

STUDI KOMPARATIF XGBOOST DAN MLP BERDASARKAN EVALUASI FITUR PADA KLASIFIKASI ANGKA BAHASA ISYARAT SIBI

Benedict Michael Pepper¹, Gilbert Ronaldo Triswanto²

^{1,2}Teknik Informatika, Universitas Ma Chung, Jl. Villa Puncak Tidar No. 1, Malang, Indonesia, 65151

***Email korespondensi:** 312310007@student.machung.ac.id

Received: 27 05 2026 – Revised: 29 07 2026 - Accepted: 04 08 2026 - Published: 05 08 2026

Abstrak. Sistem pengenalan bahasa isyarat berbasis sarung tangan sensor dikembangkan sebagai alat bantu komunikasi bagi komunitas tunarungu. Meskipun demikian, penentuan kombinasi fitur pada sistem tersebut sering kali dilakukan secara intuitif tanpa validasi empiris. Penelitian ini mengevaluasi validitas pemilihan fitur menggunakan algoritma *ExtraTreesClassifier* pada dataset angka Sistem Isyarat Bahasa Indonesia (SIBI) yang terdiri dari 20 kelas gestur. Hasil analisis menunjukkan bahwa saluran magnetometer, yang dikecualikan dalam penelitian acuan sebelumnya, merupakan fitur paling informatif. Sebaliknya, saluran akselerometer yang dipertahankan dalam studi acuan tersebut berada di peringkat terbawah kepentingan fitur. Selanjutnya, dua model klasifikasi, yaitu *Extreme Gradient Boosting* (XGBoost) dan *Multilayer Perceptron* (MLP) berbasis PyTorch, dioptimalkan menggunakan pencarian hiperparameter Bayesian melalui Optuna dengan 100 percobaan per konfigurasi pada dua himpunan fitur berbeda. Menggunakan seluruh 16 fitur, XGBoost mencapai akurasi uji sebesar 99,83% dan MLP mencapai 99,08%. Kedua hasil tersebut melampaui *baseline k-Nearest Neighbors* (kNN) yang mencatat akurasi 98,90%. Ambiguitas klasifikasi pada Kelas 16 dan 17, yang menjadi kelemahan utama kNN akibat kemiripan karakteristik data, dapat diselesaikan. Nilai *F1-score* kedua kelas tersebut meningkat masing-masing dari 0,97 dan 0,95 menjadi 0,9972 dan 0,9978. Temuan ini menunjukkan bahwa metode ansambel pohon keputusan mengungguli kNN dan MLP pada data sensor tabular. Hasil eksperimen ini juga menunjukkan bahwa tahapan evaluasi kepentingan fitur secara objektif merupakan fondasi penting dalam perancangan sistem klasifikasi gestur.

Kata kunci: flexpoint bend sensor, seleksi fitur, XGBoost, MLP, SIBI

PENDAHULUAN

Hambatan komunikasi antara komunitas tunarungu dan masyarakat umum merupakan tantangan nyata dalam mewujudkan tatanan sosial yang inklusif. Berdasarkan data *Long Form Sensus Penduduk 2020*, Badan Pusat Statistik (BPS) mencatat sekitar 1,82 juta jiwa penyandang tunarungu di Indonesia, atau sekitar 0,36% dari total populasi (BPS, 2024). Komunitas ini berkomunikasi terutama melalui Sistem Isyarat Bahasa Indonesia (SIBI), sistem bahasa isyarat terstandar yang ditetapkan pada tahun 1994 dan diterapkan di Sekolah Luar Biasa Tipe B di seluruh Indonesia (Depdiknas RI, 1994).

Sistem Pengenalan Bahasa Isyarat (*Sign Language Recognition*/SLR) otomatis menawarkan pendekatan teknologi untuk menjembatani kesenjangan komunikasi ini. Sistem SLR menerjemahkan gestur tangan menjadi teks atau ucapan secara langsung,

sehingga memungkinkan interaksi tanpa mengharuskan lawan bicara mempelajari bahasa isyarat. Penerapannya pada perangkat bersumber daya terbatas ditentukan oleh akurasi, latensi inferensi, dan efisiensi komputasi.

A. Pendekatan Perangkat Keras: Sarung Tangan Sensor vs. Visi Komputer

Pengenalan bahasa isyarat secara umum dibagi ke dalam dua paradigma perangkat keras: sistem berbasis kamera atau sensor kedalaman, dan sistem berbasis sarung tangan sensor yang dilengkapi transduser. Sistem berbasis kamera memiliki keterbatasan terkait kondisi pencahayaan, sudut pandang tetap, dan oklusi. Sarung tangan sensor mengatasi keterbatasan tersebut dengan mengukur langsung pergerakan fisik sendi jari dan orientasi tangan. Untuk klasifikasi angka SIBI yang seluruh gesturnya dapat dibedakan dari konfigurasi tangan saja, sarung tangan sensor merupakan pilihan yang tepat (Amin et al., 2022). Ambar et al. (2023) membuktikan bahwa data sensor tekuk dan *Inertial Measurement Unit* (IMU) memberikan informasi pembeda yang saling melengkapi dalam sistem berbasis sarung tangan sensor.

