

STUDI KLASIFIKASI GERAK DUDUK-BERDIRI DENGAN ALGORITMA TRANSFORMER

Ferdinand Vincent Halim ¹, Jason Christov Kesuma ¹, Benedictus Marcello Nathanael ²,
Rivian Naufal Dzakki Pradana ², Romy Budhi Widodo ¹

¹ Pusat Studi Human-Machine Interaction-Teknik Informatika, Universitas Ma Chung

² Teknik Informatika, Universitas Ma Chung
312310022@student.machung.ac.id

ABSTRAK

Gerakan Sit-to-Stand (STS) merupakan aktivitas fungsional penting yang sering digunakan untuk mengevaluasi kemampuan mobilitas dan keseimbangan, khususnya pada kelompok lanjut usia. Informasi mengenai fase-fase gerakan STS diperlukan untuk mendukung pengembangan sistem pemantauan dan alat bantu mobilitas yang lebih aman. Penelitian ini mengusulkan penggunaan algoritma Transformer untuk mengklasifikasikan empat fase gerakan STS, yaitu Sit, Flexion, Extension, dan Stand, berdasarkan data sensor Inertial Measurement Unit (IMU). Data dikumpulkan dari delapan partisipan menggunakan dua sensor IMU yang dipasang pada korpus sterni dan tibia dengan frekuensi pengambilan data 100 Hz. Dataset terdiri atas 32.000 sampel dengan empat fitur utama berupa nilai roll dan pitch dari kedua sensor. Sebelum proses pelatihan, data melalui tahapan preprocessing yang meliputi penanganan missing value, winsorization, normalisasi min-max, encoding label, dan pembentukan data sekuensial menggunakan metode sliding window sepanjang 30 timestep. Model Transformer dikembangkan menggunakan TensorFlow dan dievaluasi menggunakan metode 10-Fold Stratified Cross Validation. Hasil penelitian menunjukkan bahwa model mampu mencapai akurasi sebesar 94%, meningkat signifikan dibandingkan pendekatan Artificial Neural Network (ANN) berbasis Multi-Layer Perceptron yang sebelumnya dengan akurasi 72%. Temuan ini menunjukkan bahwa mekanisme Self-Attention pada Transformer efektif dalam menangkap dependensi temporal data sensor sehingga mampu meningkatkan performa klasifikasi fase gerakan duduk-berdiri.

Kata Kunci : *Transformer Encoder, Sit-to-Stand, Inertial Measurement Unit, Human Activity Recognition, Klasifikasi Gerakan*

1. PENDAHULUAN

Seiring dengan proses penuaan, kelompok lanjut usia (lansia) sering mengalami penurunan fungsi gerak otot dan keseimbangan yang menyulitkan mereka untuk melakukan aktivitas esensial seperti transisi duduk-ke-berdiri atau *Sit-to-Stand* (STS). Gerakan STS bukanlah aktivitas sederhana, melainkan melibatkan pergeseran titik berat tubuh (*Center of Gravity / CoG*) yang terbagi menjadi empat fase, yaitu *Sit* (duduk), *Flexion* (titik berat bergerak maju), *Extension* (titik berat bergerak naik), dan *Stand* (berdiri) [1]. Informasi akurat mengenai fase mana lansia mengalami kesulitan sangat dibutuhkan untuk mengembangkan alat bantu mobilitas medis.

Berbagai upaya telah dilakukan oleh peneliti medis untuk menganalisis fase STS ini, diantaranya dengan menggunakan *pressure plate* yang dilengkapi *Electromyography* (EMG) [2] hingga penggunaan kombinasi sensor inersia dan kamera inframerah [3]. Pada perkembangannya, penelitian terbaru mulai mengimplementasikan *Machine Learning* untuk mendeteksi fase tersebut secara otomatis dan *real-time* [4] [5]. Salah satu implementasi yang telah dilakukan adalah pemanfaatan algoritma *Artificial Neural Network* (ANN) *Multi-Layer Perceptron* yang memproses data masukan dari dua buah sensor *Inertial Measurement Unit* (IMU) pada dada (*Korpus Sterni*)

