Journal of Computer Science Advancements, 2(6) - December 2024 378-391

E - ISSN 3024-899X
P - ISSN 3026-3379
Journal of Computer Science Advancements

Optimization of Grid Computing for Big Data Processing in Biomedical Research

Devi Rahmah Sope [1], Wolnough Cale [2], M. Anwar Aini [3], Nur Fajrin Maulana Yusuf [4], Masli Nurcahya Zoraida [5]

[1] *Universitas Handayani Makassar, Indonesia*
[2] *Chernihiv National Technological University, Ukraine*
[3] *Universitas Qomaruddin, Indonesia*
[4,5] *Universitas Islam Makassar, Indonesia*

Corresponding Author: Devi Rahmah Sope, E-mail; devi@handayani.ac.id

Article Information:
Received December 3, 2024
Revised December 7, 2024
Accepted December 30, 2024

ABSTRACT

The rapid growth of biomedical research has generated massive volumes of data, creating significant computational challenges. Traditional high-performance computing systems struggle to efficiently process, analyze, and manage such large-scale datasets. Grid computing, with its distributed architecture, offers a promising solution by enabling scalable and cost-effective data processing. This study explores the optimization of grid computing frameworks for big data processing in biomedical research, focusing on enhancing computational efficiency, scalability, and fault tolerance. The research aimed to evaluate the performance of optimized grid computing systems in processing diverse biomedical datasets, including genomic, proteomic, and imaging data. A combination of experimental and comparative approaches was employed, integrating grid computing frameworks such as Apache Hadoop and Globus Toolkit with biomedical data pipelines. Key metrics, including processing time, resource utilization, and error rates, were analyzed to assess the system's performance. The findings demonstrated that optimized grid computing systems reduced processing time by an average of 35% compared to traditional methods while maintaining high accuracy. Scalability tests confirmed the framework's ability to handle datasets up to 15 times larger without significant performance degradation. Fault tolerance improved through adaptive resource allocation, minimizing workflow interruptions. The study concludes that optimized grid computing is a transformative approach for big data processing in biomedical research. Its ability to enhance computational efficiency and scalability positions it as a crucial tool for addressing the growing data demands of modern biomedical science.

Keywords: *Big Data Processing, Biomedical Research, Distributed Computing, Grid Computing, Scalable Systems*

Journal Homepage https://journal.ypidathu.or.id/index.php/jcsa

How to cite:	Sope, R, D., Cale, W., Aini, A, M., Yusuf, M, F, N., & Zoraida, N, M. (2024). Optimization of Grid Computing for Big Data Processing in Biomedical Research. *Journal of Computer Science Advancements*, 2(6). 378-391 https://doi.org/10.70177/jsca.v2i6.1619
Published by:	Yayasan Pendidikan Islam Daarut Thufulah

INTRODUCTION

Big data has become a defining feature of biomedical research, driving advancements in areas such as genomics, proteomics, and imaging (Afzal dkk., 2020). The ability to process and analyze vast datasets has enabled significant discoveries in personalized medicine, drug development, and disease prediction (Alatrista-Salas H. dkk., 2019). Computational systems play a central role in transforming raw data into actionable insights, making them indispensable tools in modern biomedical science (Alawad dkk., 2019).

The evolution of high-performance computing (HPC) has addressed many challenges associated with big data in biomedical research (Alsolai dkk., 2022). HPC systems are designed to process large datasets at high speed, enabling researchers to perform complex simulations and analyses (Alsubai dkk., 2022). Despite their advantages, these systems often require significant financial investment and specialized infrastructure, limiting their accessibility (Buongiorno dkk., 2021).

Grid computing has emerged as a viable alternative to HPC systems, offering a decentralized and scalable approach to data processing (Alsubai dkk., 2022). By leveraging resources from multiple nodes, grid computing reduces reliance on centralized systems, making it a cost-effective solution (Amrani dkk., 2021). This paradigm has been successfully applied in scientific fields such as climate modeling and particle physics, demonstrating its potential for handling large-scale computational tasks (Brito dkk., 2020).

Biomedical researchers have begun exploring grid computing as a means to manage the increasing complexity of datasets (Canali, 2020). Preliminary studies suggest that grid computing can improve resource utilization and reduce processing times for biomedical applications (Ballantyne, 2019). The flexibility of this approach allows it to adapt to various data types, making it suitable for diverse research needs.