B. Permasalahan Seleksi Fitur pada Penelitian Sebelumnya

Penelitian ini mengembangkan studi milik Widodo (2022) yang menggunakan pengklasifikasi kNN untuk mengklasifikasikan 20 tanda angka SIBI. Perangkat keras yang digunakan berupa sarung tangan sensor dengan lima sensor tekuk *Flexpoint* (10 saluran, x_1 – x_{10}) dan IMU MEMS KMX62 yang menyediakan akselerometer 3-sumbu serta magnetometer 3-sumbu (x_{11} – x_{16}), sehingga menghasilkan total 16 fitur. *Baseline kNN* dilatih pada 11 fitur yang dipilih secara manual, dengan mengecualikan saluran x_1 , x_7 , x_{14} , x_{15} , dan x_{16} tanpa dasar analitis yang eksplisit. Pengklasifikasi tersebut mencapai akurasi validasi 98,90%, namun menunjukkan kelemahan pada Kelas 16 ($F1=0,97$) dan Kelas 17 ($F1=0,95$), dua gestur yang secara morfologis paling berdekatan dalam himpunan angka SIBI. Cooper et al. (2011) mencatat bahwa klasifikasi gestur terisolasi dapat diperlakukan sebagai masalah klasifikasi multikelas standar apabila fitur yang digunakan mampu merepresentasikan data dengan baik, sehingga pemilihan fitur menjadi faktor penentu kinerja sistem.

C. Dasar Pemilihan Model dan Pertanyaan Penelitian

Grinsztajn et al. (2022) menunjukkan secara empiris bahwa pada data tabular, metode pohon keputusan secara konsisten melampaui jaringan saraf tiruan. XGBoost (Chen & Guestrin, 2016) memperluas kerangka ini dengan ekspansi Taylor orde kedua dan regularisasi L1/L2, sedangkan MLP mewakili pendekatan pembelajaran mendalam. Optuna (Akiba et al., 2019) menyediakan optimasi hiperparameter Bayesian berbasis *Tree-Parzen Estimator* (TPE) yang lebih efisien dibanding *grid search* atau *random search*. Penelitian ini menangani dua pertanyaan utama: (1) Apakah XGBoost dan MLP mampu melampaui akurasi *baseline* kNN pada *dataset* sensor sarung tangan yang sama? (2) Apakah pengurangan dari 16 menjadi 11 fitur pada penelitian sebelumnya dapat dibenarkan secara empiris? Kontribusi penelitian meliputi: analisis tingkat kepentingan fitur menggunakan *ExtraTreesClassifier*, studi ablasi terkontrol antara konfigurasi 11 dan 16 fitur, optimasi hiperparameter Bayesian melalui Optuna, serta analisis mendalam per kelas yang menyelesaikan ambiguitas Kelas 16/17.

METODE PELAKSANAAN

A. Perangkat Keras dan Pengumpulan Data

Data dikumpulkan menggunakan sarung tangan sensor dengan dua jenis komponen: (1) lima sensor tekuk Flexpoint yang ditempatkan di permukaan dorsal tangan, satu per jari, menghasilkan dua bacaan per jari pada sendi *Metacarpophalangeal* (MCP) dan *Proximal Interphalangeal* (PIP), untuk total 10 saluran sensor tekuk (x_1 - x_{10}); (2) IMU MEMS KMX62 yang menyediakan akselerometer 3-sumbu (x_{11} - x_{13}) dan magnetometer 3-sumbu (x_{14} - x_{16}). Pemetaan lengkap seluruh 16 fitur disajikan pada Tabel 1.

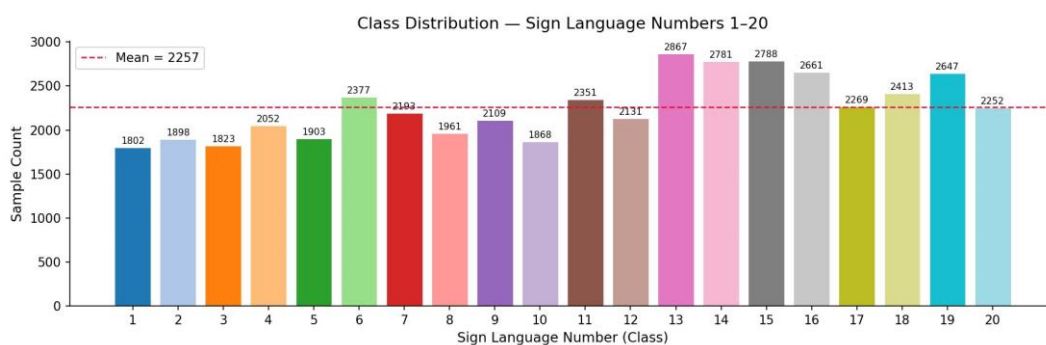
Tabel 1. Pemetaan Fitur Sensor

Fitur	Nama	Anatomi	Jenis	Digunakan di Widodo (2022)
x_1	<i>Pinky</i> MCP	Kelingking - MCP	Sensor Tekuk	Tidak
x_2	<i>Pinky</i> PIP	Kelingking - PIP	Sensor Tekuk	Ya
x_3	<i>Ring</i> MCP	Jari Manis - MCP	Sensor Tekuk	Ya
x_4	<i>Index</i> PIP	Telunjuk - PIP	Sensor Tekuk	Ya
x_5	<i>Middle</i> MCP	Jari Tengah - MCP	Sensor Tekuk	Ya
x_6	<i>Middle</i> PIP	Jari Tengah - PIP	Sensor Tekuk	Ya

