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

Universitas Pendidikan Indonesia

Juli 2025

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 (±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 (±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

DAFTAR TABEL

DAFTAR GAMBAR

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 Kunsi 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. Gruchoła, "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. Elkawkagy 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 1*, 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.