dan betis (*Tibia*) [6]. Meskipun sistem berbasis ANN tersebut mampu beroperasi secara *real-time* dari data rekaman yang diambil dengan kecepatan rekam 100 data per detik (100 Hz), evaluasi metrik kinerja model menunjukkan kelemahan yang kritis [7] [8]. Model ANN tersebut mengalami kesulitan besar dalam mengklasifikasikan fase *Extension* secara akurat, terbukti dari nilai sensitivitas (*recall*) yang paling rendah di antara fase lainnya, yakni sebesar 55%.

Keterbatasan ANN berjenis *Multi-Layer Perceptron* (MLP) dalam memodelkan ketergantungan temporal jangka panjang menuntut adanya arsitektur algoritma baru yang memiliki memori sekuensial yang lebih unggul. Pemilihan metode *Transformer* sebagai pengganti ANN secara mendasar didasarkan pada kapabilitas strukturalnya dalam memecahkan masalah pemrosesan data deret waktu (*time-series*). Berbeda dengan ANN konvensional yang memproses aliran data titik-demi-titik secara independen tanpa mengikat konteks riwayat gerakan sebelumnya, arsitektur *Transformer* tidak dibatasi oleh komputasi yang berurutan secara kaku. Sebaliknya, metode ini menggunakan pendekatan terintegrasi yang memungkinkan untuk mengamati dan mengevaluasi keseluruhan rentetan data historis pergerakan sensor secara bersamaan [9]. Keunggulan inilah yang menjadikannya sangat ideal untuk membedakan

pola dinamika sensor yang kompleks antar satu fase gerakan dengan fase lainnya.

Untuk menutupi kesenjangan pada penelitian [6] tersebut, penelitian ini mengusulkan penerapan metode pembelajaran mendalam berbasis *Transformer* sebagai pengganti *ANN*. Berbeda dengan jaringan saraf tiruan konvensional, *Transformer* memiliki mekanisme *Self-Attention* yang secara khusus sangat kuat dalam memproses dan menemukan korelasi urutan data *time-series* yang panjang [10]. Mekanisme ini diharapkan memungkinkan model untuk secara lebih akurat membaca konteks peralihan dari gerakan condong ke depan (*Flexion*) menuju gerakan naik (*Extension*), yang pada akhirnya akan secara signifikan meningkatkan nilai *recall* pada fase *Extension* dan menjamin keamanan alat bantu bagi para lansia.

2. TINJUAN PUSTAKA

Pemodelan data sekuensial mengalami perubahan paradigma yang besar sejak diperkenalkannya arsitektur *Transformer* oleh Vaswani et al. dalam makalah berjudul *Attention Is All You Need* [9]. Berbeda dengan model pendahulunya yang menggunakan *Recurrent Neural Networks (RNN)* atau *Convolutional Neural Networks (CNN)*, arsitektur ini beralih sepenuhnya ke mekanisme *self-attention*. Bagian inti dari sistem ini digerakkan oleh fungsi *Scaled Dot-Product Attention*, yang didefinisikan melalui persamaan:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Dengan penjelasan variabel:

Q (query): Mewakili kata atau token saat ini yang sedang mencari konteks.

K (key): Mewakili semua kata dalam urutan yang berfungsi sebagai titik acuan.

V (value): Berisi konten semantik atau makna aktual dari kata-kata tersebut.

$\sqrt{d_k}$: Faktor skala berdasarkan dimensi *key* (d_k) untuk mencegah angka menjadi terlalu besar dan menjaga gradien tetap stabil, dan

softmax: Mengubah skor kemiripan menjadi nilai probabilitas yang totalnya berjumlah 1.

Melalui persamaan tersebut, *Transformer* mampu mendeteksi hubungan jarak jauh antar-elemen dalam suatu urutan dengan sangat efisien. Selain itu, formula ini mendukung proses paralelisasi data, sehingga waktu pelatihan menjadi jauh lebih singkat dibandingkan dengan model berbasis *RNN*.

