HOAQ: JURNAL TEKNIK INFORMATIKA, Volume 17 Nomor 2, Desember 2026, p341-350
p-ISSN: 2337-5280, e-ISSN: 26207427
DOI: https://doi.org/10.52972/hoaq.vol17no2

EVALUASI KINERJA ALGORITMA *MACHINE LEARNING* DALAM KLASIFIKASI POSTUR DUDUK MENGGUNAKAN YOLOV8N-POSE

Naufal Aqil Hidayat[1], Muhammad Rafli[2], Andre Saputra[3] , Khahfiana Rosida[4], Muhammad Robis Alfaf [5], Jordy Lasmana Putra[6]

Program Studi Informatika, Universitas Bina Sarana Informatika
Jl. Kamal Raya No.18, Cengkareng, Jakarta Barat – DKI Jakarat, Indonesia
Email: [1]hinaufal29@gmail.com, [2]raflicubetv@gmail.com, [3]andresaputranomor2@gmail.com,
[4]khahfianar@gmail.com, [5]muhammadrobisalfaf@gmail.com, [6]jordy.jlp@bsi.ac.id

ABSTRAK

Penggunaan perangkat digital yang meningkat memperpanjang durasi duduk dan berisiko menimbulkan gangguan muskuloskeletal akibat postur yang tidak ergonomis. Penelitian ini mengembangkan sistem klasifikasi postur duduk berbasis citra menggunakan YOLOv8n-Pose (*Nano*) untuk ekstraksi 17 titik sendi tubuh, yang kemudian diratakan (*flattening*) menjadi vektor 51 fitur numerik dan dinormalisasi dengan *Standard Scaler*. *Dataset* gabungan dari Roboflow dan Kaggle, yang terdiri atas 3.558 citra dengan rincian 1.044 *good_form* dan 2.514 *bad_form*, diuji pada lima algoritma yaitu *Random Forest*, KNN, SVM, MLP, dan Deep MLP, menggunakan skema *hold-out* 80:20 berbasis *stratified sampling*. Pada data uji, *Random Forest* mencatat performa generalisasi paling seimbang dengan akurasi 0,9775 dan *Macro F1-Score* 0,9728, sedikit mengungguli *MLP Neural Network* yang menunjukkan performa yang juga kuat, meskipun *Random Forest* sebelumnya menunjukkan indikasi *overfitting* pada fase pelatihan. Sistem kemudian diimplementasikan sebagai aplikasi web bernama SitSmart di *Hugging Face Spaces*. Penelitian ini mengindikasikan bahwa kombinasi ekstraksi koordinat pose dan model klasifikasi *ensemble* mampu memberikan hasil kompetitif untuk pemantauan ergonomi otomatis, meski validasi lapangan *real-time* masih diperlukan pada tahap lanjut.

Kata kunci: ergonomi digital, computer vision, YOLOv8n-Pose, *Random Forest*, klasifikasi postur duduk

ABSTRACT

Rising use of digital devices has extended sitting durations, thereby increasing the risk of musculoskeletal disorders from poor posture. This study develops an image-based sitting posture classification system using YOLOv8n-Pose (Nano) to extract 17 body keypoints, which are then flattened into a 51-feature numeric vector and normalized with Standard Scaler. A combined Roboflow and Kaggle dataset, consisting of 3,558 images with 1,044 good_form and 2,514 bad_form samples, was evaluated across five algorithms, namely Random Forest, KNN, SVM, MLP, and Deep MLP, using an 80:20 stratified hold-out split. On the test set, Random Forest achieved the most balanced generalization, with an accuracy of 0.9775 and a Macro F1-Score of 0.9728, narrowly ahead of MLP Neural Network, which also showed strong performance, although Random Forest had previously shown signs of overfitting during training. The best-performing model was then deployed as a web application called SitSmart on Hugging Face Spaces. These findings suggest that combining pose-based coordinate extraction with an ensemble classifier can yield competitive results for automated ergonomic monitoring, although real-world and real-time field validation remains necessary as a direction for future work.

Keywords: digital ergonomics, computer vision, YOLOv8n-Pose, Random Forest, sitting posture classification.

1. PENDAHULUAN

Perkembangan teknologi informasi yang masif telah mengubah lanskap aktivitas manusia secara fundamental, khususnya pada sektor pendidikan dan lingkungan kerja profesional yang kini sangat bergantung pada intervensi perangkat digital. Transisi menuju gaya hidup pasif (*sedentary*) ini memaksa individu untuk menghabiskan waktu dalam posisi duduk dengan durasi yang sangat panjang, sehingga berpotensi memicu risiko gangguan kesehatan kronis apabila tidak diimbangi dengan postur tubuh yang benar [1], [2]. Postur duduk yang salah dan tidak ergonomis secara berkelanjutan diyakini sebagai pemicu utama munculnya *Musculoskeletal Disorders* (MSDs), yang bermanifestasi pada keluhan nyeri punggung bagian bawah, ketegangan bahu, hingga sindrom leher kaku (*text neck*) [3], [4], [5]. Beberapa literatur penelitian mengonfirmasi bahwa posisi anatomis tubuh yang terdistorsi dalam waktu lama tidak hanya berdampak pada keluhan fisiologis, tetapi juga berkontribusi secara langsung terhadap kelelahan mental dan penurunan tingkat produktivitas kerja [2], [3]. Oleh karena itu,

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

intervensi berupa sistem pemantauan postur tubuh secara presisi menjadi salah satu langkah preventif yang krusial dalam penerapan konsep ergonomi digital [1], [6].

Ergonomi digital hadir sebagai pendekatan multidisiplin yang bertujuan untuk mengharmonisasikan interaksi manusia dengan perangkat digital agar tetap aman, nyaman, dan efisien secara biomekanik. Penerapan standar ergonomi yang proaktif terbukti mampu menekan risiko cedera fisik sekaligus meningkatkan efektivitas kerja pengguna[1], [6]. Dalam konteks ini, pesatnya inovasi di bidang kecerdasan buatan, khususnya *Computer Vision*, memungkinkan pengembangan sistem analisis postur tubuh yang mampu bekerja secara otomatis dan cerdas melalui interpretasi citra digital. Dibandingkan dengan penggunaan perangkat sensor fisik terpakai (*wearable sensors*) yang sering kali membatasi ruang gerak pengguna, teknologi visi komputer dinilai jauh lebih fleksibel, efisien dari segi biaya, mudah diimplementasikan, dan tangguh dalam memproses inferensi secara *real-time* [7], [8], [9]. Salah satu pendekatan mutakhir yang banyak diadopsi untuk tujuan ini adalah teknologi estimasi pose (*pose estimation*), yang berfungsi melacak dan mendeteksi titik-titik persendian (kerangka) tubuh manusia di dalam ruang dua dimensi [10], [11], [12], [13].