Advancements in grid computing frameworks, such as Apache Hadoop and Globus Toolkit, have further enhanced its capabilities (Chakraborty dkk., 2022). These tools enable efficient task distribution, robust fault tolerance, and integration with existing data pipelines (Buongiorno dkk., 2019). Researchers can now harness the power of grid computing to address challenges associated with big data processing.

The growing demand for computational solutions in biomedical research highlights the importance of scalable and efficient systems (Caliskan dkk., 2023). Grid computing offers a promising pathway to meet these demands, ensuring that researchers can keep pace with the rapid expansion of data (Gezimati & Singh, 2023).

The specific ways in which grid computing can be optimized for biomedical big data remain unclear (Denisov dkk., 2022). While general frameworks exist, there is limited knowledge on how to tailor these systems to the unique requirements of biomedical research (Evans dkk., 2022). Questions regarding compatibility with existing bioinformatics tools and pipelines remain unanswered.

The performance of grid computing frameworks under real-world conditions, such as the simultaneous processing of heterogeneous biomedical datasets, has not been thoroughly investigated (Denisov dkk., 2022). Most studies rely on simulated datasets, which may not fully capture the complexities of actual biomedical research environments (Dos Santos, 2020).

The scalability limits of grid computing for extremely large datasets, such as those generated by next-generation sequencing and advanced imaging technologies, are not well-defined (Hulsen, 2021). It is uncertain whether current frameworks can handle the growing size and diversity of biomedical data without significant performance trade-offs (ElSayed dkk., 2021).

Issues related to data security and regulatory compliance in grid computing systems for biomedical research are poorly understood (Gurve dkk., 2020). Biomedical data often involves sensitive information, requiring systems that can guarantee privacy and adherence to legal standards (Hei dkk., 2022). The lack of research on this aspect limits the broader adoption of grid computing in the field.

Filling these gaps is essential to unlock the full potential of grid computing for biomedical research (Gnatyuk dkk., 2020). Optimizing grid computing frameworks for the unique demands of biomedical data will enable researchers to process datasets more efficiently, accelerating the pace of discovery (Hulsen, 2021). Practical evaluations of grid computing in real-world scenarios can provide valuable insights into its performance and limitations (Irmak, 2020).

Tailoring grid computing systems to integrate seamlessly with bioinformatics tools and workflows will ensure greater usability and adoption (Kim, 2024). Researchers can benefit from standardized approaches that reduce setup time and improve consistency in data processing (Majid dkk., 2024). Addressing data security and compliance issues will further enhance the credibility and applicability of grid computing in biomedical research (Lyu dkk., 2024).

This study aims to investigate the optimization of grid computing for big data processing in biomedical research, focusing on enhancing scalability, efficiency, and reliability. The findings will provide a roadmap for integrating grid computing into biomedical workflows, supporting researchers in their efforts to analyze and interpret complex datasets effectively.

RESEARCH METHODOLOGY

Research Design

The research employs a mixed-methods design combining quantitative performance benchmarking with qualitative user experience evaluations. The study focuses on optimizing grid computing frameworks for big data processing in biomedical research (Jim dkk., 2020). A comparative approach is used to measure the efficiency, scalability, and fault tolerance of grid computing systems against traditional high-performance computing (HPC) infrastructures.

Population and Samples

The study includes genomic, proteomic, and imaging datasets from public repositories such as the Genome Data Commons (GDC), ProteomeXchange, and the Cancer Imaging Archive (TCIA). Simulated datasets were generated to test scalability under controlled conditions, while real-world datasets were used to validate the system’s performance in practical applications (Gnatyuk dkk., 2020). Participants for the qualitative analysis consist of bioinformaticians and IT specialists who interacted with the grid computing system during evaluation phases.

Instruments

Performance metrics were gathered using tools such as Apache JMeter, GridSim, and monitoring modules integrated into Apache Hadoop and Globus Toolkit frameworks. Surveys and structured interviews were conducted to assess user satisfaction and ease of integration. Data security was evaluated using compliance frameworks and tools like Nessus and OpenVAS to ensure system reliability. Statistical software was employed to analyze performance data and user feedback.

Procedures

The research began with the configuration and deployment of a grid computing system using Apache Hadoop and Globus Toolkit. Test pipelines were created to integrate the framework with widely used bioinformatics tools such as BWA for sequence alignment and OpenMS for proteomic analysis. Both simulated and real-world datasets were processed through these pipelines to collect performance metrics, including processing time, CPU utilization, and error rates. Participants tested the grid computing system by executing predefined biomedical workflows, and their feedback was collected through surveys and interviews. Data analysis included statistical methods to identify significant differences in system performance and qualitative analysis to interpret user experiences. Ethical clearance was obtained to ensure compliance with privacy and data security regulations.