Pemanfaatan arsitektur *Transformer* kemudian diperluas ke ranah pengenalan bahasa isyarat oleh Camgoz et al. [13] melalui pengembangan *Sign Language Transformer (SLT)*. Sistem end-to-end yang mereka rancang mampu mengeksekusi proses *Continuous Sign Language Recognition (CSLR)*

dan *Sign Language Translation (SLT)* secara simultan. Berkat penerapan komponen *encoder-decoder Transformer*, model ini berhasil mencetak performa state-of-the-art pada dataset penanda benchmark *RWTH-PHOENIX-Weather-2014T*.

Penerapan arsitektur *Transformer* juga ditemukan pada penelitian Chaudhary et al. [14], yang mengembangkan *SignNet II* sebagai model penerjemahan bahasa isyarat dua arah (*sign-to-text* dan *text-to-sign*). Melalui integrasi mekanisme *dual learning* serta *metric embedding learning*, pendekatan ini mampu mendongkrak kualitas translasi secara signifikan, terutama dalam menjaga konsistensi antartanda pada representasi pose. Temuan tersebut kian menegaskan peran *Transformer* sebagai arsitektur yang adaptif dan efektif untuk pemrosesan multimodal, yang kelak akan diimplementasikan dalam klasifikasi STS ini.

Penyisipan *transformer* diharapkan mampu memberikan peningkatan performa, khususnya dalam menangkap pola *spatio-temporal* yang kompleks pada data gerakan STS, sekaligus menjadi kontribusi yang selaras dengan tren penelitian terkini.

3. METODE PENELITIAN

3.1. Pengambilan Data

Tahap pertama dalam pengambilan data adalah dengan melakukan pemasangan sensor pada bagian tubuh subjek dan sinkronisasi peralatan dengan aplikasi. Pengambilan data dilakukan dengan menggunakan sensor *inertial measurement unit (IMU)* tipe MTw Awinda dari Xsens Technologies B.V dengan *sampling rate* sebesar 100 Hz. MTw Awinda dilengkapi dengan sebuah stasiun yang berfungsi sebagai penerima (*receiver*) dan pengisi daya baterai IMU. Pengambilan data ini menggunakan satu buah laptop dan sebuah kursi. Perangkat lunak yang digunakan adalah MT Manager, Microsoft Excel, dan Google Colab.

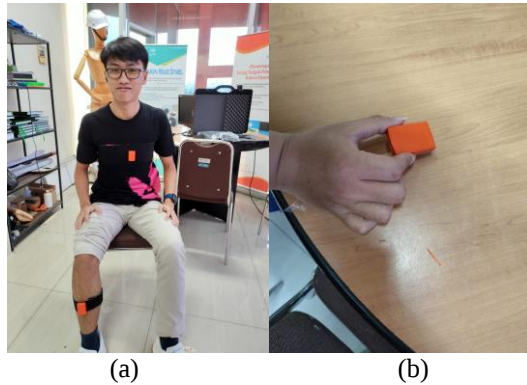
3.2. Subject

Data dikumpulkan dari delapan volunteer mahasiswa yang terdiri dari dua wanita dan enam pria berusia 20-25 tahun.

3.3. Penempatan Sensor

Tiga buah IMU digunakan selama pengambilan data. Dua IMU dipasang pada bagian tubuh Korpus Sterni (tulang dada) dan Tibia (tulang kering) seperti pada Gambar 1a. Sedangkan IMU ke tiga sebagai penanda pergantian fase, dipegang tangan seperti yang diilustrasikan pada Gambar 1b. Sensor pertama dan kedua mengukur Pitch dan Roll dari orientasi dada dan tulang kering. Sensor ketiga yang dipegang tangan akan digerakkan secara manual oleh peneliti, sambil mengamati Subjek yang sedang diujicoba. Saat pergantian fase dari Sit-Flexion-Extension-Stand; diantara pergantian fase-fase tersebut sensor ketiga ini digoyang-goyangkan

sebagai penanda. Hasil bacaan sensor ketiga pada spreadsheet akan dipakai untuk proses pelabelan kelas output.