Fitur	Nama	Anatomi	Jenis	Digunakan di Widodo (2022)
x7	<i>Index</i> MCP	Telunjuk - MCP	Sensor Tekuk	Tidak
x8	<i>Thumb</i> MCP	Ibu Jari - MCP	Sensor Tekuk	Ya
x9	<i>Ring</i> PIP	Jari Manis - PIP	Sensor Tekuk	Ya
x10	<i>Thumb</i> PIP	Ibu Jari - PIP	Sensor Tekuk	Ya
x11	Akseleromete r X	IMU - Sumbu X	Akseleromete r	Ya
x12	Akseleromete r Y	IMU - Sumbu Y	Akseleromete r	Ya
x13	Akseleromete r Z	IMU - Sumbu Z	Akseleromete r	Ya
x14	Magnetomete r X	IMU - Sumbu X	Magnetomete r	Tidak
x15	Magnetomete r Y	IMU - Sumbu Y	Magnetomete r	Tidak
x16	Magnetomete r Z	IMU - Sumbu Z	Magnetomete r	Tidak

B. Karakteristik *Dataset* dan Praproses

Dataset terdiri dari 45.147 sampel yang terdistribusi pada 20 kelas (angka SIBI 1-20). Audit integritas mengonfirmasi tidak adanya nilai kosong dan baris duplikat, dengan distribusi kelas hampir seragam antara 1.802 hingga 2.867 sampel per kelas (rata-rata 2.257 sampel). Gambar 1 menampilkan distribusi kelas tersebut.



Gambar 1. Distribusi kelas angka bahasa isyarat SIBI 1-20 (rata-rata = 2.257 sampel/kelas)

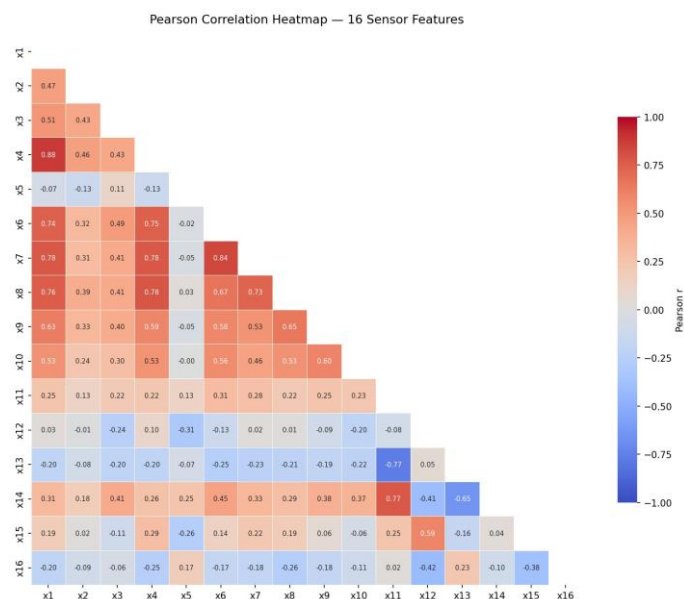
Dataset dipartisi menggunakan pemisahan acak terstratifikasi (*random_state=42*). Tabel 2 merangkum ukuran setiap partisi. *StandardScaler* diterapkan khusus pada masukan MLP, disesuaikan pada data latih kemudian diterapkan pada data validasi dan uji, untuk mencegah kebocoran data. XGBoost tidak memerlukan penskalaan karena sifatnya yang invarian terhadap transformasi monoton.

Tabel 2. Ringkasan Partisi Dataset

Partisi	Sampel	Persentase	Tujuan
Latih	27.087	60%	Pelatihan model
Validasi	9.029	20%	Seleksi hiperparameter dan early stopping
Uji	9.030	20%	Evaluasi akhir (tidak diakses selama pengembangan)

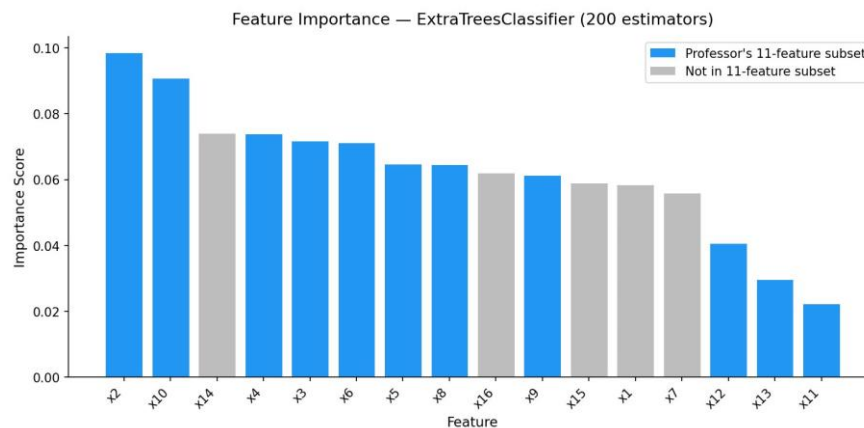
C. Analisis Data Eksploratif

Peta panas korelasi Pearson pada seluruh 16 fitur dihitung untuk mengidentifikasi redundansi linear. Sebagian besar pasangan fitur menunjukkan korelasi rendah. Satu pengecualian menonjol adalah pasangan (x_1, x_4), yaitu *Pinky MCP* dan *Index PIP*, dengan nilai $r = 0,88$. Fitur yang berkorelasi tinggi tetap dapat membawa informasi pembeda yang saling melengkapi pada model non-linear. Gambar 2 menampilkan peta panas korelasi lengkap.