Beberapa penelitian sebelumnya telah menunjukkan bahwa hibridisasi antara ekstraksi *pose estimation* dan algoritma klasifikasi *machine learning* mampu menghasilkan performa komputasi yang menjanjikan dalam domain pengenalan postur tubuh. Salah satu studi rujukan mengusulkan algoritma *Weighted Random Forest* untuk klasifikasi postur duduk berbasis koordinat dan sudut wajah-bahu dari citra, dengan akurasi mencapai 98,3% [7]. Studi rujukan lain memanfaatkan pustaka MediaPipe Pose untuk ekstraksi titik koordinat tubuh yang dipadukan dengan pengklasifikasi berbasis aturan (rule-based) mengikuti metode OWAS untuk penilaian risiko ergonomi, dengan akurasi berkisar 85,53%–92,71% pada pengujian per bagian tubuh [14]. Berangkat dari kedua pendekatan tersebut, penelitian ini mengusulkan kebaruan (*novelty*) berupa integrasi ekstraksi titik koordinat menggunakan algoritma YOLOv8n-Pose (*Nano*) yang dipadukan dengan algoritma *Random Forest* untuk klasifikasi postur duduk. Perlu dicatat bahwa perbandingan angka akurasi antar studi rujukan dan penelitian ini bersifat indikatif, mengingat perbedaan dataset, kondisi pengambilan citra, dan jumlah sampel yang digunakan, sehingga bukan merupakan perbandingan yang sepenuhnya setara (apple-to-apple). Pemilihan varian *Nano* ini didasarkan pada karakteristik *State-of-the-Art* (SOTA) dari arsitektur model YOLOv8 yang sangat unggul dalam menjaga keseimbangan antara kecepatan deteksi objek (*framerate*) dan preservasi akurasi ekstraksi titik sendi [10], [11], [12], [13].

Alih-alih menggunakan ekstraktor fitur visual berbasis matriks piksel konvensional, representasi data pada penelitian ini direkayasa secara khusus dengan mengekstrak murni koordinat spasial dari titik sendi tubuh [7], [15]. Data spasial tersebut kemudian diratakan ke dalam bentuk vektor numerik satu dimensi melalui teknik *flattening*, sebelum akhirnya dinormalisasi secara menyeluruh menggunakan fungsi *Standard Scaler* guna memastikan peningkatan kualitas distribusi data sebelum memasuki proses klasifikasi. Untuk mendapatkan lanskap evaluasi yang komprehensif, data koordinat tabular hasil ekstraksi YOLO Pose tersebut diuji secara komparatif menggunakan lima algoritma berbeda, yakni *Random Forest* (RF), *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), *Multi-Layer Perceptron* (MLP) standar, serta penambahan arsitektur jaringan saraf tiruan berlapis, yakni *Deep* MLP, sebagai model pembanding utama [15], [8], [9].

Melalui studi perbandingan yang ketat dan menyeluruh ini, nilai metrik performa serta pola kesalahan prediksi (*confusion matrix*) dari setiap algoritma akan dibedah secara empiris dan mendalam. Tujuan akhir dari penelitian ini adalah untuk menemukan dan memvalidasi algoritma klasifikasi postur duduk yang paling presisi, sehingga siap diimplementasikan sebagai mesin *engine* utama dalam membangun sistem pemantauan ergonomi digital yang aplikatif dan andal pada ekosistem produksi nyata [1], [15], [6].

2. METODE PENELITIAN

Rangkaian penelitian ini dijalankan dengan berpedoman pada kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) guna memastikan setiap tahapan berjalan secara sistematis, terstruktur, dan dapat diuji ulang (*reproducible*). Penggunaan metodologi CRISP-DM sangat relevan untuk penelitian berbasis kecerdasan buatan karena mencakup siklus yang komprehensif, mulai dari pemahaman masalah ergonomi (*Business Understanding*), pengumpulan citra (*Data Understanding*), prapemrosesan koordinat (Data Preparation), komparasi algoritma (*Modeling*), penilaian metrik performa (*Evaluation*), hingga bermuara pada tahap implementasi sistem sebagai aplikasi berbasis web (*Deployment*).

Dataset Penelitian

Dataset utama yang digunakan dalam penelitian ini diekstraksi dari dua repositori publik, yaitu *Roboflow Sitting Posture Dataset* dan *dataset* tambahan dari *platform* Kaggle. *Dataset* tersebut memuat ribuan citra digital subjek manusia yang tengah beraktivitas di meja kerja dengan berbagai variasi kondisi pencahayaan, sudut pengambilan gambar (*camera angle*), dan postur tubuh. Seluruh data gambar dikelompokkan ke dalam dua kategori kelas utama: "Postur Baik" (tubuh tegak, sudut kelengkungan punggung melebihi 90 derajat, pandangan sejajar monitor) dan "Postur Buruk" (tubuh membungkuk ke depan, bersandar secara berlebihan, atau posisi leher terlalu condong ke depan).

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

Dataset yang digunakan telah memiliki label *good_form* dan *bad_form* bawaan sehingga proses anotasi manual tidak diperlukan. Total dataset hasil penggabungan berjumlah 3.558 citra. *Dataset* ini memiliki karakteristik ketidakseimbangan kelas (*imbalanced data*), di mana sampel postur buruk jauh lebih mendominasi dibandingkan postur baik, yang secara langsung menguji ketangguhan model dalam menangani bias minoritas. Rincian distribusi kelas dari dataset dapat dilihat pada Tabel 1.