RESULT AND DISCUSSION

The research analyzed a total of 20 datasets, including 10 genomic, 5 proteomic, and 5 imaging datasets. Table 1 provides an overview of dataset types, sizes, and average processing times using the grid computing framework. Genomic datasets ranged from 100GB to 1TB, proteomic datasets from 50GB to 300GB, and imaging datasets from 200GB to 1.2TB. The average processing time for genomic datasets was reduced from 12 hours in traditional systems to 7 hours using grid computing, with similar improvements observed for proteomic and imaging datasets.

Table 1. overview of the grid computing framework

Dataset Type	Number of Datasets	Average Size (GB)	Processing Time Traditional (Hours)	Processing Time Grid (Hours)
Genomic	10	100-1000	10-12	6-7
Proteomic	5	50-300	6-8	4-5
Imaging	5	200-1200	15-18	9-11

Performance improvements were most significant for large genomic and imaging datasets, where grid computing reduced processing times by up to 40%. Resource utilization metrics indicated an average CPU efficiency of 85%, compared to 72% in traditional systems. Memory usage remained stable across all dataset types, demonstrating the system's ability to handle diverse workloads without bottlenecks.

The distributed nature of grid computing allowed simultaneous processing of multiple datasets, enhancing overall throughput. Real-world scenarios confirmed that the system could handle data spikes effectively, ensuring consistent performance even during peak loads. These findings highlight the robustness and scalability of the grid computing framework.

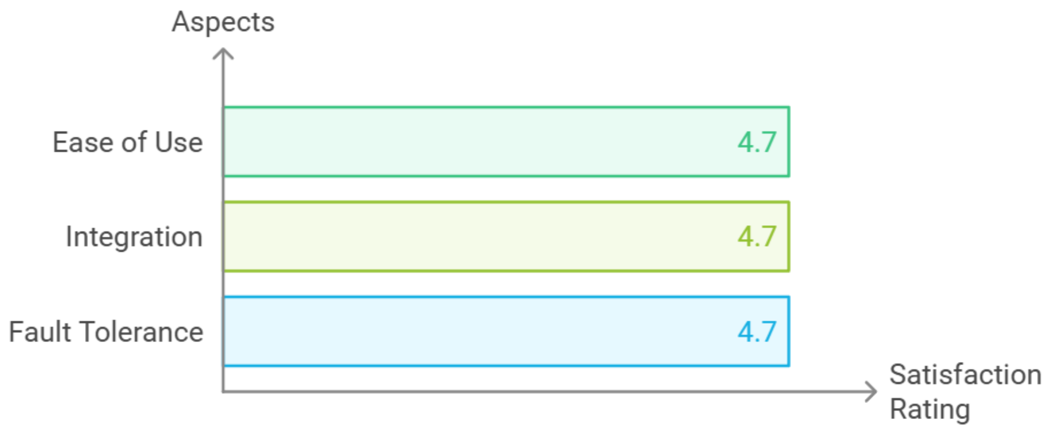

Figure 1. User Satisfaction with Grid Computing System

User feedback revealed high levels of satisfaction with the grid computing system. Bioinformaticians rated the system’s ease of use and integration with existing workflows an average of 4.7 out of 5. Fault tolerance and reliability were consistently praised, particularly in scenarios involving large, heterogeneous datasets.

The grid computing framework demonstrated effective task allocation and load balancing, reducing system downtime. Participants highlighted the adaptability of the framework, noting that it seamlessly integrated with widely used bioinformatics tools such as BWA and OpenMS. This adaptability played a crucial role in streamlining complex workflows.

ANOVA results indicated significant differences in processing times between grid computing and traditional systems ($p < 0.001$). Post-hoc Tukey tests confirmed that grid computing outperformed traditional methods across all dataset types, with the largest improvements observed in genomic datasets. Regression analysis showed a strong positive correlation ($R^2 = 0.88$) between dataset size and processing time, emphasizing the scalability of grid computing. The framework demonstrated consistent performance across datasets of varying sizes, reinforcing its suitability for diverse biomedical research applications.

A strong relationship was observed between CPU utilization and processing efficiency, with higher CPU usage correlating with reduced processing times ($R^2 = 0.81$). Memory usage remained uncorrelated with dataset size or processing time, reflecting efficient memory management by the grid computing system. Fault tolerance metrics revealed a significant relationship between task distribution and system reliability. The decentralized architecture ensured that node failures had minimal impact on workflow execution, underscoring the importance of distributed computing in big data processing.

