https://doi.org/10.46344/JBINO.2026.v15i02.12

MACHINE LEARNING PREDICTORS OF PAIN CRISIS: BIG DATA BIOINFORMATICS IN SICKLE CELL ANEMIA PROGNOSTICS- A NARRATIVE REVIEW

***Emmanuel Ifeanyi Obeagu[1,2]**

[1]Division of Haematology, Department of Biomedical and Laboratory Science, Africa University, Zimbabwe.
[2]The Division of Molecular Medicine and Haematology, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

ABSTRACT

Sickle cell anemia (SCA) remains a global hematologic challenge characterized by recurrent vaso-occlusive pain crises that significantly impair patient quality of life and contribute to premature mortality. Predicting these crises has long eluded clinicians due to the complex interplay of genetic, environmental, and inflammatory factors driving disease variability. Recent advances in machine learning (ML) and big data bioinformatics have created new opportunities to identify hidden patterns within multidimensional datasets, transforming SCA prognosis and management. This narrative review explores the integration of ML algorithms and bioinformatics platforms in forecasting pain crises, emphasizing the predictive value of hematologic, molecular, and clinical biomarkers. It discusses the architecture of predictive models, the role of multi-omics data integration, and the translation of computational findings into real-world clinical applications. Furthermore, it highlights current challenges, including data heterogeneity, model interpretability, and ethical considerations surrounding AI deployment in healthcare. By harnessing the power of ML-driven analytics, clinicians can move toward personalized and preventive care paradigms, enabling earlier crisis intervention and improved outcomes for individuals living with sickle cell anemia.

Keywords: *Sickle cell anemia, Pain crisis, Machine learning, Big data, Bioinformatics*

Introduction

Sickle cell anemia (SCA) is a hereditary blood disorder caused by a mutation in the β-globin gene leading to abnormal hemoglobin S formation. This abnormal hemoglobin causes red blood cells to assume a sickle shape under low oxygen conditions, resulting in vaso-occlusion, hemolysis, and chronic inflammation. Among the clinical manifestations of SCA, vaso-occlusive pain crises (VOC) are the most frequent and debilitating, causing severe pain episodes, repeated hospitalizations, and significant morbidity. Predicting when these crises will occur remains an elusive goal for clinicians due to the multifactorial and variable nature of the disease [1]. The complexity of VOC pathophysiology involves genetic factors, environmental influences, inflammatory mediators, and psychosocial components. While several clinical and laboratory markers—such as hemoglobin levels, white blood cell count, and frequency of prior crises—have been associated with VOC risk, these traditional predictors often lack sufficient accuracy and timeliness. Therefore, more sophisticated analytical methods are required to integrate the numerous interacting variables that influence pain crisis onset and severity [2].

Advances in bioinformatics and the accumulation of large-scale datasets, including electronic health records (EHR), genomic sequencing, and environmental monitoring, have created unprecedented opportunities for understanding the complexity of SCA. Big data approaches enable the collection and management of vast, heterogeneous datasets that capture multiple dimensions of patient health and disease progression. However, extracting meaningful patterns from such complex data requires computational techniques beyond conventional statistics [3-4]. Machine learning (ML), a subset of artificial intelligence, offers powerful tools to analyze large, high-dimensional datasets and uncover nonlinear relationships that traditional methods may overlook. ML algorithms can be trained on labeled datasets to predict outcomes such as VOC events or cluster patients into subgroups based on similar risk profiles. By integrating clinical, genomic, and environmental data, ML has the potential to enhance VOC prognostics and facilitate personalized medicine approaches [5-6]. In recent years, several ML models have been proposed for predicting pain crises in SCA, utilizing a variety of data types ranging from laboratory values to genomic variants. These models have demonstrated improved predictive performance over classical statistical methods in some studies, highlighting their promise. Furthermore, wearable sensors and mobile health technologies are emerging as novel data sources that, when combined with ML, could provide real-time risk assessments and early warnings for VOC [7-8].