Tabel 1. Distribusi dataset postur duduk

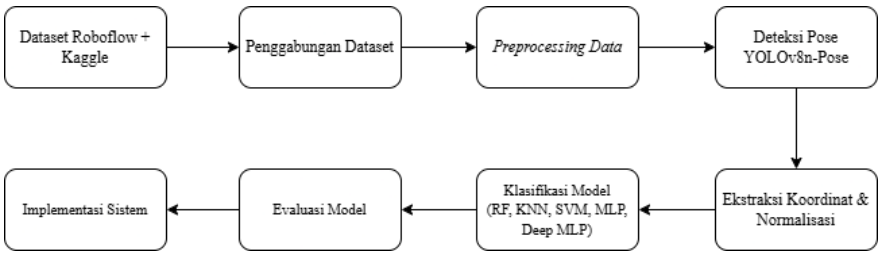
Kelas Postur	Jumlah Citra	Label Target
Postur Baik	1.044	*good_form*
Postur Buruk	2.514	*bad_form*
Total	3.558	

Kedua sumber *dataset* memiliki skema label yang berbeda sehingga perlu diharmonisasi sebelum digunakan bersama. *Dataset* dari Kaggle memiliki label kelas yang lebih spesifik, yaitu *CorrectSitting*, *LeanBackward*, *LeanForward*, dan *LegTwisting*, sedangkan dataset dari Roboflow menggunakan skema label yang lebih sederhana berupa *good_posture* dan sejenisnya. Untuk menyeragamkan kedua sumber tersebut, seluruh label dipetakan ulang (*relabeling*) ke dalam skema klasifikasi biner: kelas *CorrectSitting* dan *good_posture* dipetakan menjadi *good_form*, sedangkan kelas *LeanBackward*, *LeanForward*, dan *LegTwisting* dipetakan menjadi *bad_form*.

Seluruh citra yang digunakan merupakan data sekunder yang bersifat terbuka (*open-source*), bersumber dari dua repositori publik. Sumber pertama adalah dataset "*Sitting-Posture-Recognition*" yang didistribusikan oleh Huy Quang Chu melalui platform Kaggle di bawah lisensi Apache License 2.0 [16]. Sumber kedua adalah dataset "Sitting Posture v2 Computer Vision Dataset" yang didistribusikan melalui Roboflow di bawah lisensi CC BY 4.0 [17]. Kedua *dataset* digunakan sesuai dengan ketentuan lisensi masing-masing untuk kepentingan riset dan pengembangan akademis.

Tahapan penelitian

Secara garis besar, alur penelitian ditunjukkan pada Gambar 1. Tahapan penelitian diawali dengan ekstraksi koordinat pose dari dataset citra menggunakan algoritma YOLOv8n-Pose (*Nano*). Varian model ini dipilih karena kemampuannya mendeteksi 17 titik kerangka tubuh (*keypoints*) dengan sangat akurat sekaligus mempertahankan kecepatan komputasi yang tinggi, sehingga ideal untuk diimplementasikan pada sistem *real-time*.



Gambar 1. *Flowchart* Penelitian

Proses ekstraksi fitur pada penelitian ini dilakukan dengan mengambil koordinat spasial (X dan Y) serta tingkat keyakinan (*Confidence Score*) dari setiap titik sendi yang terdeteksi oleh YOLOv8n-Pose. Representasi data ini kemudian melalui proses *flattening* untuk diratakan menjadi vektor fitur satu dimensi yang berisi 51 atribut numerik (17 titik x 3 nilai). Selanjutnya, fitur tersebut dinormalisasi menggunakan fungsi *Standard Scaler* untuk memastikan seluruh rentang nilai terstandarisasi sebelum memasuki tahap klasifikasi. Kelima model yang dikomparasikan dalam penelitian ini meliputi *Random Forest* (RF), *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), *Multi-Layer Perceptron* (MLP) Tradisional, dan model *Deep Learning* berbentuk *Deep MLP*.

Guna memastikan penelitian ini transparan dan dapat direproduksi oleh peneliti lain (*reproducible*), konfigurasi parameter dari masing-masing algoritma klasifikasi ditetapkan secara spesifik. Pengaturan parameter ini didasarkan pada konfigurasi baseline yang umum digunakan dalam klasifikasi fitur tabular berdimensi rendah, serta disesuaikan untuk mencapai konvergensi yang optimal. Rincian parameter utama untuk setiap model ditunjukkan pada Tabel 2.

Tabel 2. Parameter utama tabel klasifikasi

Model	Parameter	Nilai
Random Forest	n_estimators	100
	max_depth	None
MLP Neural Network	hidden_layer_sizes	(100,)

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

Model	Parameter	Nilai
	max_iter	500
SVM	kernel	rbf
	C	1.0
KNN	n_neighbors	5
	weights	uniform
Deep MLP	hidden_layer_sizes	(128, 64, 32)
	activation	relu

Pembagian data dan penanganan ketidakseimbangan kelas

Sebelum tahap pelatihan, *dataset* yang telah melalui proses ekstraksi fitur dibagi secara langsung menjadi data latih (*train*) dan data uji (*test*) dengan rasio 80:20 (tanpa alokasi terpisah untuk validation set khusus, mengingat arsitektur model menggunakan parameter baseline standar). Pembagian dilakukan menggunakan teknik *Stratified Random Sampling* (parameter *stratify* pada fungsi *train_test_split*), sehingga proporsi kelas *good_form* dan *bad_form* pada data latih dan data uji tetap konsisten dengan proporsi pada *dataset* keseluruhan. Skema evaluasi yang digunakan adalah *hold-out validation* (satu kali pembagian data), bukan *k-fold cross-validation*; pemilihan ini didasarkan pada pertimbangan efisiensi komputasi, namun peneliti menyadari bahwa evaluasi *hold-out* tunggal memiliki keterbatasan dalam mengukur variabilitas performa model, sehingga hasil generalisasi yang dilaporkan pada bagian selanjutnya perlu dimaknai dengan kehati-hatian yang proporsional.

Pembagian data pada penelitian ini dilakukan secara acak berbasis citra (*image-based split*), bukan berbasis subjek (*subject-based split*). Hal ini disebabkan oleh keterbatasan dataset publik yang digunakan (Roboflow dan Kaggle), yang tidak menyertakan informasi identitas subjek pada setiap citra sehingga pemisahan data berdasarkan subjek tidak dapat dilakukan. Konsekuensinya, terdapat kemungkinan citra dari subjek yang sama muncul baik pada data latih maupun data uji, yang berpotensi menyebabkan estimasi performa model sedikit optimistis (data leakage antar subjek). Keterbatasan ini menjadi salah satu arah pengembangan penelitian selanjutnya, misalnya dengan mengumpulkan dataset yang menyertakan label subjek.