Gambar 2. Peta panas korelasi Pearson pada 16 fitur sensor

Untuk mengevaluasi kontribusi prediktif setiap fitur secara objektif, *ExtraTreesClassifier* dengan 200 estimator dilatih pada *dataset* 16 fitur penuh, kemudian penurunan ketidakmurnian rata-rata (*Gini importance*) diekstraksi dari setiap saluran. Sebagaimana ditunjukkan pada Gambar 3, sumbu magnetometer (x_{14} , x_{15} , x_{16}) memiliki kepentingan tertinggi di antara seluruh fitur, sementara sumbu akselerometer yang dipertahankan dalam 11 fitur *baseline* justru berada di peringkat terbawah. Temuan ini bertentangan dengan dasar pemilihan fitur pada penelitian acuan.



Gambar 3. Tingkat kepentingan fitur via *ExtraTreesClassifier* (200 estimator) pada 16 fitur

Diagram kotak semua 16 fitur, dikelompokkan berdasarkan kelas, dihasilkan untuk memeriksa pemisahan bacaan sensor pada 20 gestur. Gambar 4 menampilkan distribusi per kelas. Fitur seperti x_5 (*Middle Finger MCP*) dan x_2 (*Pinky PIP*) menunjukkan distribusi multimodal yang jelas. Fitur magnetometer menunjukkan distribusi yang lebih tumpang tindih antar kelas, namun kemampuannya mengkodekan orientasi tangan pada bidang horizontal memberikan informasi pembeda yang tidak dimiliki akselerometer.

D. Desain Eksperimen

Dua eksperimen paralel independen dilaksanakan. Eksperimen A: kedua model XGBoost dan MLP dilatih pada subset 11 fitur (x_2 , x_3 , x_4 , x_5 , x_6 , x_8 , x_9 , x_{10} , x_{11} , x_{12} , x_{13}) yang mereplikasi seleksi fitur asli, guna membangun perbandingan terkontrol terhadap *baseline* kNN. Eksperimen B: kedua model dilatih ulang menggunakan himpunan 16 fitur penuh, menambahkan lima saluran yang sebelumnya dikecualikan (x_1 , x_7 , x_{14} , x_{15} , x_{16}). Dalam setiap eksperimen, fase *baseline* dengan hiperparameter bawaan dan fase penyetelan melalui Optuna dilaksanakan secara berurutan.

E. Arsitektur MLP

MLP *fully-connected* diimplementasikan menggunakan *PyTorch* v2.0+. Arsitektur *baseline* terdiri dari lapisan masukan (D neuron, dengan $D = 11$ atau 16 sesuai eksperimen), tiga lapisan tersembunyi berukuran 256, 128, dan 64 neuron (setiap lapisan dilengkapi *BatchNorm*, *ReLU*, dan *Dropout* $p=0,3$), serta lapisan keluaran 20 neuron. Pelatihan menggunakan optimizer *Adam* ($lr=10^{-3}$), *CrossEntropyLoss*, ukuran *mini-batch* 64, dan *early stopping* dengan kesabaran 15 *epoch*. *Baseline* berkonvergensi pada *epoch* 79 (dari maksimum 100) dengan akurasi validasi 98,50%. Setelah optimasi Optuna, konfigurasi optimal pada 16 fitur menggunakan urutan lapisan tersembunyi [128, 64, 256] dengan *dropout* $p=0,3$.

F. Arsitektur XGBoost

XGBoost membangun ansambel sekuensial pohon keputusan. Pada setiap ronde *boosting*, pohon baru disesuaikan terhadap gradien negatif (*pseudo-residual*) dari fungsi kerugian ansambel yang ada, sehingga kapasitas terkonsentrasi pada sampel yang salah diklasifikasikan. Tabel 3 menyajikan konfigurasi hiperparameter *baseline* yang digunakan.

Tabel 3. Konfigurasi hiperparameter baseline XGBoost

Parameter	Nilai	Alasan
<i>n_estimators</i>	300	Jumlah pohon sekuensial dalam ansambel
<i>learning_rate</i>	0,1	Memperkecil kontribusi setiap pohon untuk generalisasi lebih baik
<i>max_depth</i>	6	Kedalaman maksimum per pohon
<i>subsample</i>	0,8	Pengambilan sampel baris stokastik per pohon untuk mengurangi varian
<i>colsample_bytree</i>	0,8	Pengambilan sampel fitur stokastik per pohon
<i>objective</i>	<i>multi:softprob</i>	Keluaran multikelas probabilistik
<i>eval_metric</i>	<i>merror</i>	Laju kesalahan klasifikasi multikelas
<i>tree_method</i>	<i>hist</i>	Algoritma perkiraan berbasis histogram

G. Optimasi Hiperparameter Bayesian

Sampler TPE Optuna menjalankan 100 percobaan per kombinasi arsitektur dan himpunan fitur. Tabel 4 menyajikan ruang pencarian lengkap. Riwayat optimasi untuk XGBoost dan MLP ditampilkan pada Gambar 4(a) dan Gambar 4(b).