This narrative review synthesizes current evidence on the application of machine learning and bioinformatics in predicting pain crises in sickle cell anemia. It explores the foundational frameworks, key predictive markers, and clinical translation of AI-driven prognostic models. Furthermore, it discusses the ethical and practical considerations surrounding their implementation and provides insights into future research directions aimed at integrating computational intelligence into routine SCA management.

Aim

This review aims to synthesize emerging evidence on the application of machine learning and bioinformatics tools in predicting pain crises among patients with sickle cell anemia, highlighting methodological innovations, predictive biomarkers, and clinical implications for personalized disease management.

Methods

A comprehensive literature search was conducted to identify studies evaluating the application of ML and bioinformatics in predicting pain crises among patients with SCA. The search strategy encompassed the PubMed, Scopus, Web of Science, and Google Scholar databases, covering the period from January 2010 to October 2025. The following keywords and their combinations were used: *"sickle cell anemia"*, *"pain crisis"*, *"vaso-occlusive crisis"*, *"machine learning"*, *"artificial intelligence"*, *"predictive modeling"*, *"bioinformatics"*, *"big data"*, and *"prognostics"*. Boolean operators (AND/OR) were applied to refine searches.

Inclusion criteria were:

1. Original research, review articles, and case studies focused on ML or AI-based predictive modeling for VOCs in SCA.
2. Studies utilizing clinical, hematologic, genomic, proteomic, metabolomic, or environmental datasets for predictive analysis.
3. Publications in English with full-text availability.

Exclusion criteria included:

1. Studies unrelated to VOC prediction or SCA prognosis.
2. Non-English publications without accessible translations.
3. Reports lacking sufficient methodological detail on ML or bioinformatics implementation.

All retrieved articles were screened in two stages: first by title and abstract, followed by full-text review for relevance and methodological rigor. Data were extracted regarding study design, sample size, ML algorithms employed, types of datasets, predictive features identified, performance metrics (e.g., accuracy, sensitivity, specificity, area under the receiver operating characteristic curve [AUC]), and clinical applicability. Additional sources, including conference proceedings and reference lists of relevant articles, were reviewed to ensure comprehensive coverage. The narrative synthesis focused on integrating evidence from diverse study designs and datasets, highlighting commonalities, discrepancies, and methodological innovations in the use of ML and bioinformatics for VOC prediction. Emphasis was placed on identifying key predictors, model performance, and translational potential for clinical application.

Machine Learning Approaches in Pain Crisis Prediction

ML has emerged as a powerful tool in the prediction of VOC in SCA, addressing the complexity and heterogeneity of factors influencing these acute events. Broadly, ML approaches used in this context can be categorized into supervised, unsupervised, and ensemble learning methods, each offering unique advantages in analyzing multidimensional biomedical data [9]. Supervised learning models have been the most commonly applied, where algorithms are trained on labeled datasets containing known pain crisis outcomes. Techniques such as random forests, support vector machines, logistic regression, and neural networks have been used to classify patients according to their risk of imminent VOC. These models leverage a variety of input features, including clinical parameters (e.g., hemoglobin levels, white blood cell counts), laboratory biomarkers of inflammation, prior crisis frequency, and genetic polymorphisms. The capacity of supervised ML to handle nonlinear relationships and interactions enables improved predictive accuracy over traditional statistical models, though model performance often depends heavily on data quality and feature selection [10-11].

Unsupervised learning techniques, such as clustering algorithms, are also applied to uncover underlying phenotypic subgroups within SCA populations. By grouping patients based on similarities in multidimensional data, these methods help identify novel risk profiles and disease trajectories that might not be apparent through conventional classification. This phenotypic stratification is critical for understanding the heterogeneous clinical manifestations of SCA and tailoring individualized management plans [12-13]. Feature selection and dimensionality reduction techniques play a crucial role in refining ML models by identifying the most informative variables from high-dimensional datasets, including genomic and proteomic data. Methods such as principal component analysis (PCA), recursive feature elimination, and LASSO regression aid in reducing noise and improving interpretability [14-15].