Terkait penanganan ketidakseimbangan kelas (*imbalanced data*), penelitian ini tidak menerapkan teknik penambahan data sintetis (SMOTE) maupun pembobotan kelas (*class_weight*) pada tahap pemodelan. Mitigasi bias kelas mayoritas dilakukan melalui dua langkah: (1) penerapan *stratified sampling* pada tahap pembagian data agar proporsi kelas tetap terjaga baik di data latih maupun data uji, dan (2) penggunaan metrik evaluasi yang lebih tahan terhadap ketidakseimbangan kelas, yaitu Balanced Accuracy, Macro F1-Score, dan Recall per kelas, sebagaimana disajikan pada bagian Hasil dan Pembahasan.

Evaluasi Model

Evaluasi performa model dilakukan secara komprehensif dengan mengukur empat metrik kinerja utama, yaitu *Accuracy, Precision, Recall*, dan F1-*score*. Pemilihan metrik ini didasarkan pada kebutuhan untuk melihat keandalan model dari berbagai sudut pandang. *Accuracy* digunakan untuk mengukur persentase ketepatan prediksi model secara keseluruhan. Namun, mengingat distribusi kelas pada dataset penelitian ini tidak seimbang (*imbalanced data*), di mana jumlah citra *bad_form* mendominasi kelas *good_form*, metrik *Accuracy* saja tidak cukup representatif untuk dijadikan tolak ukur utama.

Oleh karena itu, evaluasi difokuskan pada *Precision* dan *Recall*. *Precision* mengukur seberapa presisi model saat memprediksi kelas positif guna menekan margin kesalahan deteksi palsu (*False Positive*). Sementara itu, *Recall* mengukur sensitivitas kemampuan model dalam mengenali seluruh data yang memang bernilai positif guna menekan angka kelolosan deteksi (*False Negative*). Untuk menyeimbangkan *trade-off* antara kedua metrik tersebut, digunakan *F1 - Score* yang merupakan rata-rata harmonik dari *Precision* dan *Recall*. Secara matematis, metrik-metrik tersebut dievaluasi menggunakan formulasi berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Selain metrik matematis, penelitian ini juga menggunakan *Confusion Matrix* sebagai instrumen evaluasi visual. *Confusion Matrix* memetakan nilai aktual dan nilai prediksi ke dalam komponen *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) dalam bentuk matriks dua dimensi. Visualisasi ini krusial untuk membedah secara spesifik kelas postur mana yang paling sering diprediksi dengan keliru oleh sistem, sehingga analisis komparatif terhadap pola kelemahan masing-masing algoritma dapat diurai secara mendalam.

3. HASIL DAN PEMBAHASAN

Pengujian komprehensif dilakukan terhadap lima algoritma klasifikasi guna menentukan arsitektur model terbaik dalam mengenali dan mengkategorikan postur duduk. Pengenalan ini didasarkan pada hasil ekstraksi 51 atribut koordinat spasial linier yang diproduksi oleh YOLOv8n-Pose (*Nano*) dan telah melalui tahap normalisasi *Standard Scaler*. Proses evaluasi ini dirancang tidak hanya untuk mencari model dengan akurasi metrik tertinggi, tetapi juga untuk menguji sejauh mana strategi pra-pemrosesan data yang efisien, yaitu berupa ekstraksi koordinat spasial dan normalisasi fitur, dapat memengaruhi performa relatif antara algoritma *machine learning* konvensional dan arsitektur jaringan saraf tiruan pada data tabular berdimensi rendah.

Hasil evaluasi model pada data uji

Tahap pengujian dibagi menjadi dua fase utama untuk menjamin validitas dan keandalan model, yaitu fase validasi menggunakan data latih (*training*) dan fase pengujian menggunakan data uji (*test set*). Evaluasi pertama pada fase validasi bertujuan untuk mengukur kemampuan awal masing-masing algoritma dalam mempelajari representasi fitur koordinat tubuh dan melakukan konvergensi terhadap pola data. Hasil metrik kinerja pada tahap validasi ditunjukkan pada Tabel 3.

Tabel 3. Hasil Perbandingan Model

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	1,0000	1,0000	1,0000	1,0000
MLP Neural Network	0,9965	0,9965	0,9965	0,9965
KNN	0,9677	0,9677	0,9677	0,9677
SVM	0,9487	0,9484	0,9487	0,9484
Deep MLP	0,9174	0,9180	0,9174	0,9177

Selanjutnya, kelima model tersebut dievaluasi kembali menggunakan data uji (*test set*) yang berisi variasi citra baru dan belum pernah dilatih sebelumnya. Pengujian tahap kedua ini sangat krusial guna mengukur kemampuan generalisasi model saat dihadapkan pada kasus deteksi postur di dunia nyata. Hasil pengujian empiris pada data uji disajikan pada Tabel 4.

Tabel 4. Hasil evaluasi model pada data uji

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0,9775	0,9775	0,9775	0,9775
MLP Neural Network	0,9705	0,9704	0,9705	0,9704
KNN	0,9382	0,9384	0,9382	0,9383
SVM	0,9424	0,9421	0,9424	0,9418
Deep MLP	0,9017	0,9038	0,9017	0,9025

Berdasarkan komparasi metrik kinerja antara Tabel 3 (fase pelatihan) dan Tabel 4 (fase uji), terlihat adanya dinamika generalisasi pada masing-masing algoritma. Model *Random Forest* mencatatkan skor sempurna (1,0000) pada seluruh metrik di data latih. Sesuai dengan karakteristik algoritma berbasis *tree*, skor sempurna ini merupakan indikasi klasik dari kecenderungan model untuk menghafal data (*overfitting*) akibat pertumbuhan kedalaman pohon yang tidak dibatasi (max_depth=None). Namun demikian, *overfitting* ini justru tidak berdampak negatif