Gambar 1. (a) Penempatan sensor di tulang dada dan tulang kering, (b) IMU penanda fase.

Pemasangan sensor dengan tali Velcro (perekat) pada Korpus Sterni dan Tibia dilakukan dengan pertimbangan bahwa kedua bagian tersebut melakukan kontraksi otot yang minimal, sehingga dapat memastikan bahwa MTw terpasang dengan kuat dan kokoh pada kulit. Adapun sensor ketiga digunakan sebagai penanda dari tiap pergerakan fase. Menurut Jin et.al dalam [1], pemasangan sensor pada dada akan menjadi yang terbaik karena tingkat kemiringan tubuh subjek dan tibia. Selain itu, pemasangan sensor pada Korpus Sterni dan Tibia, seperti yang diilustrasikan pada Gambar 1a, menyebabkan titik berat (*Center of Gravity/CoG*) bergerak maju selama perpindahan dari fase Duduk (Sit) ke fase Fleksi (Flexion), yang mengarah pada perubahan sudut Korpus Sterni dan Tibia. Pada postur ini, posisi Korpus Sterni dan Tibia condong ke depan dan membentuk sudut dengan sumbu-y. Penempatan sensor pada Korpus Sterni juga dilakukan oleh peneliti lain pada [2].

3.4. Proses Pengembangan Model Transformer

Proses pengembangan model Transformer diawali dengan proses pelabelan data dan preprocessing data. Model Transformer dikembangkan menggunakan bahasa pemrograman Python dengan framework Tensorflow dan Keras. Dataset yang digunakan untuk proses pelatihan terdiri atas 32.000 sampel yang dikumpulkan dari delapan subjek. Dataset memiliki empat fitur utama (roll_dada, pitch_dada, roll_betis, dan pitch_betis) dan empat kelas target (Sit, Flexion, Extension, Stand). Berbeda dengan pendekatan ANN (MLP) pada penelitian sebelumnya [6], model Transformer memerlukan tahap preprocessing tambahan berupa Sliding Window untuk mengubah data tabular per-frame menjadi input sekuensial guna memanfaatkan dependensi temporal antar timestep melalui mekanisme Self-Attention. Contoh ilustrasi data pelatihan ditampilkan pada Tabel 1.

Tabel 1. Data sensor IMU dari dataset penelitian

roll_dada (°)	pitch_dada (°)	roll_betis (°)	pitch_betis (°)	phase-class
-0.37	-0.36	0.05	0.08	Sit
-0.36	-0.36	0.05	0.09	Sit
-0.36	-0.35	0.05	0.08	Sit
:	:	:	:	:
-12.35	-1.55	-1.76	-0.39	Flexion
-12.36	-1.56	-1.76	-0.38	Flexion
-12.35	-1.54	-1.77	-0.36	Flexion
:	:	:	:	:
-29.85	-27.24	-2.36	3.71	Extension
-29.91	-27.17	-2.34	3.69	Extension
-29.87	-26.99	-2.29	3.66	Extension
:	:	:	:	:
-27.34	-8.99	-4.16	-1.54	Stand
-27.19	-8.67	-4.24	-1.55	Stand
-27.04	-8.37	-4.31	-1.55	Stand
:	:	:	:	:

3.5. Preprocessing Data untuk Transformer

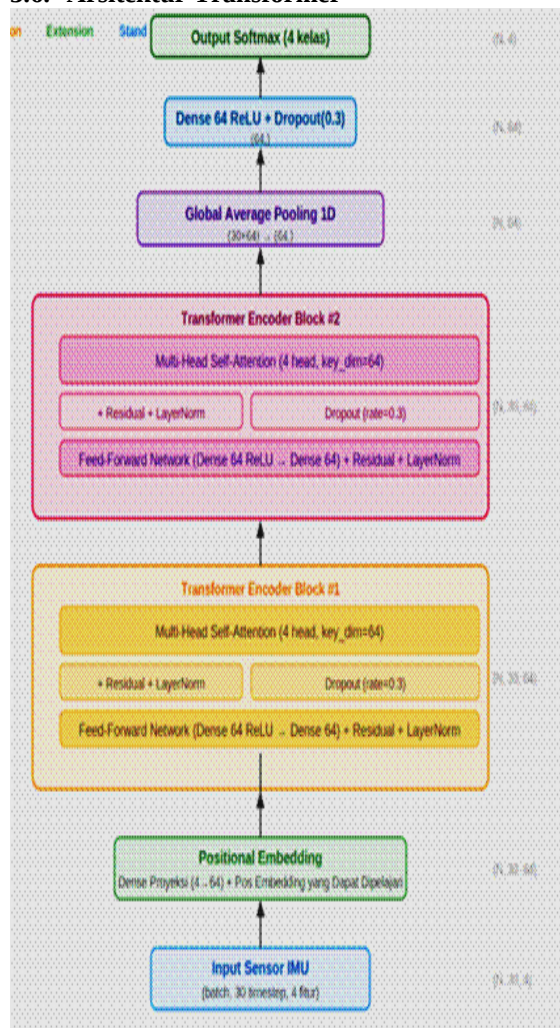
Pipeline preprocessing diadaptasi dari [6] dengan penambahan satu tahap baru untuk mendukung input sekuensial yang dibutuhkan arsitektur Transformer. Tabel 1 menunjukkan bahwa dataset memiliki empat fitur utama, yaitu roll_dada, pitch_dada, roll_betis, dan pitch_betis. Pipeline preprocessing secara lengkap terdiri dari enam tahap sebagai berikut:

- Pemanggilan dan penggabungan dataset: Tiga file Excel dataset sensor dimuat dan digabungkan, menghasilkan 32.000 sampel dari delapan subjek dengan distribusi kelas Sit, Flexion, Extension, dan Stand.
- Penanganan *missing value*: Baris yang mengandung nilai kosong dihapus menggunakan fungsi `dropna()`.
- Winsorize (penanganan outlier): Diterapkan per-partisipan dengan batas bawah 10% dan batas atas 15%, konsisten dengan metode yang digunakan pada [6]. Hal ini untuk menghilangkan nilai ekstrem yang dapat mengganggu proses pelatihan.
- Normalisasi min-max: Nilai fitur diskalakan ke rentang [0,1] menggunakan scaler yang di-fit pada seluruh data gabungan dan diterapkan per-partisipan. Normalisasi sangat penting agar mekanisme Self-Attention pada Transformer bekerja pada skala fitur yang seragam.
- Encoding label: Label fase berbentuk string diubah menjadi integer. Extension=0, Flexion=1, Sit=2, Stand=3.
- Sliding window: Data tabular per-frame diubah menjadi jendela sekuensial dengan panjang $T=30$ timestep dan langkah geser (step) sebesar 1 (overlap maksimal). Windowing dilakukan per-partisipan untuk mencegah kontaminasi temporal lintas partisipan. Setiap window menghasilkan satu sampel berukuran (30, 4) dan label ditentukan berdasarkan fase pada timestep terakhir dalam window tersebut. Langkah ke-6

ini adalah kontribusi baru dibandingkan penelitian sebelumnya [6].

Transformasi sliding window mengubah N baris data tabular menjadi $N-29$ window sekuensial per partisipan. Jumlah ini diperoleh dari rumus umum jumlah window = $N - T + 1$, dengan panjang window $T=30$ dan langkah geser (step) sebesar 1, sehingga $N - 30 + 1 = N - 29$. Penggunaan step = 1 menghasilkan overlap maksimal, yaitu setiap dua window yang berdekatan berbagi 29 dari 30 timestep; strategi ini memperbanyak jumlah sampel pelatihan sekaligus menjaga kontinuitas informasi temporal antar window. Setiap window berbentuk tensor dua dimensi berukuran (30, 4), yang merepresentasikan 30 timestep berurutan dengan 4 fitur orientasi IMU (roll_dada, pitch_dada, roll_betis, pitch_betis) pada masing-masing timestep. Label setiap window ditentukan berdasarkan fase pada timestep terakhir, dengan asumsi bahwa fase pada akhir jendela merepresentasikan kondisi gerakan terkini yang ingin diklasifikasikan.

3.6. Arsitektur Transformer



Gambar 2. Arsitektur Transformer untuk klasifikasi fase STS

Arsitektur Transformer diimplementasikan berdasarkan [7]. Model Transformer dilengkapi mekanisme Multi-Head Self-Attention (MHSA), sehingga memungkinkan model menangkap dependensi temporal di seluruh timestep dalam sebuah window secara bersamaan. Arsitektur keseluruhan ditampilkan pada Gambar 2.

- Input Layer: Ukuran (30, 4) — 30 timestep dengan empat fitur orientasi IMU per timestep.
- Positional Embedding: Proyeksi Dense yang dapat dipelajari memetakan empat fitur input ke ruang dimensi $d_{\text{model}}=64$. Kemudian ditambahkan vektor positional embedding yang dapat dipelajari (*learnable*) per timestep untuk memberikan informasi urutan temporal kepada model. Hal ini diperlukan karena mekanisme Self-Attention bersifat *permutation-invariant* (tidak mengetahui urutan secara inheren).
- Transformer Encoder Block x2: Setiap blok terdiri atas “
 - Multi-Head Self-Attention (MHSA) dengan empat head dan $\text{key_dim}=64$
 - Residual connection+Layer Normalization
 - Feed-Forward Network (FFN) dengan dimensi tersembunyi $\text{ff_dim}=64$ dan aktivasi ReLU
 - Residual connection+Layer Normalization
 - Dropout (rate=0,3) setelah setiap sub-layer.
- Global Average Pooling 1D: Merata-ratakan output dari seluruh 30 timestep menjadi satu vektor representasi berdimensi 64.
- Dense Layer tersembunyi: Lapisan fully-connected dengan 64 unit dan aktivasi ReLU, diikuti Dropout (0,3).
- Output Layer: Lapisan Dense dengan empat unit dan aktivasi Softmax, menghasilkan probabilitas kelas untuk keempat fase STS.

Sedangkan konfigurasi hiperparameter lengkap model Transformer ditampilkan pada Tabel 2.

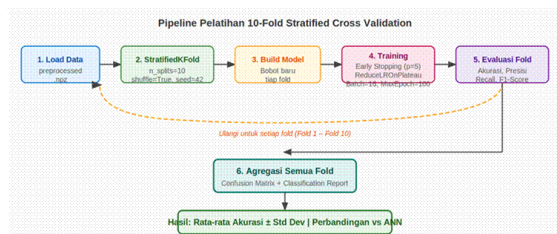
Tabel 2. Konfigurasi hiperparameter Transformer

Hiperparameter	Nilai	Keterangan
Panjang sekuens (T)	30	Timestep per window
Jumlah fitur	4	Roll dan pitch pada dada dan betis
Dimensi model (d_{model})	64	Dimensi embedding
Jumlah head attention	4	Multi-Head Self-Attention
Key dimension (key_dim)	64	Dimensi per head attention
Dimensi FFN (ff_dim)	64	Unit tersembunyi Feed-Forward
Jumlah encoder block	2	Transformer block bertumpuk
Dropout rate	0,3	Regularisasi dropout
Jumlah kelas output	4	Sit, Flexion, Extension, Stand

Hiperparameter	Nilai	Keterangan
Learning rate	0,001	Adam optimizer LR awal
Batch size	16	Ukuran batch pelatihan
Maksimum epoch	100	Dengan Early Stopping (patience=5)

3.7. Prosedur Pelatihan dan Evaluasi Model

Model dilatih menggunakan 10-Fold Stratified Cross Validation (K-Fold CV) untuk mengevaluasi kemampuan generalisasi dan memastikan konsistensi metodologi evaluasi dengan [6]. Stratified K-Fold dari scikit-learn digunakan dengan parameter `shuffle=True` dan `random_state=42` untuk menjaga distribusi kelas yang proporsional di setiap fold. Diagram pipeline pelatihan ditampilkan pada Gambar 3.



Gambar 3. Pipeline pelatihan 10-Fold Stratified Cross Validation untuk Transformer Encoder

Untuk setiap fold, model baru dibangun dari awal dengan bobot yang diinisialisasi ulang. Dua callback digunakan selama pelatihan. Pertama, Early Stopping memonitor `val_loss` dengan patience sebesar 5 epoch dan memulihkan bobot terbaik saat pelatihan dihentikan guna mencegah overfitting. Kedua, `ReduceLROnPlateau` mengurangi learning rate sebesar faktor 0,5 apabila `val_loss` tidak membaik selama 3 epoch berturut-turut, dengan learning rate minimum sebesar $1e-6$.

Model dikompilasi menggunakan optimizer Adam ($lr=0,001$) dan fungsi loss `sparse_categorical_crossentropy`, dengan akurasi sebagai metrik evaluasi. Prediksi dihasilkan melalui operasi `argmax` pada vektor probabilitas Softmax. Metrik evaluasi per-fold dan agregat (akurasi, presisi, recall, F1-score) dihitung dan dibandingkan terhadap baseline ANN [6] (akurasi training: 86%, akurasi testing: 72%). Akurasi training per-fold dicatat pada epoch terbaik (`val_loss` minimum) sebagaimana ditentukan oleh Early Stopping. Confusion matrix dibuat dengan mengagregasi prediksi dari semua 10 fold, dan laporan klasifikasi per kelas dihasilkan untuk analisis performa yang lebih rinci.

Fungsi loss yang digunakan adalah `sparse_categorical_crossentropy` yang setara dengan `categorical_crossentropy` pada ANN baseline,

dengan perbedaan bahwa `sparse_categorical_crossentropy` menerima label integer secara langsung tanpa memerlukan one-hot encoding eksplisit.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Hasil penelitian menunjukkan bahwa model Transformer mampu mengklasifikasikan empat fase Sit-to-Stand (STS), yaitu Sit, Flexion, Extension, dan Stand dengan akurasi keseluruhan sebesar 94%. Nilai tersebut menunjukkan peningkatan yang signifikan dibandingkan pendekatan Artificial Neural Network (ANN) pada penelitian [6] yang memperoleh akurasi pengujian sebesar 72%. Peningkatan performa ini menunjukkan bahwa pemodelan temporal melalui mekanisme Self-Attention mampu menangkap hubungan antar-timestep secara lebih efektif dibandingkan pendekatan per-frame yang digunakan oleh Multi-Layer Perceptron (MLP).

Secara rinci, fase Flexion memperoleh performa terbaik dengan F1-score sebesar 0,96, sedangkan fase Sit dan Extension memperoleh F1-score masing-masing sebesar 0,94 dan 0,93. Fase Stand memiliki performa paling rendah dengan F1-score sebesar 0,91, namun masih berada pada kategori sangat baik. Peningkatan paling penting terjadi pada fase Extension yang sebelumnya menjadi kelemahan pada [6].

4.2. Pembahasan

Perbandingan hasil penelitian ini dengan [6] yang menggunakan ANN berbasis MLP pada dataset STS. Penelitian tersebut menghasilkan akurasi pengujian sebesar 72%, sedangkan penelitian ini mencapai 94%. Perbandingan kedua dilakukan terhadap [11] yang menerapkan Transformer pada Human Activity Recognition berbasis wearable sensor dan memperoleh akurasi sekitar 99,2%. Hasil tersebut memperkuat bahwa Transformer sangat efektif dalam menangani data sensor time-series. Perbandingan ketiga dilakukan terhadap [12] yang menggunakan Transformer-based untuk pengenalan aktifitas manusia menghasilkan akurasi 93,69%. Hasil penelitian ini yang mencapai 94% menunjukkan performa yang sangat kompetitif dan konsisten dengan tren penelitian terkini.

4.3. Keterbatasan Penelitian

Penelitian ini masih memiliki beberapa keterbatasan. Pertama, jumlah subjek yang digunakan delapan orang sehingga belum merepresentasikan populasi yang lebih luas. Kedua, seluruh subjek berasal dari kelompok usia muda sehingga kemampuan generalisasi terhadap lansia belum dapat dipastikan. Ketiga, proses pelabelan masih dilakukan secara semi-manual menggunakan sensor ketiga sehingga berpotensi menimbulkan

variasi subjektif. Keempat, model baru diuji secara offline menggunakan 10-Fold Stratified Cross Validation dan belum diimplementasikan pada sistem real-time.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan algoritma Transformer untuk mengklasifikasikan empat fase gerakan Sit-to-Stand (Sit, Flexion, Extension, dan Stand) menggunakan data sensor IMU pada korpus sterni dan tibia. Dengan memanfaatkan mekanisme Self-Attention dan teknik sliding window, model mampu menangkap hubungan temporal antar-timestep secara lebih efektif dibandingkan ANN berbasis Multi-Layer Perceptron. Hasil evaluasi menggunakan 10-Fold Stratified Cross Validation menunjukkan akurasi sebesar 94%, meningkat signifikan dibandingkan penelitian sebelumnya yang mencapai 72%. Peningkatan performa ini menunjukkan bahwa Transformer merupakan pendekatan yang efektif dan menjanjikan untuk klasifikasi fase gerakan duduk-berdiri berbasis sensor IMU serta berpotensi diterapkan pada sistem pemantauan mobilitas dan pengembangan alat bantu bagi lansia.

Penelitian selanjutnya disarankan menambah jumlah partisipan yang melibatkan lansia, mengotomatisasi proses pelabelan, serta mengimplementasikan Transformer pada aplikasi real-time.

DAFTAR PUSTAKA

- [1] F. Jin, T. Nagasaki, and C. Wada, "An estimation of knee and ankle joint angles during extension phase of standing up motion performed using an inertial sensor," *Journal of Physical Therapy Science*, vol. 29, p. 1172, 2017.
- [2] C. Wada and S. Ino, "Study of the Relationship between Sit-to-Stand Activity and Seat Orientation," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 21, no. 2, p. 337, 2017.
- [3] A. Gómez *et al.*, "Analysis of sit-to-stand biomechanics using inertial sensors and infrared cameras," *Journal of Biomechanics*, 2021.
- [4] J. C. Farah *et al.*, "Machine learning for realtime sit-to-stand detection using wearable sensors," *Sensors*, 2019.
- [5] A. Arifin *et al.*, "Activity recognition including sit-to-stand phases using Multi-Layer Perceptron and IMU sensors," *IEEE Access*, 2021.
- [6] H.D. Winataharto, "Rancang Bangun Aplikasi Klasifikasi Gerakan Duduk-Berdiri Menggunakan Artificial Neural Network Secara Realtime," Program Studi Teknik Informatika, Universitas Ma Chung, Malang, 2022.
- [7] J. Wang *et al.*, "Deep learning for sensor-based human activity recognition: Overview and future challenges," *Pattern Recognition*, 2019.
- [8] N. Millor *et al.*, "Kinematic parameters to evaluate sit-to-stand using an inertial sensor," *Gait & Posture*, 2014.
- [9] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [10] Q. Wen *et al.*, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125*, 2022.
- [11] Dirgová Luptáková I, Kubovčík M, Pospíchal J. Wearable Sensor-Based Human Activity Recognition with Transformer Model. *Sensors (Basel)*. Mar 1;22(5):1911. doi: 10.3390/s22051911. PMID: 35271058; PMCID: PMC8914677, 2022
- [12] Oussema Kirmani, Sazia Mahfuz. Transformer-Based Human Activity Recognition Using Sensor Data. *Procedia Computer Science*. Volume 280. 2026.
- [13] Camgoz, N.C., Koller, O., Hadfield, S. and Bowden, R., Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10023-10033), 2020.
- [14] Chaudhary, L., Ananthanarayana, T., Hoq, E. and Nwogu, I., Signnet ii: A transformer-based two-way sign language translation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11), pp.12896-12907, 2022.