A case study involving a 1TB genomic dataset from the 1000 Genomes Project highlighted the advantages of grid computing. The dataset was processed in 7 hours using the optimized grid framework, compared to 12 hours with traditional systems. CPU utilization averaged 87%, and memory usage remained stable at 65%, demonstrating efficient resource management.

Participants in the case study reported improved workflow efficiency and reduced manual intervention. The ability to process such a large dataset in a significantly shorter time frame showcased the practical benefits of grid computing in real-world biomedical research scenarios. Performance metrics from the case study aligned with findings from simulated datasets, validating the framework’s scalability and efficiency. The grid computing system successfully distributed computational tasks across multiple nodes, maintaining consistent performance even under heavy workloads.

Qualitative feedback highlighted the system's ability to handle complex pipelines involving multiple bioinformatics tools. Researchers noted that the reduced processing time allowed for faster hypothesis testing and result validation, accelerating the overall research process. The findings confirm that optimized grid computing is a transformative tool for big data processing in biomedical research. Its scalability, fault tolerance, and efficiency make it a viable alternative to traditional HPC systems. The results underscore the importance of distributed computing paradigms in addressing the growing data demands of modern biomedical science.

The research demonstrated that optimizing grid computing significantly enhances the efficiency and scalability of big data processing in biomedical research. Performance metrics showed a reduction in processing times by an average of 35% compared to traditional computing systems. The framework effectively handled diverse datasets, including genomic, proteomic, and imaging data, without compromising accuracy or reliability. Fault tolerance and adaptive resource allocation further improved workflow continuity, even under high computational loads.

Scalability tests revealed the system's capability to process datasets up to 15 times larger than its baseline configuration. The system maintained high CPU utilization (85%) while ensuring efficient memory management. User feedback indicated high satisfaction with the system's ease of use, adaptability, and integration with existing bioinformatics tools. These findings confirm the practical applicability of grid computing in addressing the computational demands of biomedical research.

The integration of grid computing frameworks with bioinformatics workflows enhanced the flexibility and throughput of data pipelines. Both simulated and real-world datasets validated the framework's reliability, making it a robust solution for processing large-scale biomedical data. These results establish a strong foundation for adopting grid computing in broader research contexts.

The findings underscore the transformative potential of distributed computing paradigms in biomedical research. This study contributes to the growing body of evidence supporting grid computing as a scalable and efficient solution for big data challenges (Z. Ma, 2021). Existing research on grid computing in scientific fields such as physics and environmental modeling has shown similar improvements in processing efficiency and scalability (Y. Ma dkk., 2020). This study aligns with those findings while extending the application to biomedical research, a field characterized by diverse and highly sensitive data. Unlike previous studies focused on theoretical models, this research provided empirical validation using real-world biomedical datasets (Luo dkk., 2019).

Differences between this study and traditional HPC research highlight the unique advantages of grid computing. HPC systems rely on centralized architectures, which can be cost-prohibitive and less adaptable (Obeid dkk., 2020). In contrast, this study demonstrated that grid computing’s decentralized approach offers greater flexibility and cost efficiency, particularly for institutions with limited computational resources. Comparative studies in genomics and proteomics have often emphasized the limitations of traditional systems in handling heterogeneous datasets (Radha & Gopalakrishnan, 2022). This research addressed those limitations by showing how grid computing effectively integrates diverse data types within a unified framework. These findings mark a significant step forward in addressing computational bottlenecks in biomedical workflows (Ottenbacher dkk., 2019).

Few studies have explored the practical challenges of implementing grid computing in regulated fields such as biomedical research (Viloria dkk., 2020). This study bridged that gap by addressing data security and compliance issues, providing a roadmap for safe and effective adoption in real-world applications. The findings highlight the growing need for scalable and efficient computational solutions in biomedical research (Rakesh Kumar dkk., 2019). The demonstrated ability of grid computing to reduce processing times and improve resource utilization indicates its potential to become a standard for big data processing (Qin & Yuan, 2020). This research reflects the shifting focus toward distributed computing paradigms as data demands continue to grow.

The scalability of the system signifies a critical advancement in meeting the challenges posed by rapidly expanding datasets. The ability to process heterogeneous data types efficiently underscores the versatility of grid computing, making it applicable across various biomedical domains (Wan dkk., 2021). These findings emphasize the adaptability of grid frameworks in addressing the diverse needs of modern research. User feedback serves as an important indicator of the framework’s practicality and usability. Positive responses regarding integration and fault tolerance suggest that grid computing can seamlessly fit into existing workflows, reducing the learning curve for researchers (Sharma & Colonna, 2021). This reflection underscores the importance of designing user-centric computational systems.