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

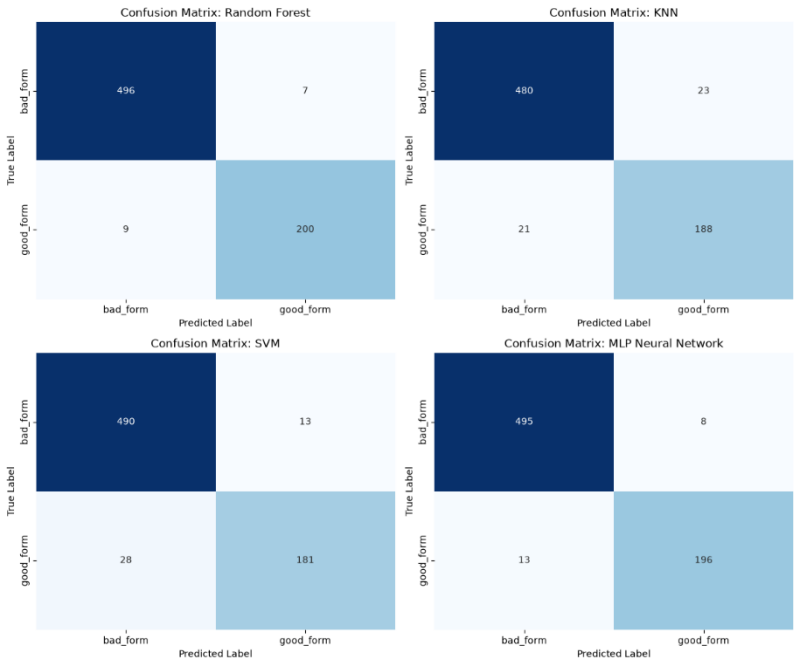
terhadap kemampuan generalisasinya; model tetap mampu mempertahankan performa terbaik di antara seluruh model saat dihadapkan pada data uji, dengan perolehan akurasi sebesar 0,9775 dan F1-Score sebesar 0,9775. Di sisi lain, *MLP Neural Network* juga menunjukkan generalisasi yang kuat, dengan penurunan skor dari 0,9965 (data latih) menjadi 0,9705 (data uji), menempatkannya di posisi kedua terbaik namun dengan selisih yang cukup terlihat dari *Random Forest*.

Pada pengujian data uji, model *Random Forest* berhasil menunjukkan performa generalisasi terbaik di antara seluruh model yang diuji, dengan tingkat akurasi mencapai 0,9775 dan nilai F1-score sebesar 0,9775. Hasil ini mengindikasikan bahwa meskipun algoritma ensemble berbasis pohon menunjukkan indikasi *overfitting* pada data latih, kemampuannya dalam mengagregasi ratusan pohon keputusan tetap menghasilkan generalisasi yang sangat kuat terhadap data baru. Di posisi kedua, model *MLP Neural Network* juga menampilkan kinerja yang solid dengan perolehan akurasi 0,9705 dan F1-score 0,9704, menjadikannya alternatif yang tangguh meskipun tertinggal tipis dari *Random Forest*.

Sebaliknya, model *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) memperlihatkan performa di kelas menengah dengan perolehan F1-score masing-masing sebesar 0,9418 dan 0,9383, di mana SVM tampil sedikit lebih unggul dibandingkan KNN pada data uji ini. Hal ini mengindikasikan bahwa metode komputasi jarak ketetanggaan pada KNN dan konstruksi batas keputusan non-linier berbasis kernel RBF pada SVM mulai menemui hambatan saat membedakan sampel kelas postur baik dan buruk yang memiliki pergeseran koordinat sendi (*noise*) yang saling berdekatan. Sementara itu, model Deep MLP menempati urutan terendah dengan akurasi 0,9017 dan F1-*score* 0,9025, yang mengisyaratkan kembali temuan bahwa kompleksitas arsitektur *deep learning* berlapis tanpa dukungan volume data berskala masif justru kurang optimal dalam mengekstraksi dataset berdimensi rendah (51 atribut).

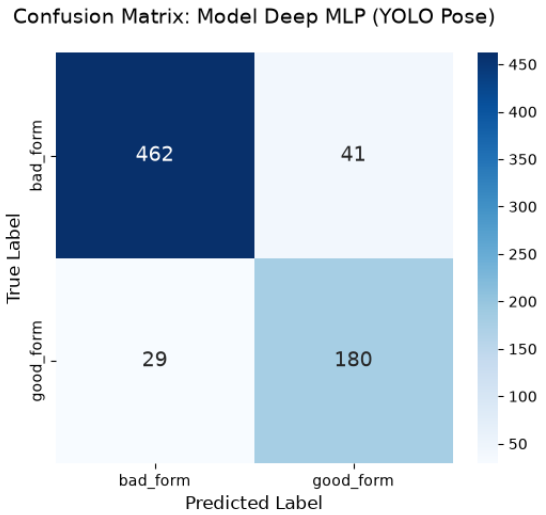
Visualisasi Confusion Matrix

Untuk menganalisis pola kesalahan prediksi secara lebih rinci, *confusion matrix* dari seluruh model dievaluasi untuk melihat distribusi prediksi yang benar maupun yang meleset pada kelas *good_form* dan *bad_form*.



Gambar 2. *Confusion Matrix Model* (RF, KNN, SVM, MLP)