Tabel 4. Ruang pencarian hiperparameter Optuna

Model	Hiperparameter	Rentang	Skala
XGBoost	<i>n_estimators</i>	[50, 200]	Integer
XGBoost	<i>learning_rate</i>	[0,05; 0,20]	Log-uniform
XGBoost	<i>max_depth</i>	[3, 6]	Integer
XGBoost	<i>subsample</i>	[0,6; 1,0]	Uniform
XGBoost	<i>colsample_bytree</i>	[0,6; 1,0]	Uniform
MLP	<i>learning_rate</i>	[10 ⁻⁴ ; 5×10 ⁻³]	Log-uniform
MLP	<i>dropout_rate</i>	[0,1; 0,4]	Uniform
MLP	<i>hidden_layer_config</i>	{[256,128,64], [128,64,32], [128,64,256], [256,64,128]}	Kategorikal
MLP	<i>batch_size</i>	{32, 64, 128}	Kategorikal



(a)



(b)

Gambar 4. Riwayat optimasi Optuna: (a) XGBoost dan (b) MLP (100 percobaan; sumbu Y: akurasi validasi)

H. Protokol Evaluasi

Seluruh metrik kinerja akhir berasal dari himpunan uji yang disimpan (9.030 sampel), yang tidak diakses dalam fase pengembangan model mana pun. Metrik yang dihitung meliputi akurasi global, *Precision/Recall/F1-Score* per kelas untuk semua 20 kelas, *Macro-Averaged F1*, dan Matriks Konfusi Ternormalisasi (baris dinormalisasi berdasarkan total sampel kelas sebenarnya).

HASIL DAN PEMBAHASAN

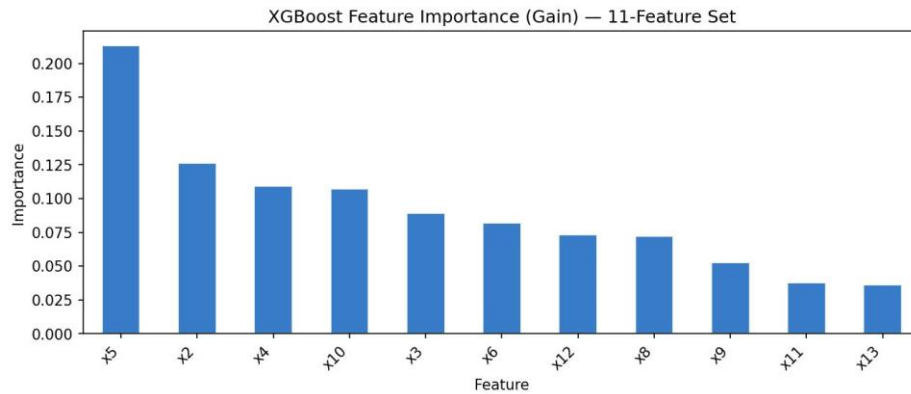
A. Kinerja *Baseline* (11 Fitur, Eksperimen A)

Baseline XGBoost dengan parameter bawaan dilatih dalam waktu sekitar 11,5 detik pada CPU. MLP *baseline* mencapai akurasi validasi 98,50%, di bawah referensi kNN; hal ini sejalan dengan Grinsztajn et al. (2022) yang menunjukkan bahwa MLP pada data tabular umumnya kurang optimal tanpa penyetelan hiperparameter. Model MLP dilatih selama 94 *epoch* dengan *early stopping* (terbaik pada *epoch* 79) dalam waktu 257 detik pada CPU. Tabel 5 merangkum perbandingan.

Tabel 5. Perbandingan hasil *baseline* pada 11 fitur (Eksperimen A)

Model	Akurasi Validasi	F1 Kelas 16	F1 Kelas 17
kNN (Widodo, 2022)	98,90%	0,9700	0,9500
XGBoost <i>baseline</i>	99,53%	0,9898	0,9878
MLP <i>baseline</i>	98,50%	0,9619	0,9569

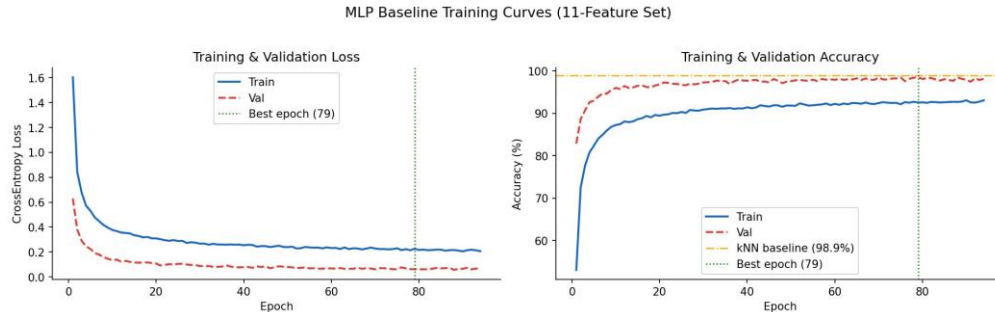
Tingkat kepentingan fitur XGBoost *baseline* pada 11 fitur (Gambar 5) mengonfirmasi temuan eksploratif: *x5* (*Middle Finger MCP*) muncul sebagai fitur paling penting (nilai kepentingan 0,213), diikuti *x2* (*Pinky PIP*, 0,126). Sebaliknya, *x11* (Akselerometer *X*) dan *x13* (Akselerometer *Z*) menempati peringkat terbawah dengan nilai masing-masing 0,038 dan 0,036, menunjukkan bahwa saluran yang dipertahankan dalam *baseline* dengan mengorbankan magnetometer ternyata memberikan informasi pembeda paling sedikit.