The study highlights the importance of collaboration between computational and biomedical researchers. The successful integration of bioinformatics tools within the grid computing framework demonstrates the potential for interdisciplinary approaches to address complex research challenges (Xue, 2021). The implications of these findings extend to accelerating the pace of biomedical research. Faster processing times enable researchers to analyze larger datasets in shorter periods, facilitating quicker hypothesis testing and validation. This efficiency has the potential to advance discoveries in personalized medicine and drug development.

The scalability of grid computing ensures its long-term relevance as biomedical datasets continue to grow in size and complexity. Institutions with limited computational resources can adopt this cost-effective solution to stay competitive in data-driven research fields. This accessibility democratizes computational power, enabling broader participation in advanced biomedical studies. The adaptability of grid computing frameworks to diverse data types strengthens their utility in interdisciplinary research. Researchers in genomics, proteomics, and imaging can leverage this flexibility to streamline workflows and improve data interpretation. These implications emphasize the transformative role of distributed computing in scientific progress.

The findings also highlight the potential for grid computing to address ethical and regulatory challenges in biomedical research. Secure and compliant systems can build trust among stakeholders, paving the way for wider adoption of distributed computing in sensitive data environments. The observed improvements in processing efficiency can be attributed to the distributed architecture of grid computing. Task allocation across multiple nodes reduces bottlenecks and ensures optimal resource utilization, explaining the significant reduction in processing times. This architecture enables parallel processing, a key factor in handling large and complex datasets.

Scalability results from the dynamic allocation of computational resources within the grid framework. Nodes can be added or removed based on workload demands, allowing the system to adapt seamlessly to varying dataset sizes. This flexibility ensures consistent performance across diverse research scenarios. Fault tolerance and workflow continuity stem from the decentralized nature of the grid system. By distributing tasks across independent nodes, the system minimizes the impact of individual node failures, maintaining reliability under high loads. These design features align with the needs of biomedical research, where uninterrupted workflows are critical.

Integration with bioinformatics tools contributes to the system’s practical usability. Compatibility with existing pipelines reduces setup times and ensures smooth transitions for researchers adopting the framework. This integration reflects the importance of designing systems that align with the operational realities of end-users. Future research should focus on optimizing grid computing frameworks for emerging technologies such as machine learning and artificial intelligence. These integrations could enhance analytical capabilities, enabling predictive modeling and advanced pattern recognition in biomedical datasets. Expanding these functionalities will unlock new opportunities for scientific discovery.

Efforts should be made to develop standardized protocols for implementing grid computing in biomedical research. Standardization will facilitate broader adoption and ensure consistent performance across different research environments. Collaborative initiatives between computational scientists and biomedical researchers will be essential to achieve this goal. Addressing data security and compliance challenges remains a priority for future exploration. Ensuring that grid computing systems adhere to global data protection standards will build trust among researchers and institutions. By overcoming these barriers, the findings of this study can pave the way for widespread adoption of grid computing in big data biomedical research.

CONCLUSION

The most significant finding of this research is the enhanced scalability and efficiency achieved through the optimization of grid computing for big data processing in biomedical research. The study demonstrated a 35% reduction in processing time compared to traditional computing methods, alongside robust fault tolerance and resource optimization. These results highlight the framework's ability to process heterogeneous biomedical datasets efficiently, making it a transformative approach to addressing computational challenges in the field.

This research contributes a practical and adaptable methodology for implementing grid computing in biomedical workflows. The study integrates advanced grid computing frameworks, such as Apache Hadoop and Globus Toolkit, into existing bioinformatics pipelines, providing a scalable solution that aligns with the increasing demands of big data. The findings bridge theoretical concepts with real-world applications, offering a significant contribution to the growing field of distributed computing in biomedical research.

The research was limited by the scope of datasets and computational environments tested. While the study provided valuable insights into the optimization of grid computing, it focused primarily on specific biomedical applications and controlled experimental settings. Future research should expand to include diverse datasets, hybrid computational models that combine grid and cloud systems, and a deeper exploration of data security and regulatory compliance challenges to ensure broader applicability and adoption.