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

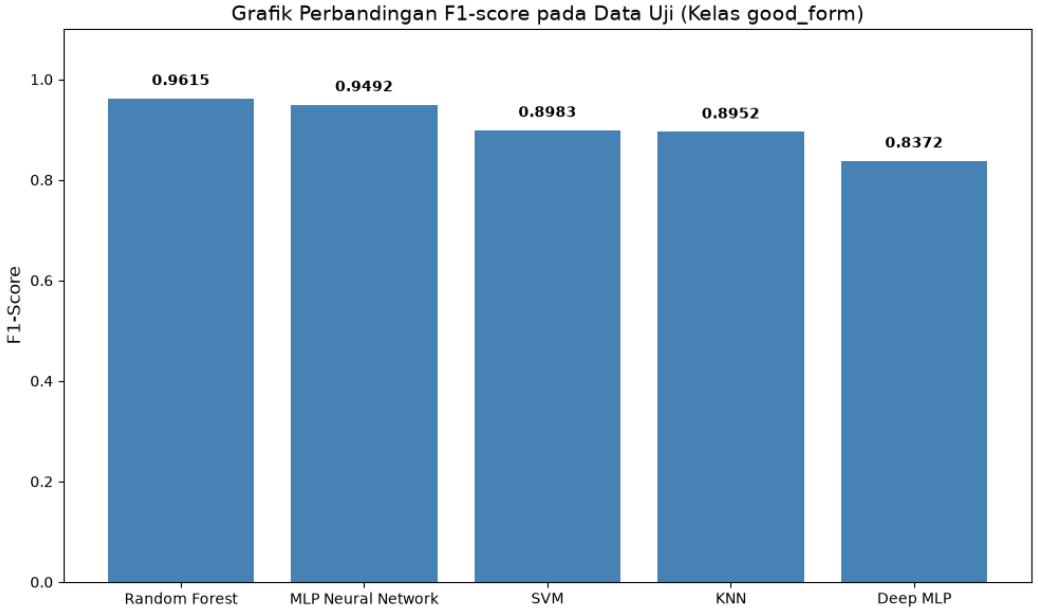


Gambar 3. *Confusion Matrix Model Deep* MLP

Visualisasi *confusion matrix* memperlihatkan kembali keunggulan *Random Forest* yang menunjukkan tingkat keseimbangan prediksi paling optimal di antara kedua kelas, dengan margin misklasifikasi yang minim baik pada kelas mayoritas (*bad_form*) maupun kelas minoritas (*good_form*). Model *MLP Neural Network* juga mencatatkan performa yang sangat kompetitif dan mendekati *Random Forest*, meskipun menghasilkan angka *False Positive* yang sedikit lebih tinggi pada kelas *good_form*. Sebaliknya, *Deep MLP* mencatatkan tingkat kesalahan tertinggi secara keseluruhan. Hal ini memvalidasi indikasi bahwa penambahan kompleksitas arsitektur pada *deep learning* tanpa diimbangi volume data berskala masif justru memicu peningkatan misklasifikasi saat mengekstraksi fitur tabular linier.

Perbandingan F1-Score Antar Model

Untuk memperjelas perbandingan performa secara visual dan melihat stabilitas model dalam menangani klasifikasi, dilakukan visualisasi menggunakan grafik batang berbasis nilai F1-*score* sebagaimana ditunjukkan pada Gambar 4. Penggunaan metrik F1-score sangat krusial dalam tahapan evaluasi ini karena metrik tersebut memberikan gambaran yang lebih komprehensif terkait keseimbangan antara *Precision* dan *Recall*, sehingga dapat dipastikan bahwa tidak ada model yang bias dalam memprediksi salah satu kelas postur secara sepihak.



Gambar 4. Grafik perbandingan F1-*score* kelas minoritas (*good_form*) pada data uji

Berdasarkan Gambar 4, grafik secara spesifik menyoroti performa metrik F1-*score* dalam mengklasifikasikan kelas minoritas (*good_form*). Pemfokusan evaluasi pada kelas ini sangat krusial untuk memastikan bahwa model tidak bias dan tidak hanya mengandalkan dominasi jumlah sampel pada kelas mayoritas. Hasil visualisasi mengindikasikan bahwa *Random Forest* memimpin dengan perolehan F1-score tertinggi sebesar 0,9615, disusul oleh *MLP Neural Network* pada angka 0,9492. Pencapaian ini mengindikasikan bahwa mekanisme agregasi ratusan pohon keputusan pada *Random Forest* justru menghasilkan keseimbangan sensitivitas yang sangat

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

baik terhadap kelas minoritas, sekaligus mempertegas bahwa indikasi *overfitting* yang teramati pada data latih tidak menghambat kemampuan model dalam mengenali pola postur duduk yang benar di tengah tantangan ketidakseimbangan *dataset* (*imbalanced data*).

Di sisi lain, algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) mencatatkan performa di tingkat menengah dengan skor masing-masing 0,8983 dan 0,8952, keduanya berdekatan cukup rapat. Sementara itu, *Deep MLP* kembali berada pada posisi terendah dengan skor 0,8372. Pola penurunan performa pada model-model tersebut konsisten dengan evaluasi metrik sebelumnya, yang memperlihatkan bahwa tanpa adanya intervensi sintesis data (seperti SMOTE), model yang rentan terhadap jarak spasial dan kompleksitas jaringan akan mengalami kesulitan signifikan dalam mendeteksi kelas minoritas.

Evaluasi metrik seimbang (Balanced Accuracy, Macro F1-Score, dan Recall per kelas)

Mengingat karakteristik dataset yang tidak seimbang (*imbalanced*), metrik *Accuracy, Precision, Recall*, dan F1-*score* pada Tabel 4 dihitung secara *weighted average*, sehingga performa pada kelas mayoritas (*bad_form*) berpotensi mendominasi skor akhir dan menutupi kelemahan model dalam mengenali kelas minoritas (*good_form*). Untuk mengantisipasi bias evaluasi tersebut, dilakukan analisis tambahan menggunakan metrik yang lebih tahan terhadap ketidakseimbangan kelas, yaitu *Balanced Accuracy, Macro* F1-*Score*, dan *Recall* per kelas. Hasil evaluasi ini disajikan pada Tabel 5.

Tabel 5. Evaluasi metrik seimbang pada data uji

Model	*Balanced Accuracy*	*Macro F1-Score*	*Recall good_form*	*Recall bad_form*
Random Forest	0,9715	0,9728	0,9569	0,9861
MLP Neural Network	0,9609	0,9642	0,9378	0,9841
KNN	0,9269	0,9257	0,8995	0,9543
SVM	0,9201	0,9291	0,8660	0,9742
Deep MLP	0,8899	0,8834	0,8612	0,9185

Berdasarkan Tabel 5, model *Random Forest* menunjukkan kemampuan generalisasi yang paling seimbang di antara seluruh model yang diuji, dengan nilai *Balanced Accuracy* sebesar 0,9715 dan *Macro F1-Score* sebesar 0,9728. Model *MLP Neural Network* menyusul di peringkat berikutnya dengan performa yang juga kuat (*Balanced Accuracy* sebesar 0,9609 dan *Macro F1-Score* 0,9642). Di sisi lain, model *KNN* dan *SVM* mencatatkan kinerja metrik seimbang yang lebih rendah, di mana *SVM* menunjukkan bias kelas mayoritas yang paling ekstrem dengan Recall kelas *bad_form* mencapai 0,9742 namun Recall untuk kelas minoritas (*good_form*) merosot ke 0,8660. Model *Deep MLP* mencatatkan kinerja metrik seimbang terendah secara keseluruhan, dengan *Balanced Accuracy* 0,8899. Mengingat sifat dataset yang *imbalanced*, temuan ini secara akademis mengindikasikan bahwa *Random Forest* merupakan arsitektur yang paling andal dalam menjaga stabilitas prediksi lintas kelas, terutama karena prioritas utama sistem pemantauan ergonomi adalah memastikan postur tubuh yang benar (*good_form*) dapat terdeteksi secara presisi tanpa rasio kelolosan (*false negative*) yang tinggi, sebuah kemampuan yang justru tidak terhambat oleh indikasi *overfitting* yang teramati pada fase pelatihan model ini.