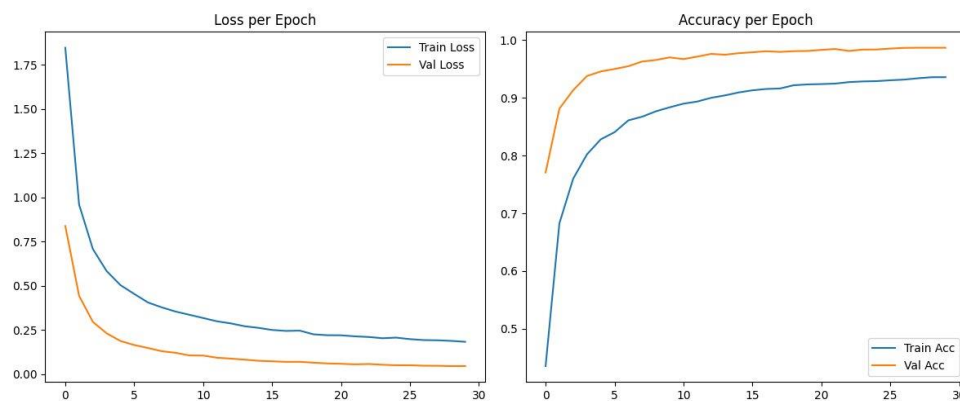
Gambar 5. Tingkat kepentingan fitur XGBoost (Gini Gain) pada himpunan 11 fitur

B. Dinamika Pelatihan MLP

Kurva pelatihan *baseline* MLP (Gambar 6) menunjukkan konvergensi yang baik: pembelajaran cepat dalam 20 *epoch* pertama diikuti penyempurnaan bertahap. Selisih antara *loss* latih dan validasi tetap kecil sepanjang pelatihan, yang mengindikasikan efektivitas regularisasi *dropout* dalam mencegah *overfitting*. Gambar 7 menampilkan kurva MLP yang telah dioptimalkan pada 16 fitur, yang konvergen lebih cepat dan mencapai akurasi validasi lebih tinggi.



Gambar 6. Kurva pelatihan MLP *baseline* (11 fitur): *loss* dan akurasi per *epoch*



Gambar 7. Kurva pelatihan MLP yang dioptimalkan (16 fitur): *loss* dan akurasi per *epoch*

C. Perbandingan Lengkap Setelah Penyetelan

Setelah optimasi Optuna, seluruh kombinasi model dan himpunan fitur dievaluasi pada himpunan uji, sebagaimana dirangkum pada Tabel 6. XGBoost pada 16 fitur penuh mencapai akurasi terbaik 99,83%, melampaui *baseline* kNN sebesar 0,93 poin persentase. MLP yang dioptimalkan mencapai 99,08%.

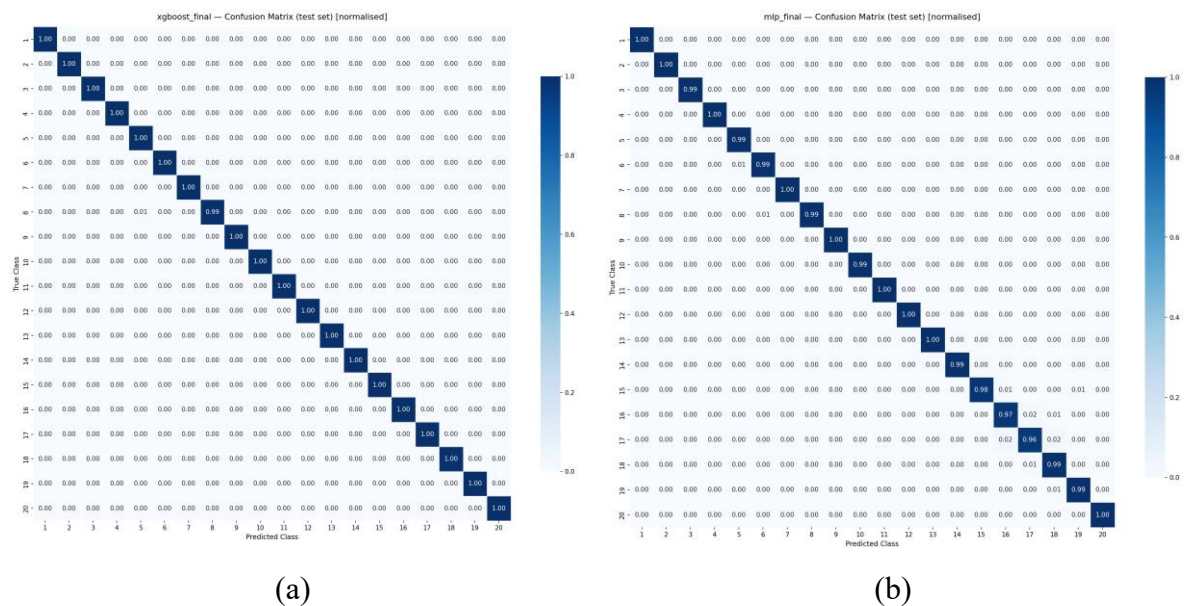
Tabel 6. Hasil semua model dan himpunan fitur pada himpunan uji

