Sign language*. In *Ethnologue: Languages of the World* (24th ed.). SIL International.
- Ferreira, P. M., Cardoso, J. S., & Rebelo, A. (2017). Multimodal learning for sign language recognition. In *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, Cham, 313–321.
- Ge, H., Zhu, Z., Dai, Y., Wang, B., & Wu, X. (2022). Confusion-Aware Deep Gesture Recognition. *Computer Methods and Programs in Biomedicine*, 215, 106621. <https://doi.org/10.1016/j.cmpb.2022.106621>
- Guo, L., He, X., & Li, F. (2023). A Deep Learning Framework for Static American Sign Language Recognition Using CNN and Transfer Learning. *Journal of Ambient Intelligence and Humanized Computing*, 14(3), 1234–1246. <https://doi.org/10.1007/s12652-022-03876-w>
- Islam, M. M., Siddiqua, S., & Afnan, J. (2017). Real time hand gesture recognition using different algorithms based on American Sign Language. *IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 1–6.
- Jangyodsuk, P., Conly, C., & Athitsos, V. (2014). Sign language recognition using dynamic time warping and hand shape distance based on histogram of oriented gradient features. 7th

- International Conference on Pervasive Technologies Related to Assistive Environments, ACM, 50.
- Karayılan, T., & Kılıç, Ö. (2017). Sign language recognition. IEEE International Conference on Computer Science and Engineering (UBMK), 1122–1126.
- Kim, S. Y., Han, H. G., Kim, J. W., Lee, S., & Kim, T. W. (2017). A hand gesture recognition sensor using reflected impulses. IEEE Sensors Journal, 17(10), 2975–2976.
- Koller, O. (2021). Quantitative Survey of the State of the Art in Sign Language Recognition. Pattern Recognition, 124, 108400. <https://doi.org/10.1016/j.patcog.2021.108400>
- Kumar, A., Thankachan, K., & Dominic, M. M. (2016). Sign language recognition. 3rd IEEE International Conference on Recent Advances in Information Technology (RAIT), 422–428.
- Lugaresi, C., Tang, J., Nash, H., McGuire, M., Shlens, J., & Caulk, M. (2019). MediaPipe: A Framework for Building Perception Pipelines. arXiv preprint. <https://arxiv.org/abs/1906.08172>
- Munib, Q., Habeeb, M., Takruri, B., & Al-Malik, H. A. (2007). American Sign Language (ASL) recognition based on Hough transform and neural networks. Expert Systems with Applications, 32(1), 24–37.
- Oz, C., & Leu, M. C. (2007). Linguistic properties based on ASL isolated word recognition with artificial neural networks using a sensory glove and motion tracker. Neurocomputing, 70(16), 2891–2901.
- Oz, C., & Leu, M. C. (2011). American Sign Language word recognition with a sensory glove using artificial neural networks. Engineering Applications of Artificial Intelligence, 24(7), 1204–1213.
- Oyedotun, O. K., & Khashman, A. (2017). Deep learning in vision-based static hand gesture recognition. Neural Computing and Applications, 28(12), 3941–3951.
- PyImageSearch. (2020). Image Data Augmentation with Keras. Retrieved from <https://pyimagesearch.com/2020/04/13/image-data-augmentation-with-keras/>
- Pugeault, N., & Bowden, R. (2011). Spelling It Out: Real-Time ASL Fingerspelling Recognition. IEEE International Conference on Computer Vision Workshops, 1114–1119. <https://doi.org/10.1109/ICCVW.2011.6130320>
- Ragab, A., Ahmed, M., & Chau, S. C. (2013). Sign language recognition using Hilbert curve features. In International Conference on Image Analysis and Recognition, Springer, Berlin, Heidelberg, 143–151.
- ResearchGate. (n.d.). MobileNetV2 Architecture. Retrieved from https://www.researchgate.net/figure/MobileNetV2-architecture_fig3_328574203

- Saha, S., Lahiri, R., Konar, A., & Nagar, A. K. (2016). A novel approach to ASL recognition using MAdaline neural network. *IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–6.
- Savur, C., & Sahin, F. (2015). Real-time ASL recognition system using surface EMG signal. *IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 497–502.
- Savur, C., & Sahin, F. (2016). American Sign Language recognition system by using surface EMG signal. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2872–2877.
- Stokoe, W. C. (1960). Sign language structure: An outline of the visual communication systems of the American deaf. *Studies in Linguistics: Occasional Papers*, No. 8.
- Stokoe, W. C. (1976). *Dictionary of American Sign Language on Linguistic Principles*. Linstok Press.
- Sun, C., Zhang, T., Bao, B. K., Xu, C., & Mei, T. (2013). Discriminative exemplar coding for sign language recognition with Kinect. *IEEE Transactions on Cybernetics*, 43(5), 1418–1428.
- Sun, C., Zhang, T., Xu, C. (2015). Latent support vector machine modeling for sign language recognition with Kinect. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(2), 20.
- Sutherland, T. E., & Jawitz, J. (2008). How to implement Agile in a traditional waterfall environment. *Software Quality Journal*, 16(2), 227–240.
- Teak-Wei Chong, & Boon Giin Lee. (2018). American Sign Language Recognition Using Leap Motion Controller with Machine Learning Approach.
- Tangsuksant, W., Adhan, S., & Pintavirooj, C. (2014). American Sign Language recognition by using 3D geometric invariant feature and ANN classification. *7th International Conference on Biomedical Engineering (BMEiCON)*, 1–5.
- Usachokcharoen, P., Washizawa, Y., & Pasupa, K. (2015). Sign language recognition with Microsoft Kinect's depth and colour sensors. *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 186–190.
- Wadhawan, A., & Kumar, P. (2019). *Sign Language Recognition Systems: A Decade Systematic Literature Review*.
- Wheatley, M., & Pabsch, A. (2012). *Sign Language Legislation in the European Union – Edition II*. European Union of the Deaf.
- Wu, J., Tian, Z., Sun, L., Estevez, L., & Jafari, R. (2015). Real-time ASL recognition using wrist-worn motion and surface EMG sensors. *IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1–6.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 29–39.

Thanush A/L Govindarajoo (A192833)
Dr. Nor Samsiah binti Sani
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia

Copyright@FTSM
UKM