Pembahasan perbandingan model

Hasil evaluasi komparatif secara menyeluruh menempatkan model ensemble *Random Forest* sebagai arsitektur paling andal untuk klasifikasi postur duduk, disusul secara ketat oleh *MLP Neural Network* konvensional. Keberhasilan ini mengindikasikan bahwa ekstraksi fitur menggunakan YOLOv8n-Pose yang dipadukan dengan transformasi *flattening* mampu menghasilkan representasi data spasial murni yang sangat kuat. Penggunaan 51 atribut koordinat krusial (17 titik sendi) hasil ekstraksi YOLOv8n-Pose yang telah dinormalisasi menggunakan *Standard Scaler* memastikan kesetaraan skala antarfitur spasial, sehingga model dapat memproses informasi geometris secara optimal, tanpa memerlukan ekstraktor fitur visual berbasis piksel konvensional.

Keunggulan *Random Forest* pada penelitian ini bertumpu pada mekanisme agregasi ratusan pohon keputusan (*ensemble learning*), yang terbukti mampu menghasilkan generalisasi sangat kuat terhadap data uji meskipun sempat menunjukkan indikasi *overfitting* pada fase pelatihan. Di sisi lain, *MLP Neural Network* tetap menampilkan kemampuan yang solid dalam menemukan representasi bobot yang optimal dan adaptif pada data fitur berdimensi rendah, menjadikannya alternatif yang tangguh meski tertinggal tipis dari *Random Forest* pada hampir seluruh metrik evaluasi.

Di sisi lain, penurunan performa Deep MLP ke posisi terendah memberikan temuan akademis yang krusial terkait penerapan arsitektur *deep learning* pada data koordinat spasial. Temuan ini mengindikasikan bahwa

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

penambahan *hidden layer* berlapis justru memicu kompleksitas yang tidak perlu ketika dihadapkan pada dataset berformat tabular dengan dimensi yang terbatas (hanya 51 fitur). Alih-alih meningkatkan performa, arsitektur yang terlampau kompleks pada jumlah sampel yang relatif menengah justru menyulitkan model untuk melakukan konvergensi secara optimal, sehingga lebih rentan mengalami misklasifikasi dibandingkan algoritma yang lebih sederhana.

Secara keseluruhan, arsitektur *machine learning* berbasis *ensemble* yang ringkas, jika dipadukan dengan teknik pra-pemrosesan data yang tepat, menunjukkan performa yang lebih unggul pada penelitian ini. *Random Forest* tidak hanya unggul dari segi akurasi prediktif dan metrik F1-score pada data uji maupun metrik seimbang, tetapi juga memiliki kompleksitas komputasi yang relatif ringan dibandingkan jaringan saraf tiruan. Karakteristik komputasi yang efisien inilah yang menjadikannya kandidat utama untuk diimplementasikan sebagai *engine* klasifikasi pada sistem monitoring postur SitSmart, meskipun *MLP Neural Network* tetap menjadi alternatif yang layak dipertimbangkan mengingat performanya yang konsisten mendekati *Random Forest* di seluruh skema evaluasi.

4. KESIMPULAN

Penelitian ini telah berhasil mengembangkan sistem analisis dan klasifikasi postur duduk berbasis citra melalui integrasi algoritma YOLOv8n-Pose untuk ekstraksi titik kerangka tubuh, transformasi fitur *flattening*, serta normalisasi data koordinat menggunakan *Standard Scaler*. Pengujian empiris terhadap lima model klasifikasi menunjukkan bahwa *Random Forest* menduduki peringkat performa generalisasi tertinggi dan paling seimbang pada data uji. Model ini mencapai tingkat akurasi 0,9775 dan *Macro F1-Score* 0,9728, mengungguli model *MLP Neural Network* yang berada di posisi kedua dengan performa yang juga kuat. Perlu dicatat bahwa *Random Forest* sebelumnya menunjukkan indikasi *overfitting* pada fase pelatihan, ditandai dengan skor sempurna pada seluruh metrik data latih; namun demikian, indikasi tersebut tidak bersifat destruktif, karena model tetap mampu mempertahankan kemampuan generalisasi terbaik saat dihadapkan pada data uji. Pembagian data latih dan data uji dilakukan dengan rasio 80:20 menggunakan *stratified random sampling* melalui skema *hold-out validation*. Pembagian ini juga diakui bersifat berbasis citra dan bukan berbasis subjek, mengingat keterbatasan identitas subjek pada dataset publik yang digunakan.

Hasil evaluasi komparatif ini memberikan indikasi penting bahwa algoritma *ensemble* berbasis pohon keputusan, apabila dikombinasikan dengan teknik pra-pemrosesan data numerik yang tepat, lebih efektif dibandingkan model *deep learning* berlapis (seperti *Deep* MLP) dalam memproses dataset berformat tabular linier berdimensi rendah. Terkait peningkatan performa yang dihasilkan, perlu dicatat bahwa perbandingan tingkat akurasi antara sistem yang diusulkan dengan studi rujukan sebelumnya (seperti studi berbasis MediaPipe) bukanlah komparasi yang sepenuhnya setara. Perbedaan karakteristik dataset, variasi kondisi pengujian, serta jumlah sampel lintas studi menjadi salah satu batasan (*limitation*) dalam penelitian ini. Oleh karena itu, peningkatan batas akurasi tersebut mengindikasikan potensi metode secara komparatif, namun tidak dapat secara mutlak diklaim hanya karena faktor perubahan algoritma semata.