Model	Himpunan Fitur	Akurasi Uji	F1 Kelas 16	F1 Kelas 17
kNN (Widodo, 2022)	11 fitur	98,90%*	0,9700	0,9500
XGBoost (dioptimalkan)	11 fitur	99,60%	0,9898	0,9878
XGBoost (dioptimalkan)	16 fitur	99,83%	0,9972	0,9978
MLP <i>baseline</i>	11 fitur	97,90%	0,9619	0,9569
MLP (dioptimalkan)	16 fitur	99,08%	0,9662	0,9604

* Diambil dari Widodo (2022); skema evaluasi tidak identik dengan penelitian ini.

D. Analisis Matriks Konfusi

Gambar 8 menyajikan matriks konfusi ternormalisasi untuk seluruh kombinasi model dan himpunan fitur.



Gambar 8. Matriks konfusi ternormalisasi himpunan uji: (a) XGBoost dioptimalkan 16 fitur dan (b) MLP dioptimalkan 16 fitur

Matriks konfusi XGBoost final pada Gambar 8(a) menampilkan diagonal yang hampir sempurna di seluruh 20 kelas. Matriks MLP final pada Gambar 8(b) mengungkapkan kesalahan klasifikasi minor di antara Kelas 15-17, dengan sekitar 3% hingga 4% sampel data meleset ke sel di luar diagonal utama antara pasangan kelas berdekatan, yang menjelaskan nilai F1 lebih rendah pada kelas-kelas tersebut.

E. Penyelesaian Ambiguitas Kelas 16/17

Dalam SIBI, angka 16 melibatkan jari telunjuk yang menunjuk ke atas dengan ibu jari terbuka ke samping, sedangkan angka 17 menampilkan konfigurasi serupa dengan variasi halus pada tumpang tindih ibu jari dan kemiringan telunjuk ke depan. Gestur-gestur ini menghasilkan bacaan sensor yang menempati wilayah geometris berdekatan dalam ruang fitur 16 dimensi, sehingga pengklasifikasi berbasis jarak Euklid seperti kNN sulit menetapkan batas keputusan yang tegas.

Mekanisme pohon keputusan XGBoost menyelesaikan permasalahan ini dengan mempelajari kondisi pemisahan searah sumbu yang mengisolasi saluran sensor dan ambang nilai tertentu. Mekanisme ini bekerja dengan menetapkan kondisi spesifik, contohnya: jika *Magnetometer_Z* melebihi ambang tertentu dan *Index_PIP* berada di bawah ambang lain, model memprediksi Kelas 16. Hasilnya terukur: F1 Kelas 16 meningkat dari 0,97 menjadi 0,9972, dan F1 Kelas 17 dari 0,95 menjadi 0,9978.

F. Validasi Empiris terhadap Strategi Seleksi Fitur

Penelitian acuan mengecualikan saluran magnetometer dengan alasan bahwa akselerometer lebih mudah dipahami secara fisik, sementara magnetometer dianggap lebih mudah terganggu oleh derau lingkungan. Seluruh hasil eksperimen menunjukkan sebaliknya.

Perbedaan fisik kedua jenis sensor tersebut adalah: akselerometer mengukur komponen percepatan gravitasi sepanjang sumbu tangan lokal, sehingga sensitif terhadap kemiringan statis namun tidak sensitif terhadap rotasi *yaw*. Magnetometer mengukur komponen medan magnet bumi sepanjang sumbu lokal, merepresentasikan orientasi rotasional tangan pada bidang horizontal. Untuk angka SIBI yang dibedakan terutama oleh arah putaran pergelangan tangan, magnetometer adalah satu-satunya saluran yang merekam perbedaan orientasi tersebut secara langsung.

Secara kuantitatif, penambahan kembali lima saluran yang dikecualikan pada XGBoost menghasilkan kenaikan akurasi +0,23 poin persentase (99,60% ke 99,83%), F1

Kelas 16 +0,0074, dan F1 Kelas 17 +0,0100. Pada MLP, peningkatannya +1,18 poin persentase (97,90% ke 99,08%). Selisih yang lebih besar pada MLP menunjukkan bahwa XGBoost lebih toleran terhadap fitur yang tidak lengkap berkat mekanisme ansambel, sedangkan MLP lebih bergantung pada kelengkapan data masukan.

G. Perbandingan Model Pohon vs. Jaringan Saraf Tiruan

Penelitian ini menambah bukti empiris atas temuan Grinsztajn et al. (2022): hierarki kinerja XGBoost > MLP > kNN konsisten di seluruh konfigurasi himpunan fitur. Keunggulan XGBoost berasal dari seleksi fitur inheren melalui kriteria pemisahan, batas keputusan non-linear yang efisien melalui partisi ruang persegi panjang, dan sensitivitas lebih rendah terhadap fitur masukan yang tidak relevan dibanding metode berbasis jarak.

H. Keterbatasan Penelitian

Perlu dicatat bahwa akurasi *baseline* kNN (98,90%) berasal dari skema validasi Widodo (2022) yang berbeda dari protokol evaluasi penelitian ini, sehingga perbandingan langsung bersifat indikatif. *Dataset* yang digunakan hanya mencakup gestur statis terisolasi; komunikasi nyata melibatkan urutan berkelanjutan dengan efek ko-artikulasi antar gestur. Data dikumpulkan dari satu desain sarung tangan dan satu subjek, sehingga generalisasi ke desain lain atau pengguna yang berbeda belum dapat dikonfirmasi. Jumlah iterasi Optuna dibatasi pada 100 percobaan per konfigurasi akibat keterbatasan komputasi. Hierarki kepentingan fitur yang ditemukan mencerminkan geometri gestur angka SIBI dan kemungkinan tidak berlaku umum pada sistem bahasa isyarat lain.

KESIMPULAN

Penelitian ini menyajikan evaluasi komparatif sistematis terhadap kNN, XGBoost, dan MLP untuk klasifikasi 20 angka bahasa isyarat SIBI dari *dataset* sarung tangan sensor 16 saluran, sekaligus menguji validitas empiris strategi seleksi fitur berbasis intuisi yang diterapkan pada penelitian sebelumnya. Lima temuan utama diperoleh: (1) XGBoost melampaui kNN pada semua konfigurasi; dengan 11 fitur, XGBoost yang dioptimalkan mencapai akurasi uji 99,60% atau 0,70 poin persentase di atas *baseline*; (2) himpunan 16 fitur penuh lebih unggul secara konsisten, dengan XGBoost mencapai 99,83% dan MLP mencapai 99,08%; (3) XGBoost menyelesaikan ambiguitas Kelas 16/17 dengan F1 masing-masing 0,9972 dan 0,9978, dibandingkan 0,97 dan 0,95 pada kNN; (4) pohon

keputusan melampaui jaringan saraf tiruan pada data sensor tabular ini, konsisten dengan Grinsztajn et al. (2022); (5) saluran magnetometer terbukti lebih informatif untuk klasifikasi dibanding saluran akselerometer yang dipertahankan dalam *baseline*, menunjukkan bahwa pemilihan fitur berdasarkan analisis data menghasilkan konfigurasi yang lebih baik dibanding pertimbangan manual. Langkah lanjutan mencakup penerapan *TinyML* untuk mengalihkan model XGBoost terlatih ke C/C++ pada mikrokontroler, pengenalan urutan isyarat berkelanjutan menggunakan arsitektur LSTM atau *Transformer*, serta validasi lintas subjek sebelum penerapan di lingkungan nyata.

UCAPAN TERIMA KASIH

Para penulis mengucapkan terima kasih kepada Prof. Dr.Eng. Romy Budhi Widodo atas penyediaan *dataset* sarung tangan sensor asli beserta penelitian dasarnya yang menjadi landasan perbandingan dalam studi ini. Penelitian ini didukung oleh Program Studi Teknik Informatika, Universitas Ma Chung.

DAFTAR PUSTAKA

- Adeyanju, I. A., Bello, O. O., & Adegboye, M. A. (2021). Machine learning methods for sign language recognition: A critical review and analysis. *Intelligent Systems with Applications*, 12, 200056. <https://doi.org/10.1016/j.iswa.2021.200056>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2623–2631. <https://doi.org/10.1145/3292500.3330701>
- Ambar, R., Salim, S., Abd Wahab, M. H., Abdul Jamil, M. M., & Phing, T. C. (2023). Development of a wearable sensor glove for real-time sign language translation. *Annals of Emerging Technologies in Computing (AETiC)*, 7(5), 25–38. <https://doi.org/10.33166/AETiC.2023.05.003>
- Amin, M. S., Rizvi, S. T. H., & Hossain, M. M. (2022). A comparative review on applications of different sensors for sign language recognition. *Journal of Imaging*, 8(4), 98. <https://doi.org/10.3390/jimaging8040098>

- Badan Pusat Statistik. (2024). *Potret penyandang disabilitas di Indonesia: Hasil Long Form Sensus Penduduk 2020*. BPS.
<https://www.bps.go.id/id/publication/2024/12/20/43880dc0f8be5ab92199f8b9/potret-penyandang-disabilitas-di-indonesia--hasil-long-form-sp2020.html>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cooper, H., Holt, B., & Bowden, R. (2011). Sign language recognition. In T. B. Moeslund, A. Hilton, V. Krüger, & L. Sigal (Eds.), *Visual analysis of humans* (pp. 539–562). Springer. https://doi.org/10.1007/978-0-85729-997-0_27
- Departemen Pendidikan Nasional Republik Indonesia. (1994). *Kamus sistem isyarat bahasa Indonesia (SIBI)*. Depdiknas RI.
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why tree-based models still outperform deep learning on tabular data. *Advances in Neural Information Processing Systems*, 35, 507–520. <https://doi.org/10.48550/arXiv.2207.08815>
- Malika, M., & Berlianti. (2025). Penggunaan bahasa isyarat SIBI dan BISINDO untuk siswa tuli di Sekolah Luar Biasa Negeri Pembina Medan. *Concept: Journal of Social Humanities and Education*, 4(3), 78–87.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, 8024–8035. <https://doi.org/10.48550/arXiv.1912.01703>
- Widodo, R. B. (2022). *Machine learning metode k-Nearest Neighbors: Klasifikasi angka bahasa isyarat*. Media Nusa Creative.