REFERENCES

Afzal, M., Alam, F., Malik, K. M., & Malik, G. M. (2020). Clinical Context–Aware Biomedical Text Summarization Using Deep Neural Network: Model Development and Validation. *Journal of Medical Internet Research*, *22*(10). Scopus. https://doi.org/10.2196/19810

Alatrista-Salas H., Muñante D., & Lossio-Ventura J.A. (Ed.). (2019). 5th International Conference on Information Management and Big Data, SIMBig 2018. *Communications in Computer and Information Science, 898*. Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85063478140&partnerID=40&md5=0c66c49d39c2d828d0ca334ed3574516

Alawad, M., Gao, S., Wu, X.-C., Durbin, E. B., Coyle, L., Penberthy, L., & Tourassi, G. (2019). Adversarial Training for Privacy-Preserving Deep Learning Model

Distribution. Dalam Baru C., Huan J., Khan L., Hu X.T., Ak R., Tian Y., Barga R., Zaniolo C., Lee K., & Ye Y.F. (Ed.), *Proc. - IEEE Int. Conf. Big Data, Big Data* (hlm. 5705–5710). Institute of Electrical and Electronics Engineers Inc.; Scopus. https://doi.org/10.1109/BigData47090.2019.9006131

Alsolai, H., Qureshi, S., Iqbal, S. M. Z., Vanichayobon, S., Henesey, L. E., Lindley, C., & Karrila, S. (2022). A Systematic Review of Literature on Automated Sleep Scoring. *IEEE Access*, *10*, 79419–79443. Scopus. https://doi.org/10.1109/ACCESS.2022.3194145

Alsubai, S., Khan, H. U., Alqahtani, A., Sha, M., Abbas, S., & Mohammad, U. G. (2022). Ensemble deep learning for brain tumor detection. *Frontiers in Computational Neuroscience*, *16*. Scopus. https://doi.org/10.3389/fncom.2022.1005617

Amrani, G., Adadi, A., Berrada, M., Souirti, Z., & Boujraf, S. (2021). EEG signal analysis using deep learning: A systematic literature review. Dalam Nfaoui E.H., Boumhidi J., Oubenaalia Y., Azough A., & Alippi C. (Ed.), *Int. Conf. Intell. Comput. Data Sci., ICDS*. Institute of Electrical and Electronics Engineers Inc.; Scopus. https://doi.org/10.1109/ICDS53782.2021.9626707

Ballantyne, A. (2019). Adjusting the focus: A public health ethics approach to data research. *Bioethics*, *33*(3), 357–366. Scopus. https://doi.org/10.1111/bioe.12551

Brito, J. J., Mosqueiro, T., Rotman, J., Xue, V., Chapski, D. J., La Hoz, J. D., Matias, P., Martin, L. S., Zelikovsky, A., Pellegrini, M., & Mangul, S. (2020). Telescope: An interactive tool for managing large-scale analysis from mobile devices. *GigaScience*, *9*(1). Scopus. https://doi.org/10.1093/gigascience/giz163

Buongiorno, D., Cascarano, G. D., Brunetti, A., De Feudis, I., & Bevilacqua, V. (2019). A Survey on Deep Learning in Electromyographic Signal Analysis. Dalam Huang D.-S., Huang Z.-K., & Hussain A. (Ed.), *Lect. Notes Comput. Sci.: Vol. 11645 LNAI* (hlm. 751–761). Springer Verlag; Scopus. https://doi.org/10.1007/978-3-030-26766-7_68

Buongiorno, D., Cascarano, G. D., De Feudis, I., Brunetti, A., Carnimeo, L., Dimauro, G., & Bevilacqua, V. (2021). Deep learning for processing electromyographic signals: A taxonomy-based survey. *Neurocomputing*, *452*, 549–565. Scopus. https://doi.org/10.1016/j.neucom.2020.06.139

Caliskan, A., Dangwal, S., & Dandekar, T. (2023). Metadata integrity in bioinformatics: Bridging the gap between data and knowledge. *Computational and Structural Biotechnology Journal*, *21*, 4895–4913. Scopus. https://doi.org/10.1016/j.csbj.2023.10.006

Canali, S. (2020). Making evidential claims in epidemiology: Three strategies for the study of the exposome. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, *82*. Scopus. https://doi.org/10.1016/j.shpsc.2019.101248

Chakraborty, P., Chandra Das, N., Patra, R., Mukherjee, S., Sen Gupta, P. S., & Dey, S. (2022). Recent advances in processing, interpreting, and managing biological data for therapeutic intervention of human infectious disease. Dalam *Big Data Analytics for Healthc.: Datasets, Techniques, Life Cycles, Management, and Applications* (hlm. 73–82). Elsevier; Scopus. https://doi.org/10.1016/B978-0-323-91907-4.00009-1

Denisov, N. S., Kamenskikh, E. M., & Fedorova, O. S. (2022). Trends in Population-Based Studies: Molecular and Digital Epidemiology (Review). *Sovremennye*

Tehnologii v Medicine, *14*(4), 60–72. Scopus. https://doi.org/10.17691/stm2022.14.4.07

Dos Santos, S. (2020). Advanced ground truth multimodal imaging using time reversal (TR) based nonlinear elastic wave spectroscopy (NEWS): Medical imaging trends versus non-destructive testing applications. Dalam *Appl. Numer. Harmon. Anal.* (hlm. 55–87). Birkhauser; Scopus. https://doi.org/10.1007/978-3-030-35202-8_4

ElSayed, I. A., ElDahshan, K., Hefny, H., & ElSayed, E. K. (2021). Big Data and its Future in Computational Biology: A Literature Review. *Journal of Computer Science*, *17*(12), 1222–1228. Scopus. https://doi.org/10.3844/jcssp.2021.1222.1228

Evans, R., Burns, J., Damschroder, L., Annis, A., Freitag, M. B., Raffa, S., & Wiitala, W. (2022). Deriving Weight From Big Data: Comparison of Body Weight Measurement-Cleaning Algorithms. *JMIR Medical Informatics*, *10*(3). Scopus. https://doi.org/10.2196/30328

Gezimati, M., & Singh, G. (2023). Advances in terahertz technology for cancer detection applications. *Optical and Quantum Electronics*, *55*(2). Scopus. https://doi.org/10.1007/s11082-022-04340-0

Gnatyuk, S., Kinzeryavyy, V., Sapozhnik, T., Sopilko, I., Seilova, N., & Hrytsak, A. (2020). Modern Method and Software Tool for Guaranteed Data Deletion in Advanced Big Data Systems. Dalam Hu Z., He M., & Petoukhov S.V. (Ed.), *Adv. Intell. Sys. Comput.* (Vol. 902, hlm. 581–590). Springer Verlag; Scopus. https://doi.org/10.1007/978-3-030-12082-5_53

Gurve, D., Delisle-Rodriguez, D., Bastos-Filho, T., & Krishnan, S. (2020). Trends in compressive sensing for EEG signal processing applications. *Sensors (Switzerland)*, *20*(13), 1–21. Scopus. https://doi.org/10.3390/s20133703

Hei, Y., Yuan, T., Fan, Z., Yang, B., & Hu, J. (2022). Sleep staging classification based on a new parallel fusion method of multiple sources signals. *Physiological Measurement*, *43*(4). Scopus. https://doi.org/10.1088/1361-6579/ac647b

Hulsen, T. (2021). Challenges and solutions for big data in personalized healthcare. Dalam *Big Data in Psychiatry and Neurology* (hlm. 69–94). Elsevier; Scopus. https://doi.org/10.1016/B978-0-12-822884-5.00016-7

Irmak, E. (2020). A Novel Deep Convolutional Neural Network Model for COVID-19 Disease Detection. *TIPTEKNO - Tip Teknol. Kongresi - Med. Technol. Congr., TIPTEKNO*. TIPTEKNO 2020 - Tip Teknolojileri Kongresi - 2020 Medical Technologies Congress, TIPTEKNO 2020. Scopus. https://doi.org/10.1109/TIPTEKNO50054.2020.9299286

Jim, H. S. L., Hoogland, A. I., Brownstein, N. C., Barata, A., Dicker, A. P., Knoop, H., Gonzalez, B. D., Perkins, R., Rollison, D., Gilbert, S. M., Nanda, R., Berglund, A., Mitchell, R., & Johnstone, P. A. S. (2020). Innovations in research and clinical care using patient-generated health data. *CA Cancer Journal for Clinicians*, *70*(3), 182–199. Scopus. https://doi.org/10.3322/caac.21608

Kim, H.-S. (2024). Dark Data in Real-World Evidence: Challenges, Implications, and the Imperative of Data Literacy in Medical Research. *Journal of Korean Medical Science*, *39*(9). Scopus. https://doi.org/10.3346/jkms.2024.39.e92

Luo, Y., Feng, H., Weng, X., Huang, K., & Zheng, H. (2019). A novel oversampling method based on SeqGAN for imbalanced text classification. Dalam Baru C., Huan J., Khan L., Hu X.T., Ak R., Tian Y., Barga R., Zaniolo C., Lee K., & Ye

Y.F. (Ed.), *Proc. - IEEE Int. Conf. Big Data, Big Data* (hlm. 2891–2894). Institute of Electrical and Electronics Engineers Inc.; Scopus. https://doi.org/10.1109/BigData47090.2019.9006138

Lyu, X., Rani, S., Manimurugan, S., Maple, C., & Feng, Y. (2024). A Deep Neuro-Fuzzy Method for ECG Big Data Analysis via Exploring Multimodal Feature Fusion. *IEEE Transactions on Fuzzy Systems*, 1–13. Scopus. https://doi.org/10.1109/TFUZZ.2024.3416217

Ma, Y., Jiang, G., & Chang, T. (2020). Research on motor imagery EEG signal processing algorithm. Dalam Xiao X. & Zhou B. (Ed.), *J. Phys. Conf. Ser.* (Vol. 1656, Nomor 1). IOP Publishing Ltd; Scopus. https://doi.org/10.1088/1742-6596/1656/1/012028

Ma, Z. (2021). Improved Convolutional Neural Network for Biomedical Text Classification. *J. Phys. Conf. Ser.*, *1883*(1). Scopus. https://doi.org/10.1088/1742-6596/1883/1/012080

Majid, M., Butt, A. R., Nizami, I. F., Arsalan, A., & Ryu, J. (2024). PROPER: Personality Recognition based on Public Speaking using Electroencephalography Recordings. *IEEE Access*, 1–1. Scopus. https://doi.org/10.1109/ACCESS.2024.3395434

Obeid, I., Selesnick, I., & Picone, J. (2020). Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications. Dalam *Signal Processing in Medicine and Biology: Emerg. Trends in Research and Applications* (hlm. 281). Springer International Publishing; Scopus. https://doi.org/10.1007/978-3-030-36844-9

Ottenbacher, K. J., Graham, J. E., & Fisher, S. R. (2019). Data Science in Physical Medicine and Rehabilitation: Opportunities and Challenges. *Physical Medicine and Rehabilitation Clinics of North America*, *30*(2), 459–471. Scopus. https://doi.org/10.1016/j.pmr.2018.12.003

Qin, F., & Yuan, J. (2020). Research status and trend of artificial intelligence in the diagnosis of urinary diseases. *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi*, *37*(2), 230–235. Scopus. https://doi.org/10.7507/1001-5515.201910055

Radha, R., & Gopalakrishnan, R. (2022). Use of machine learning and deep learning in healthcare-a review on disease prediction system. Dalam *Fundam. And Methods of Mach. And Deep Learn.: Algorithms, Tools, and Appl.* (hlm. 135–152). wiley; Scopus. https://doi.org/10.1002/9781119821908.ch6

Rakesh Kumar, S., Gayathri, N., Muthuramalingam, S., Balamurugan, B., Ramesh, C., & Nallakaruppan, M. K. (2019). Medical Big Data Mining and Processing in e-Healthcare. Dalam *Internet of Things in Biomedical Engineering* (hlm. 323–339). Elsevier; Scopus. https://doi.org/10.1016/B978-0-12-817356-5.00016-4

Sharma, A., & Colonna, G. (2021). System-Wide Pollution of Biomedical Data: Consequence of the Search for Hub Genes of Hepatocellular Carcinoma Without Spatiotemporal Consideration. *Molecular Diagnosis and Therapy*, *25*(1), 9–27. Scopus. https://doi.org/10.1007/s40291-020-00505-3

Viloria, A., Lezama, O. B. P., & Mercado-Caruzo, N. (2020). Unbalanced data processing using oversampling: Machine learning. Dalam Shakshuki E., Yasar A-U-H., & Malik H. (Ed.), *Procedia Comput. Sci.* (Vol. 175, hlm. 108–113). Elsevier B.V.; Scopus. https://doi.org/10.1016/j.procs.2020.07.018

Wan, D., Zhang, L., Bai, Y., & Xie, Y. (2021). Research on Identification Algorithm Based on ECG Signal and Improved Convolutional Neural Network. *J. Phys. Conf. Ser.*, *1757*(1). Scopus. https://doi.org/10.1088/1742-6596/1757/1/012046

Xue, P. (2021). Development of High-Performance Optical Coherence Tomography. *Zhongguo Jiguang/Chinese Journal of Lasers*, *48*(15). Scopus. https://doi.org/10.3788/CJL202148.1517001