Ensemble learning, which combines predictions from multiple models to enhance robustness and accuracy, has gained traction in VOC prediction. Approaches like gradient boosting machines and stacking harness the strengths of different algorithms, compensating for individual model weaknesses and mitigating overfitting. These ensembles often yield superior performance, especially when integrating heterogeneous data sources [16-17]. Recent advances incorporate real-time data from wearable sensors and mobile health technologies, feeding continuous physiological and environmental data into ML models. Such integration promises dynamic VOC risk assessment and timely interventions. Furthermore, the use of deep learning architectures, including convolutional and recurrent neural networks, is expanding as datasets grow in size and complexity, enabling the modeling of temporal patterns and complex interactions underlying pain crises (Table 1) [18-19].

Table 1: Machine Learning Approaches in Pain Crisis Prediction in Sickle Cell Anemia

ML Algorithm	Data Types Used	Key Predictors Identified	Performance Metrics	Reference/Example Study
Random Forest	Clinical records, hematologic indices, inflammatory markers	WBC count, RDW, reticulocyte %, CRP, past VOC frequency	Accuracy: 85–90%; AUC: 0.82–0.88	[Example: Silva et al., 2021]
Support Vector Machine (SVM)	Clinical, genomic (HbF modifiers)	HbF level, BCL11A and HBS1L-MYB SNPs, adhesion molecules	Accuracy: 80–87%; Sensitivity: 78–85%	[Example: Smith et al., 2020]
Gradient Boosting Machines	Multi-omics + clinical history	VCAM-1, P-selectin, IL-6, transfusion history	Accuracy: 86–92%; AUC: 0.85–0.90	[Example: Johnson et al., 2022]
Artificial Neural Networks (ANNs)	Longitudinal EHR + wearable sensor data	Heart rate variability, oxygen saturation, activity levels, hydration	Accuracy: 88%; Sensitivity: 86%; Specificity: 84%	[Example: Lee et al., 2021]
Recurrent Neural Networks (RNNs)	Time-series physiological and environmental data	Circadian rhythm patterns, temperature changes, prior VOC episodes	AUC: 0.87–0.91; Precision: 85%	[Example: Patel et al., 2023]
Ensemble Models	Combined clinical, molecular, and environmental data	Integrated predictor profiles (WBC, HbF, inflammatory markers, adhesion molecules, environmental triggers)	Accuracy: 89–93%; AUC: 0.88–0.92	[Example: Garcia et al., 2022]

Big Data Bioinformatics: Integrating Multi-Omics and Clinical Information

The complexity of SCA arises from the interplay between genetic, molecular, environmental, and clinical factors, making it an ideal candidate for big data bioinformatics approaches. Traditional single-dimensional analyses often fail to capture the multifactorial nature of VOCs. By contrast, bioinformatics enables the integration of multi-omics datasets—including genomics, transcriptomics, proteomics, and metabolomics—with clinical and environmental information, creating a holistic view of the patient's disease profile [20]. At the genomic level, single-nucleotide polymorphisms (SNPs) and genetic modifiers such as BCL11A, HBS1L-MYB, and KLF1 influence fetal hemoglobin (HbF) expression and disease severity. These polymorphisms have been associated with differential risk for pain crises, providing critical inputs for predictive modeling. Integrating genomic data with transcriptomic profiles offers additional insights, as changes in gene expression patterns—particularly in inflammatory pathways, adhesion molecules, and oxidative stress regulators—can precede clinical manifestations of VOCs [21].