Sebagai luaran akhir, model *Random Forest* terbaik tersebut telah berhasil diimplementasikan ke dalam antarmuka aplikasi web bernama SitSmart dan di-*deploy* pada ekosistem *cloud Hugging Face Spaces*. Pemilihan *Random Forest* sebagai *engine* klasifikasi utama juga didasarkan pada kompleksitas komputasinya yang relatif ringan dibandingkan jaringan saraf tiruan, menjadikannya kandidat yang efisien untuk implementasi *real-time*. Berdasarkan hasil evaluasi metrik klasifikasi secara *offline*, sistem ini menunjukkan potensi komputasi yang menjanjikan untuk inferensi postur secara *real-time*. Meskipun demikian, disadari bahwa pengujian lapangan, evaluasi latensi komputasi, serta interaksi dengan pengguna nyata di lingkungan kerja (*real-world*) masih sangat diperlukan sebagai arah pengembangan penelitian selanjutnya guna memvalidasi klaim keandalan dan aplikabilitas sistem ini secara komprehensif.

DAFTAR PUSTAKA

[1] Z. Rakauskas, V. Macaitis, A. Vasjanov, and V. Barzdenas, “Ergonomic Innovation: A Modular Smart Chair for Enhanced Workplace Health and Wellness,” *Sensors*, vol. 25, no. 13, p. 4024, Jun. 2025, https://doi.org/10.3390/s25134024.

[2] B. Putsa, W. Jalayondeja, K. Mekhora, P. Bhuanantanondh, and C. Jalayondeja, “Factors associated with reduced risk of musculoskeletal disorders among office workers: a cross-sectional study 2017 to 2020,” *BMC Public Health*, vol. 22, no. 1, p. 1503, Aug. 2022, https://doi.org/10.1186/S12889-022-13940-0.

[3] A. R. Kett, T. L. Milani, and F. Sichting, “Sitting for Too Long, Moving Too Little: Regular Muscle Contractions Can Reduce Muscle Stiffness During Prolonged Periods of Chair-Sitting,” *Frontiers in Sports and Active Living*, vol. 3, p. 760533, Nov. 2021, https://doi.org/10.3389/fspor.2021.760533.

[4] J. Piruta and W. Kułak, “Physiotherapy in Text Neck Syndrome: A Scoping Review of Current Evidence and Future Directions,” *Journal of Clinical Medicine*, vol. 14, no. 4, p. 1386, Feb. 2025, https://doi.org/10.3390/jcm14041386.

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

[5] A. R. Tsantili, D. Chrysikos, and T. Troupis, "Text Neck Syndrome: Disentangling a New Epidemic," *Acta Medica Academica*, vol. 51, no. 2, pp. 123–127, Dec. 2022, https://doi.org/10.5644/ama2006-124.380.

[6] J. Leppe-Zamora, S. Ramos-Fuster, B. Muñoz-Monari, S. Roa-Alcaino, and O. L. Sarmiento, "The effect of computer prompt in breaks of sedentary behaviour among office workers: a systematic review and meta-analysis," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 22, no. 1, p. 75, Jun. 2025, https://doi.org/10.1186/s12966-025-01781-0.

[7] J. Lee, H. Choi, and J. Kim, "Classification Algorithm for Sitting Postures Using Weighted Random Forest," *IET Image Processing*, vol. 19, no. 1, p. e70126, Jan. 2025, https://doi.org/10.1049/ipr2.70126.

[8] R. Zemp *et al.*, "Application of Machine Learning Approaches for Classifying Sitting Posture Based on Force and Acceleration Sensors," *BioMed Research International*, vol. 2016, no. 1, p. 5978489, Jan. 2016, https://doi.org/10.1155/2016/5978489.

[9] J. Ahmad, J. Sidén, and H. Andersson, "A Proposal of Implementation of Sitting Posture Monitoring System for Wheelchair Utilizing Machine Learning Methods," *Sensors*, vol. 21, no. 19, p. 6349, Sep. 2021, https://doi.org/10.3390/S21196349.

[10] C. Dong and G. Du, "An enhanced real-time human pose estimation method based on modified YOLOv8 framework," *Scientific Reports*, vol. 14, no. 1, pp. 8012-, Apr. 2024, https://doi.org/10.1038/s41598-024-58146-z.

[11] T. Lu, K. Cheng, X. Hua, and S. Qin, "KSL-POSE: A Real-Time 2D Human Pose Estimation Method Based on Modified YOLOv8-Pose Framework," *Sensors*, vol. 24, no. 19, p. 6249, Sep. 2024, https://doi.org/10.3390/s24196249.

[12] F. Wang, G. Wang, and B. Lu, "YOLOv8-PoseBoost: Advancements in Multimodal Robot Pose Keypoint Detection," *Electronics*, vol. 13, no. 6, p. 1046, Mar. 2024, https://doi.org/10.3390/electronics13061046.

[13] S. Cai, H. Xu, W. Cai, Y. Mo, and L. Wei, "A human pose estimation network based on YOLOv8 framework with efficient multi-scale receptive field and expanded feature pyramid network," *Scientific Reports*, vol. 15, no. 1, pp. 15284-, May 2025, https://doi.org/10.1038/s41598-025-00259-0.

[14] C. Singhtaun, S. Natsupakpong, and P. Lorprasertkul, "Ergonomic Risk Assessment Using Human Pose Estimation with MediaPipe Pose," *ACM International Conference Proceeding Series*, vol. 1, pp. 465–471, Jul. 2025, https://doi.org/10.1145/3719384.3719453.

[15] D. F. Odesola, J. Kulon, S. Verghese, A. Partlow, and C. Gibson, "A Smart System for Continuous Sitting Posture Monitoring, Assessment, and Personalized Feedback," *Sensors*, vol. 25, no. 18, p. 5610, Sep. 2025, https://doi.org/10.3390/S25185610.

[16] H. Q. Chu, "Sitting-Posture-Recognition," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/huyquangchu/test-dataset.

[17] Min, "Sitting Posture v2 Computer Vision Dataset," Roboflow. [Online]. Available: https://universe.roboflow.com/min-bnvzi/sitting-posture-v2.

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi