

**ANALISIS *BIG DATA* SINYAL *FACIAL SURFACE*
ELECTROMYOGRAPHY MENGGUNAKAN APACHE SPARK PADA
KOMPUTER TERDISTRIBUSI UNTUK PENINGKATAN AKURASI
IDENTIFIKASI POLA EMOSI**



SKRIPSI

diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar
Sarjana Teknik pada Program Studi Mekatronika dan Kecerdasan Buatan

Oleh:

Auziah Mumtaz

2103554

**PROGRAM STUDI MEKATRONIKA DAN KECERDASAN BUATAN
KAMPUS UPI DI PURWAKARTA
UNIVERSITAS PENDIDIKAN INDONESIA**

2025

LEMBAR HAK CIPTA

ANALISIS *BIG DATA* SINYAL *FACIAL SURFACE* *ELECTROMYOGRAPHY* MENGGUNAKAN APACHE SPARK PADA KOMPUTER TERDISTRIBUSI UNTUK PENINGKATAN AKURASI IDENTIFIKASI POLA EMOSI

Oleh

Auziah Mumtaz

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Teknik pada Program Studi Mekatronika dan Kecerdasan Buatan

© Auziah Mumtaz 2025

Universitas Pendidikan Indonesia

Juli 2025

Hak Cipta dilindungi undang-undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa ijin dari penulis.

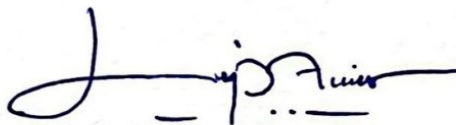
LEMBAR PENGESAHAN

AUZIAH MUMTAZ

**ANALISIS *BIG DATA* SINYAL *FACIAL SURFACE*
ELECTROMYOGRAPHY MENGGUNAKAN *APACHE SPARK* PADA
KOMPUTER TERDISTRIBUSI UNTUK PENINGKATAN AKURASI
IDENTIFIKASI POLA EMOSI**

disetujui dan disahkan oleh pembimbing:

Pembimbing I



Liptia Venica, S.T., M.T.
NIP. 920210919941203201

Pembimbing II



Dewi Indriati Hadi Putri, S. Pd., M. T.
NIP. 920190219900126201

Mengetahui

**Ketua Program Studi Mekatronika dan
Kecerdasan Buatan**



Dewi Indriati Hadi Putri, S. Pd., M. T.
NIP. 920190219900126201

ABSTRAK

Identifikasi emosi berbasis sinyal sEMG wajah menghadapi tantangan akibat volume data yang besar dan kompleksitas sinyal fisiologis. Penelitian ini mengembangkan pendekatan analisis *big data* menggunakan kerangka kerja Apache Spark dalam lingkungan komputasi terdistribusi untuk mengolah keseluruhan *dataset* EmgDataVR ($\pm 30,4$ juta baris). Jenis penelitian ini adalah eksperimental dengan pendekatan *mixed-method* dan model *iterative-incremental development* (IID). Metode mencakup *preprocessing* terdistribusi (feature selection, row filtering, dan deteksi outlier), serta penerapan algoritma *clustering* K-Means dan Bisecting K-Means. Evaluasi dilakukan dengan metrik *silhouette coefficient* untuk menilai kualitas pemisahan *cluster*. Konfigurasi terbaik diperoleh dari Bisecting K-Means dengan dua *cluster* (*silhouette* = 0,8639). Pola hasil *clustering* mengindikasikan dua kondisi emosi berdasarkan tingkat *arousal*: *low arousal* (aktivitas otot rendah) dan *high arousal* (aktivitas signifikan pada *Zygomaticus* dan *Orbicularis*, sesuai AU12 dan AU6). Temuan ini membuktikan bahwa pendekatan komputasi terdistribusi dapat meningkatkan efisiensi pemrosesan, akurasi hasil, dan skalabilitas sistem identifikasi emosi berbasis sinyal sEMG wajah.

Kata Kunci: sEMG, identifikasi emosi, *big data*, Apache Spark, *clustering*, K-Means, Bisecting K-Means

ABSTRACT

Facial emotion recognition based on surface electromyography (sEMG) signals faces significant challenges due to the large volume of data and the complexity of physiological signals. This study aims to develop a big data analysis approach using the Apache Spark framework in a distributed computing environment to process the entire EmgDataVR dataset ($\pm 30,4$ million rows). This is an experimental research employing a mixed-method approach, developed under the iterative-incremental development (IID) model. The proposed method includes a distributed preprocessing pipeline (feature selection, null filtering, and outlier detection), followed by the application of two clustering algorithms: K-Means and Bisecting K-Means. The clustering performance is evaluated using the silhouette coefficient metric to measure the quality of cluster separation. The Bisecting K-Means configuration with two clusters yields the most optimal result with a silhouette score of 0,8639. The resulting cluster patterns indicate two primary emotional states based on arousal levels: low arousal (low muscle activation) and high arousal (strong activation of the Zygomaticus and Orbicularis muscles, corresponding to AU12 and AU6). These findings demonstrate that a distributed computing approach can significantly improve processing efficiency, result accuracy, and the scalability of emotion identification systems based on sEMG signals.

Keywords: *sEMG, emotion recognition, big data, Apache Spark, distributed computing, clustering, K-Means, Bisecting K-Means*

DAFTAR ISI

SKRIPSI	i
LEMBAR HAK CIPTA	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN BEBAS PLAGIARISME	iv
UCAPAN TERIMA KASIH	vi
ABSTRAK	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvii
BAB I PENDAHULUAN	19
1.1 Latar Belakang	19
1.2 Rumusan Masalah	22
1.3 Batasan Penelitian	23
1.4 Tujuan Penelitian.....	24
1.5 Manfaat Penelitian	25
1.6 Struktur Penulisan Skripsi.....	26
BAB II TINJAUAN PUSTAKA	28
2.1 Emosi dan Ekspresi Wajah	28
2.1.1 Definisi dan Klasifikasi Emosi.....	28

2.1.2	Ekspresi Wajah sebagai Indikator Emosi	29
2.2	Elektromiografi Wajah	34
2.2.1	Prinsip Dasar Elektromiografi.....	34
2.2.2	Karakteristik Sinyal <i>Surface Electromyography</i> (sEMG) Wajah.....	36
2.3	<i>Big Data Analytics</i>	37
2.3.1	Konsep dan Karakteristik <i>Big Data</i>	37
2.3.2	Sistem Komputasi Terdistribusi untuk Pemrosesan <i>Big Data</i>	38
2.3.3	Teknologi <i>Big Data</i> untuk <i>Big Data Analytics</i>	40
2.4	Analisis <i>Big Data</i> Menggunakan Apache Spark	41
2.4.1	Arsitektur dan Komponen Utama Apache Spark	42
2.4.2	Perbandingan Apache Spark dan Hadoop <i>MapReduce</i>	44
2.5	Hadoop <i>Distributed File System</i> (HDFS)	46
2.5.1	Konsep dan Arsitektur Dasar HDFS	46
2.5.2	Mekanisme Kerja HDFS	47
2.6	Integrasi Apache Spark dengan HDFS pada Arsitektur <i>Multi-Node</i>	49
2.7	Algoritma <i>Clustering</i> untuk Identifikasi Pola Emosi.....	50
2.7.1	K-Means.....	51
2.7.2	Bisecting K-Means.....	56
2.8	Penelitian Relevan.....	59
BAB III METODE PENELITIAN.....		66
3.1	Jenis Penelitian.....	66
3.2	Alur Penelitian	67
3.3	Arsitektur Sistem.....	69
3.3.1	Alur Kerja Sistem.....	72
3.3.2	Algoritma Sistem	76

3.4	Alat dan Bahan.....	82
BAB IV HASIL DAN PEMBAHASAN.....		85
4.1	Deskripsi Data.....	85
4.2	Eksplorasi Data Awal (Exploratory Data Analysis/EDA).....	85
4.3	Hasil <i>Preprocessing</i>	86
4.3.1	<i>Feature Selection</i>	86
4.3.2	<i>Row Filtering</i>	87
4.3.3	<i>Outlier Detection</i>	88
4.4	<i>Clustering</i>	92
4.4.1	Simulasi <i>Clustering</i> dengan Data Sintetis Sederhana Menggunakan Perhitungan Manual.....	92
4.4.2	Implementasi <i>Clustering</i> Data Sintetis Sederhana di Spark	101
4.4.3	<i>Clustering</i> Data Sinyal sEMG.....	104
4.5	Analisis Hasil dan Pola <i>Cluster</i>	108
4.5.1	K-Means dengan 2 <i>Cluster</i>	108
4.5.2	K-Means dengan 4 <i>Cluster</i>	113
4.5.3	Bisecting K-Means dengan 2 <i>Cluster</i>	119
4.5.4	Bisecting K-Means dengan 4 <i>Cluster</i>	123
4.6	Pemilihan Konfigurasi <i>Cluster</i> Optimal dan Interpretasi Pola Emosi	127
4.6.1	Evaluasi Performa <i>Clustering</i>	128
4.6.2	Interpretasi Pola Emosi Berdasarkan Teori Emosi dan <i>Arousal</i>	132
4.7	Validasi Hasil dan Perbandingan dengan Studi Pendahuluan	137
4.7.1	Perbandingan Metodologi dan Cakupan Data.....	137
4.7.2	Evaluasi Kualitas Hasil <i>Clustering</i>	138

BAB V KESIMPULAN DAN SARAN.....	142
5.1 Kesimpulan	142
5.1.1 Keberhasilan Pengembangan Teknik Analisis <i>Big Data</i> untuk Pemrosesan Sinyal sEMG Wajah.....	142
5.1.2 Efektivitas <i>Big Data</i> dalam Meningkatkan Kualitas Identifikasi Pola Emosi	143
5.1.3 Keunggulan Komputasi Terdistribusi dalam Pemrosesan <i>Big Data</i> sEMG	143
5.1.4 Optimalisasi Konfigurasi <i>Clustering</i>	144
5.1.5 Identifikasi Pola Emosi Berdasarkan Tingkat <i>Arousal</i>	144
5.1.6 Kontribusi Kuantifikasi pada Aktivasi Nilai Sinyal sEMG Amplitudo	144
5.2 Saran.....	145
5.2.1 Optimalisasi Parameter <i>Initial Centroid</i>	145
5.2.2 Implementasi Algoritma <i>Clustering</i> Alternatif	146
5.2.3 Integrasi Sinyal Fisiologis Multimodal	146
5.2.4 Personalisasi Model Adaptif	147
5.2.5 Prospek Pengembangan Aplikasi	147
DAFTAR PUSTAKA	148
LAMPIRAN	154
RIWAYAT HIDUP	155

DAFTAR TABEL

Tabel 3. 1 Informasi Hardware yang Digunakan dalam Penelitian untuk Memproses Sinyal sEMG pada Sistem Komputasi Terdistribusi	82
Tabel 3. 2 Informasi Software yang Digunakan dalam Penelitian untuk Memproses Sinyal sEMG pada Sistem Komputasi Terdistribusi	82
Tabel 3. 3 Informasi Dataset Mentah Sinyal sEMG dari EmteqLabs yang Digunakan dalam Penelitian.....	83
Tabel 4. 1 Struktur Dataset Awal Sinyal sEMG dari EmteqLabs.....	85
Tabel 4. 2 Hasil Seleksi Fitur pada Dataset Awal Sinyal sEMG.....	87
Tabel 4. 3 Status Data Berdasarkan Kategori Event pada Dataset Awal Sinyal sEMG	87
Tabel 4. 4 Status Data Berdasarkan Nilai NULL dari Kolom Sinyal sEMG Amplitudo pada Dataset Awal Sinyal sEMG	88
Tabel 4. 5 Hasil Preprocessing Dataset Sinyal sEMG per Partisipan	88
Tabel 4. 6 Data Sintetis Sederhana untuk Simulasi Manual Algoritma Clustering	92
Tabel 4. 7 Perhitungan Jarak Objek ke Centroid (Iterasi ke-1 K-Means).....	94
Tabel 4. 8 Perhitungan Centroid Baru (Iterasi ke-1 K-Means)	94
Tabel 4. 9 Perhitungan Jarak Objek ke Centroid (Iterasi ke-2 K-Means).....	95
Tabel 4. 10 Perhitungan Jarak Objek ke Centroid (Iterasi ke-1 Bisecting K-Means)	96
Tabel 4. 11 Perhitungan Centroid Baru (Iterasi ke-1 Bisecting K-Means)	97
Tabel 4. 12 Perhitungan SSE (Iterasi ke-1 Bisecting K-Means).....	97
Tabel 4. 13 Perhitungan Jarak Objek ke Centroid (Iterasi ke-2 Bisecting K-Means)	98
Tabel 4. 14 Perhitungan Centroid Baru (Iterasi ke-2 Bisecting K-Means)	99
Tabel 4. 15 Perhitungan SSE (Iterasi ke-2 Bisecting K-Means).....	99
Tabel 4. 16 Perhitungan Jarak Objek ke Centroid (Iterasi ke-3 Bisecting K-Means)	100
Tabel 4. 17 Perhitungan Centroid Baru (Iterasi ke-3 Bisecting K-Means).....	100
Tabel 4. 18 Perhitungan SSE (Iterasi ke-3 Bisecting K-Means).....	100

Tabel 4. 19 Perbandingan Silhouette Coefficient pada Konfigurasi Jumlah Cluster (2–6) Menggunakan Algoritma K-Means dan Bisecting K-Means	105
Tabel 4. 20 Perbandingan Waktu Eksekusi, Alokasi Sumber Daya, dan Kualitas Clustering Berdasarkan Jumlah Worker Node	106
Tabel 4. 21 Perbandingan Waktu Eksekusi, Pemanfaatan Resource, dan Distribusi Komputasi Clustering pada Berbagai Jumlah Worker Node	107
Tabel 4. 22 Nilai Centroid Sinyal sEMG Amplitudo Hasil K-Means 2 Cluster...	109
Tabel 4. 23 Statistik Deskriptif Sinyal sEMG Amplitudo K-Means 2 Cluster ...	109
Tabel 4. 24 Distribusi Label Stimulus Hasil K-Means 2 Cluster	112
Tabel 4. 25 Nilai Centroid Sinyal sEMG Amplitudo Hasil K-Means 4 Cluster.....	113
Tabel 4. 26 Statistik Deskriptif Sinyal sEMG Amplitudo K-Means 4 Cluster.....	114
Tabel 4. 27 Distribusi Label Stimulus Hasil K-Means 4 Cluster	118
Tabel 4. 28 Nilai Centroid Sinyal sEMG Amplitudo Hasil Bisecting K-Means 2 Cluster	119
Tabel 4. 29 Statistik Deskriptif Sinyal sEMG Amplitudo Bisecting K-Means 2 Cluster	120
Tabel 4. 30 Distribusi Label Stimulus Hasil Bisecting K-Means 2 Cluster.....	122
Tabel 4. 31 Nilai Centroid Sinyal sEMG Amplitudo Hasil Bisecting K-Means 4 Cluster	123
Tabel 4. 32 Statistik Deskriptif Sinyal sEMG Amplitudo Bisecting K-Means 4 Cluster	124
Tabel 4. 33 Distribusi Label Stimulus Hasil Bisecting K-Means 4 Cluster.....	127
Tabel 4. 34 Perbandingan Performa Clustering Berdasarkan Algoritma dan Konfigurasi Jumlah Cluster.....	128
Tabel 4. 35 Keterkaitan Aktivasi Otot Wajah, AU, dan Jenis Emosi yang Diindikasikan (Cluster 1 = High Arousal)	134
Tabel 4. 36 Perbandingan Karakteristik Emosi Fisiologis antar Cluster	137
Tabel 4. 37 Perbandingan Metodologi dan Infrastruktur antara Studi Pendahuluan dan Penelitian Saat Ini.....	138
Tabel 4. 38 Perbandingan Kualitas Clustering Berdasarkan Silhouette Coefficient antara Studi Pendahuluan dan Penelitian Saat Ini.....	138

Tabel 4. 39 Perbandingan Hasil Interpretasi Pola Emosi Berdasarkan Hasil Clustering antara Studi Pendahuluan dan Penelitian Saat Ini	139
Tabel 4. 40 Perbandingan Nilai Centroid Cluster antara Studi Pendahuluan dan Penelitian Saat Ini ($\times 10^{-6}$).....	140

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian Analisis Big Data Sinyal sEMG	67
Gambar 3. 2 Arsitektur Infrastruktur Cluster Spark.....	69
Gambar 3. 3 Arsitektur Sistem Pemrosesan Terdistribusi Sinyal sEMG Menggunakan Apache Spark.....	70
Gambar 3. 4 Flowchart Sistem Pemrosesan dan Analisis Sinyal sEMG Menggunakan Apache Spark.....	73
Gambar 3. 5 Proses Distributed Computing untuk Pemrosesan Sinyal sEMG Menggunakan Apache Spark.....	75
Gambar 4. 1 Visualisasi Retensi Dataset Sinyal sEMG Setelah Preprocessing	91
Gambar 4. 2 Visualisasi Titik Data dan Centroid Awal pada Data Sintetis Sederhana (K=3).....	93
Gambar 4. 3 Tampilan File Input Data Sintetis Sederhana di Direktori HDFS ..	101
Gambar 4. 4 Tampilan Worker Node Aktif (Job Clustering K-Means) di Web UI YARN.....	102
Gambar 4. 5 Output Proses Clustering K-Means pada Data Sintetis Sederhana yang Telah Selesai Dieksekusi di Cluster Spark Terdistribusi.....	102
Gambar 4. 6 Tampilan File Output Data Sintetis Sederhana di Direktori HDFS (K-Means).....	103
Gambar 4. 7 Output Proses Clustering Bisecting K-Means pada Data Sintetis Sederhana yang Telah Selesai Dieksekusi di Cluster Spark Terdistribusi	103
Gambar 4. 8 Tampilan File Output Data Sintetis Sederhana di Direktori HDFS (Bisecting K-Means).....	104
Gambar 4. 9 Boxplot Sinyal sEMG Amplitudo K-Means 2 Cluster.....	111
Gambar 4. 10 Boxplot Sinyal sEMG Amplitudo K-Means 4 Cluster.....	118
Gambar 4. 11 Boxplot Sinyal sEMG Amplitudo Bisecting K-Means 2 Cluster..	122
Gambar 4. 12 Boxplot Sinyal sEMG Amplitudo Bisecting K-Means 4 Cluster..	126
Gambar 4. 13 Aktivasi Otot Wajah Hasil Clustering K-Means 2 Cluster.....	131

Gambar 4. 14 Aktivasi Otot Wajah Hasil Clustering Bisecting K-Means 2 Cluster	131
Gambar 4. 15 Representasi Emosi yang Terbentuk dari Hasil Clustering (Kiri = Cluster 0, Kanan = Cluster 1).....	135
Gambar 4. 16 Pemetaan Pola Emosi dari Hasil Clustering Berdasarkan Model Sirkumfleks Rusell	136

DAFTAR PUSTAKA

- [1] E. Tsopanova, “The Role of Emotions in Decision-Making Systems,” 2023.
- [2] F. H. Hariz, K. N. Upadhyaya, T. F. Sumayya, B. Bharatesh, and S. M. Gowda, “Facial emotion recognition for smart applications,” *Int. J. Eng. Res. Technol. NCCDS*, vol. 8, no. 13, 2020.
- [3] M. Ben Ayed, S. Elkosantini, S. A. Alshaya, and M. Abid, “Suspicious behavior recognition based on face features,” *IEEE Access*, vol. 7, pp. 149952–149958, 2019.
- [4] V. K. Barigala *et al.*, “Identifying the Optimal Location of Facial EMG for Emotion Detection Using Logistic Regression,” in *Healthcare Transformation with Informatics and Artificial Intelligence*, IOS Press, 2023, pp. 81–84.
- [5] L. Svecova, D. Vala, and Z. Slanina, “EMG as objective method for revealed mistakes in sport shooting,” in *Proceedings of the Second International Scientific Conference “Intelligent Information Technologies for Industry”(IITI’17) Volume 2*, Springer, 2018, pp. 439–448.
- [6] C. Farkas, A. Hamilton-Wright, H. Parsaei, and D. W. Stashuk, “A review of clinical quantitative electromyography,” *Crit. Rev. Biomed. Eng.*, vol. 38, no. 5, 2010.
- [7] C. R. and V. P. G L, “Deep Learning Based Emotion Detection System,” *Interantional J. Sci. Res. Eng. Manag.*, vol. 08, no. 07, 2024, doi: 10.55041/ijsrem36445.
- [8] P. A. Sanipatín-Díaz, P. D. Rosero-Montalvo, and W. Hernandez, “Portable Facial Expression System Based on EMG Sensors and Machine Learning Models,” *Sensors*, vol. 24, no. 11, p. 3350, 2024.
- [9] S. A. Mithbavkar and M. S. Shah, “Recognition of emotion through facial expressions using EMG signal,” in *2019 international conference on nascent technologies in engineering (ICNTE)*, IEEE, 2019, pp. 1–6.
- [10] M. S. Rashid, Z. Zaman, H. Mahmud, and M. K. Hasan, “Emotion Recognition with Forearm-based Electromyography,” *arXiv Prepr. arXiv1911.05305*, 2019.
- [11] J. J. A. van Berkum, M. Struiksma, and B. t Hart, “Using facial EMG to track emotion during language comprehension: past, present, and future,” in *Language Electrified*, Springer, 2023, pp. 687–729.
- [12] J. M. Rutkowska *et al.*, “Optimal processing of surface facial EMG to identify emotional expressions: A data-driven approach,” *Behav. Res. Methods*, pp. 1–14, 2024.

- [13] L. Venica, A. Mumtaz, M. R. Wahid, D. I. H. Putri, and M. S. Gianti, "Identifying Emotional Signatures from Facial EMG Signals: A Clustering-Based Exploration (Preliminary Study)," in *2024 IEEE International Conference on Smart Mechatronics (ICSMech)*, IEEE, 2024, pp. 305–310.
- [14] R. A. Nugroho, I. Cholissodin, and I. Indriati, "Implementasi Naïve Bayes Classifier untuk Klasifikasi Emosi Tweet Berbahasa Indonesia pada Spark," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 1, pp. 301–310, 2021.
- [15] A. Mohebi, S. Aghabozorgi, T. Ying Wah, T. Herawan, and R. Yahyapour, "Iterative big data clustering algorithms: a review," *Softw. Pract. Exp.*, vol. 46, no. 1, pp. 107–129, 2016.
- [16] M. Kang and J.-G. Lee, "An experimental analysis of limitations of MapReduce for iterative algorithms on Spark," *Cluster Comput.*, vol. 20, pp. 3593–3604, 2017.
- [17] J. G. Shanahan and L. Dai, "Large scale distributed data science using apache spark," in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2015, pp. 2323–2324.
- [18] S. Salloum, R. Dautov, X. Chen, P. X. Peng, and J. Z. Huang, "Big data analytics on Apache Spark," *Int. J. Data Sci. Anal.*, vol. 1, pp. 145–164, 2016.
- [19] C. Na and C. Xin, "The research of large scale data processing platform based on the spark," in *2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, IEEE, 2016, pp. 293–296.
- [20] S. Tang, B. He, C. Yu, Y. Li, and K. Li, "A Survey on Spark Ecosystem for Big Data Processing," pp. 1–21, 2018, doi: 10.1109/TKDE.2020.2975652.
- [21] H. K. Omar and A. K. Jumaa, "Distributed big data analysis using spark parallel data processing," *Bull. Electr. Eng. Informatics*, vol. 11, no. 3, pp. 1505–1515, 2022.
- [22] M. M. Saeed, Z. Al Aghbari, and M. Alsharidah, "Big data clustering techniques based on spark: a literature review," *PeerJ Comput. Sci.*, vol. 6, p. e321, 2020.
- [23] M. Gjoreski *et al.*, "Facial EMG sensing for monitoring affect using a wearable device," *Sci. Rep.*, vol. 12, no. 1, p. 16876, 2022.
- [24] B. Sarasati and O. Nurvia, "Emosi Dalam Tulisan," *Psibernetika*, vol. 14, no. 1, 2021.
- [25] S. Rahmadanti, "Manajemen Emosi Kungsi Keseimbangan Psikologis," *Circ. Arch.*, vol. 1, no. 4, 2024.
- [26] D. Bruni, "Evolutionary Psychology and Emotions: A Species-Typical Computational Design," *Theor. Hist. Sci.*, vol. 16, p. 29, 2019.

- [27] G. Coppin and D. Sander, "Theoretical approaches to emotion and its measurement," in *Emotion measurement*, Elsevier, 2021, pp. 3–37.
- [28] A. Dessai and H. Virani, "Multimodal and Multidomain Feature Fusion for Emotion Classification Based on Electrocardiogram and Galvanic Skin Response Signals," *Sci*, vol. 6, no. 1, p. 10, 2024.
- [29] M. Gruchola, "Conceptualisation of Emotions," *Rocz. Kult.*, vol. 14, no. 1, pp. 15–32, 2023.
- [30] R. E. Jack, "Cultural specificities in the transmission and decoding of facial expressions of emotion," *Expr. Emot. Philos. Psychol. Leg. Perspect.*, pp. 173–198, 2016.
- [31] H. C. J. Cuve, S. Sowden-Carvalho, and J. Cook, "Spatiotemporal structure and substates in emotional facial expressions," *bioRxiv*, pp. 2009–2024, 2024.
- [32] S. N. Mohammed and A. K. A. Hassan, "A survey on emotion recognition for human robot interaction," *J. Comput. Inf. Technol.*, vol. 28, no. 2, pp. 125–146, 2020.
- [33] X. Liang, A. Angelopoulou, E. Kapetanios, B. Woll, R. Al Batat, and T. Woolfe, "A multi-modal machine learning approach and toolkit to automate recognition of early stages of dementia among british sign language users," in *Computer Vision—ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, Springer, 2020, pp. 278–293.
- [34] C. Correia-Caeiro, A. Burrows, D. A. Wilson, A. Abdelrahman, and T. Miyabe-Nishiwaki, "CalliFACS: the common marmoset facial action coding system," *PLoS One*, vol. 17, no. 5, p. e0266442, 2022.
- [35] K. Savoji, M. Soleimani, and M. AJ, "A Comprehensive Review of Electromyography in Rehabilitation: Detecting Interrupted Wrist and Hand Movements with a Robotic Arm Approach," *EAI Endorsed Trans. AI Robot.*, vol. 3, 2024.
- [36] E. Guzmán-Quezada *et al.*, "Development of an Electromyography Signal Acquisition Prototype and Statistical Validation Against a Commercial Device," *Sensors*, vol. 24, no. 21, p. 6787, 2024.
- [37] J. S. Baschnagel, M. Aita, and M. McTighe, "22 Surface Electromyography".
- [38] J. A. Jarratt, *Essential Electromyography*. Cambridge University Press, 2023.
- [39] T. Rout, M. Garanayak, M. R. Senapati, and S. K. Kamilla, "Big Data And Its Applications: A Review," *Int. Conf. Electr. Electron. Signals, Commun. Optim.*, 2015.
- [40] N. Irzavika *et al.*, "Big Data." PT Penamuda Media, 2024.

- [41] F. Laganà *et al.*, “Development of an Integrated System of sEMG Signal Acquisition, Processing, and Analysis with AI Techniques,” *Signals*, vol. 5, no. 3, pp. 476–493, 2024.
- [42] V. Janev, “Chapter 1 Ecosystem of Big Data,” in *Knowledge graphs and big data processing*, Springer, 2020, pp. 3–19.
- [43] F. D. Utami and F. D. Astuti, “Comparison of Hadoop Mapreduce and Apache Spark in Big Data Processing with Hgrid247-DE,” *J. Appl. Informatics Comput.*, vol. 8, no. 2, pp. 390–399, 2024.
- [44] S. Oliviandi, A. B. Osmond, and R. Latuconsina, “Implementasi Apache Spark Pada Big Data Berbasis Hadoop Distributed File System,” *eProceedings Eng.*, vol. 5, no. 1, 2018.
- [45] P. Matai and A. Bhatia, “Architecting for Real - Time Analytics: Leveraging Stream Processing and Data Warehousing Integration,” *Int. J. Sci. Res.*, vol. 13, no. 9, pp. 1586–1590, 2024, doi: 10.21275/sr24925170923.
- [46] Q. Q. Tran, B. D. Nguyen, L. T. T. Nguyen, and O. T. T. Nguyen, “Big data processing with apache spark,” *Tra Vinh Univ. J. Sci.*, 2023.
- [47] B. Quinto, *Next-generation machine learning with spark: Covers XGBoost, LightGBM, Spark NLP, distributed deep learning with keras, and more.* Apress, 2020.
- [48] A. Diwakar Reddy and J. Geetha Reddy, “Algorithms for Iterative Applications in MapReduce Framework,” in *International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications: ASISA 2016*, Springer, 2018, pp. 51–61.
- [49] S. Ketu and S. Agarwal, “Performance enhancement of distributed K-Means clustering for big Data analytics through in-memory computation,” in *2015 Eighth International Conference on Contemporary Computing (IC3)*, IEEE, 2015, pp. 318–324.
- [50] V. Doma and M. Pirouz, “A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals,” *J. Big Data*, vol. 7, no. 1, p. 18, 2020.
- [51] M. Elkawagy and H. Elbeh, “High performance hadoop distributed file system,” *Int. J. Networked Distrib. Comput.*, vol. 8, no. 3, pp. 119–123, 2020, doi: 10.2991/ijndc.k.200515.007.
- [52] DAPLAB&UNIFR, “Workshop Data Sciences,” 2017. <https://derlin.github.io/workshop-data-sciences/hdfs/> (accessed May 07, 2025).
- [53] L. Tian, Y. Sun, and L. Yang, “Overview of Storage Architecture and Strategy of HDFS,” in *Applied Mathematics, Modeling and Computer Simulation*, IOS Press, 2022, pp. 945–954.

- [54] Z. Sun, F. Chen, M. Chi, and Y. Zhu, “A spark-based big data platform for massive remote sensing data processing,” in *Data Science: Second International Conference, ICDS 2015, Sydney, Australia, August 8-9, 2015, Proceedings 2*, Springer, 2015, pp. 120–126.
- [55] L. Venica, N. Nurhikam, M. Ayyas, and M. R. Wahid, “Menjelajahi Dunia Machine Learning.” PT Penamuda Media, 2024.
- [56] W. Xiao and J. Hu, “A survey of parallel clustering algorithms based on spark,” *Sci. Program.*, vol. 2020, no. 1, p. 8884926, 2020.
- [57] Y. Zhuang, Y. Mao, and X. Chen, “A limited-iteration bisecting k-means for fast clustering large datasets,” in *2016 IEEE Trustcom/BigDataSE/ISPA*, IEEE, 2016, pp. 2257–2262.
- [58] B. Hartono, S. Eniyati, and K. Hadiono, “Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering,” *J. Sist. Komput. dan Inform. Hal*, vol. 503, p. 509, 2023.
- [59] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Y. Wah, “A comparison study on similarity and dissimilarity measures in clustering continuous data,” *PLoS One*, vol. 10, no. 12, p. e0144059, 2015.
- [60] V. Asanza, E. Peláez, F. Loayza, I. Mesa, J. Díaz, and E. Valarezo, “EMG signal processing with clustering algorithms for motor gesture tasks,” in *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, IEEE, 2018, pp. 1–6.
- [61] T. Johnson and S. K. Singh, “Divisive Hierarchical Bisecting Min–Max Clustering Algorithm,” in *Proceedings of the International Conference on Data Engineering and Communication Technology: ICDECT 2016, Volume I*, Springer, 2017, pp. 579–592.
- [62] V. Rohilla, S. Chakraborty, and M. S. Singh, “Data clustering using bisecting k-means,” in *2019 international conference on computing, communication, and intelligent systems (ICCCIS)*, IEEE, 2019, pp. 80–83.
- [63] R. Patil and A. Khan, “Bisecting K-means for clustering web log data,” *Int. J. Comput. Appl.*, vol. 116, no. 19, 2015.
- [64] E. Burghardt, D. Sewell, and J. Cavanaugh, “Agglomerative and divisive hierarchical Bayesian clustering,” *Comput. Stat. Data Anal.*, vol. 176, p. 107566, 2022.
- [65] M. S. E. Sendi, M. Heydarzadeh, and B. Mahmoudi, “A spark-based analytic pipeline for seizure detection in EEG big data streams,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2018, pp. 4003–4006.
- [66] 李玉杰, 张晶晶, 蒋伟, and 王春晓, “Research on emotional space for movie

- and TV drama videos,” *中国邮电高校学报 (英文)*, vol. 29, no. 5, p. 73, 2022.
- [67] M. Kołodziej, A. Majkowski, and M. Jurczak, “Acquisition and Analysis of Facial Electromyographic Signals for Emotion Recognition,” *Sensors*, vol. 24, no. 15, p. 4785, 2024.
 - [68] S.-T. Lai, “An Iterative and Incremental Data Quality Improvement Procedure for Reducing the Risk of Big Data Project,” *J. Softw.*, vol. 12, no. 12, pp. 945–956, 2017.
 - [69] A. G. Bin Saadon and H. M. O. Mokhtar, “Survey on iterative and incremental approaches in distributed computing environment,” *Int. J. Data Sci.*, vol. 4, no. 1, pp. 18–30, 2019.
 - [70] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, “Incremental deep computation model for wireless big data feature learning,” *IEEE Trans. Big Data*, vol. 6, no. 2, pp. 248–257, 2019.
 - [71] B. St. Thomas, P. Chandar, C. Hosey, and F. Diaz, “Mixed method development of evaluation metrics,” in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 4070–4071.
 - [72] X. Zhang, T. Li, M. Sun, L. Zhang, C. Zhang, and Y. Zhang, “Replay-Based Incremental Learning Framework for Gesture Recognition Overcoming the Time-Varying Characteristics of sEMG Signals,” *Sensors*, vol. 24, no. 22, p. 7198, 2024.
 - [73] C. Pan *et al.*, “Phased array antenna calibration method experimental validation and comparison,” *Electronics*, vol. 12, no. 3, p. 489, 2023.
 - [74] S. Smith and E. Mörelius, “Principle-based concept analysis methodology using a phased approach with quality criteria,” *Int. J. Qual. Methods*, vol. 20, p. 16094069211057996, 2021.