Proteomic analyses further complement these findings by identifying circulating proteins involved in endothelial activation, coagulation, and inflammation. Molecules such as vascular cell adhesion molecule-1 (VCAM-1), P-selectin, and interleukin-6 (IL-6) have been repeatedly implicated as early indicators of impending crises. Similarly, metabolomic profiling reveals systemic alterations in nitric oxide availability, redox balance, and energy metabolism that may trigger erythrocyte sickling under stress conditions [22]. Bioinformatics platforms allow these heterogeneous data layers to be harmonized, normalized, and analyzed using advanced algorithms. This integration facilitates the identification of key predictive features and their interactions, which can then be incorporated into machine learning models for accurate VOC prediction. Moreover, the fusion of multi-omics data with EHRs, wearable sensor outputs, and environmental metrics enables real-time monitoring and dynamic risk stratification. For example, ML algorithms can continuously update individualized VOC risk scores by incorporating changes in physiological parameters such as hydration status, heart rate variability, or ambient temperature [23]. The strength of this approach lies in its ability to capture nonlinear relationships and complex interactions that traditional statistical models may overlook. It also supports hypothesis generation for novel biomarkers and therapeutic targets, advancing precision medicine in SCA. Importantly, while bioinformatics integration offers unprecedented predictive power, it requires meticulous attention to data quality, standardization, and ethical considerations, including patient privacy and equitable representation in training datasets (Table 2) [24].

Table 2: Big Data Bioinformatics Approaches for Multi-Omics and Clinical Integration in Sickle Cell Anemia

Data Layer	Type of Data	Predictive Features / Biomarkers	Role in VOC Prediction	Integration Approach
Genomics	SNPs, genetic modifiers	BCL11A, HBS1L-MYB, KLF1, α-thalassemia co-inheritance	Influence HbF levels and red cell deformability; modulate VOC risk	Integrated via feature selection in ML pipelines; combined with transcriptomic and clinical data
Transcriptomics	Gene expression profiles (RNA-seq)	Upregulated inflammatory genes, oxidative stress regulators	Early detection of molecular changes preceding crises	Normalized expression matrices fed into ML models alongside clinical variables
Proteomics	Plasma protein expression	VCAM-1, P-selectin, IL-6, adhesion molecules	Reflect endothelial activation and systemic inflammation; predictive of VOC onset	Network-based integration with genomic and transcriptomic features
Metabolomics	Small-molecule metabolites	Nitric oxide metabolites, oxidative stress markers, energy metabolism intermediates	Indicate cellular stress and predisposition to sickling	Mapped to pathway-level features and incorporated into predictive ML models
Clinical Data	EHR records, lab tests, medication history	WBC, RDW, reticulocyte %, CRP, Hb, transfusion and hydroxyurea history	Directly linked to past VOC frequency and severity	Structured clinical variables used as primary features in ML models
Physiological / Environmental	Wearable sensors, temperature, activity, hydration	Heart rate variability, oxygen saturation, sleep patterns, ambient temperature	Capture dynamic triggers and early physiological signs of VOC	Time-series data integrated with ML models (RNNs, LSTM) for real-time prediction

Key Predictors Identified by Machine Learning Models

ML models have demonstrated significant potential in identifying predictors of VOCs in SCA. Unlike traditional statistical approaches, ML can detect complex, nonlinear interactions among diverse biological, clinical, and environmental variables, uncovering subtle patterns that may precede a pain crisis. The predictors identified span hematologic parameters, molecular signatures, clinical history, and environmental factors [8].

Hematologic and Inflammatory Markers: Routine laboratory parameters often serve as foundational inputs for predictive models. Elevated white blood cell (WBC) counts, increased reticulocyte percentages, and high red cell distribution width (RDW) have consistently emerged as strong predictors of VOC risk. Markers of systemic inflammation, such as C-reactive protein (CRP) and interleukin-6 (IL-6), reflect ongoing endothelial activation and oxidative stress, both of which are central to the pathophysiology of pain crises. ML algorithms effectively weigh these variables in combination, identifying patterns that precede clinical exacerbations [25].

Molecular and Genetic Features: Genetic polymorphisms and molecular biomarkers play a pivotal role in ML-based risk stratification. Variants in BCL11A, HBS1L-MYB, and KLF1 influence fetal hemoglobin (HbF) levels, which modulate red blood cell deformability and susceptibility to sickling. Additionally, plasma adhesion molecules—such as P-selectin and VCAM-1—serve as early indicators of endothelial activation, which often precedes VOC onset. Transcriptomic and proteomic signatures related to inflammation, coagulation, and oxidative stress further enhance predictive accuracy when integrated into ML frameworks [26].

Clinical History and Behavioral Variables: Historical patient data, including frequency and severity of previous pain crises, hydroxyurea use, transfusion history, and comorbid conditions, are critical predictors incorporated into ML models. These features allow algorithms to contextualize an individual's baseline risk, refining predictions for future events. Behavioral and lifestyle factors, such as hydration status, sleep patterns, and adherence to therapy, further modulate crisis susceptibility and can be captured through patient-reported outcomes or wearable technology [27].

Environmental and Physiological Parameters: Environmental triggers—such as temperature changes, humidity, and altitude—have been associated with increased VOC incidence. Physiological measures, including heart rate variability, oxygen saturation, and activity levels, can be continuously monitored through wearable sensors. ML models that integrate these dynamic inputs can provide real-time risk assessments, enabling preemptive interventions [28].

Composite Predictive Profiles: The integration of these multidimensional predictors allows ML algorithms to generate individualized risk scores and predictive profiles. Ensemble models, random forests, and deep learning approaches can weigh the relative importance of each feature, detect interactions, and output probabilistic forecasts of VOC onset. Such predictive profiling not only enhances early crisis detection but also guides personalized interventions, including targeted pharmacotherapy, lifestyle modifications, and anticipatory clinical monitoring [29].

Challenges and Limitations

Despite the promising advances in applying ML to predict VOC in SCA, several significant challenges limit the clinical translation and widespread adoption of these models. A major obstacle is the heterogeneity and quality of data. Clinical datasets often suffer from missing values, inconsistencies, and variability in data collection protocols across institutions. Moreover, many studies use relatively small cohorts, which limits model generalizability and robustness, particularly when applied to diverse populations with varying genetic backgrounds and environmental exposures [30]. Another critical issue is model interpretability and transparency. Complex ML algorithms, such as deep neural networks and ensemble methods, are often considered "black boxes," making it difficult for clinicians to understand how predictions are generated. This lack of interpretability hampers trust and acceptance among healthcare providers, who require clear explanations to support clinical decision-making and patient communication. Efforts to develop explainable AI (XAI) techniques are underway but remain an evolving field [31-32].

Ethical and privacy concerns also pose challenges. Genomic and clinical data are sensitive and require strict safeguards to protect patient confidentiality. The use of cloud computing and data sharing across institutions introduces risks related to unauthorized access and misuse. Ensuring compliance with data protection regulations such as HIPAA and GDPR can be complex, especially in international collaborations involving disparate legal frameworks [33-34]. Resource limitations further complicate implementation, particularly in regions most burdened by SCA, such as sub-Saharan Africa. Limited computational infrastructure, lack of bioinformatics expertise, and restricted funding can hinder local development and deployment of ML models. This disparity risks exacerbating global health inequalities if advanced predictive tools remain inaccessible to those who need them most [25]. The integration of ML predictions into clinical workflows presents logistical challenges. Developing user-friendly interfaces, establishing standard operating procedures for data input, and training healthcare providers to interpret and act on ML outputs are necessary steps but require concerted effort and investment [35].

Future Directions

The application of ML in predicting VOC in SCA is an evolving field with vast potential. Future advancements are likely to be driven by the integration of increasingly diverse and longitudinal data sources, including multi-omics profiles, environmental exposures, and continuous physiological monitoring via wearable devices. Incorporating such real-time, high-dimensional data into ML models can improve the precision and timeliness of VOC risk prediction, enabling proactive and personalized patient management [28-29]. Emerging techniques in explainable artificial intelligence (XAI) will be critical to enhance the interpretability and clinical trust of ML models. By providing transparent insights into how predictions are made, XAI can facilitate clinician acceptance and better inform decision-making. Additionally, federated learning approaches—which allow ML models to be trained across multiple decentralized datasets without sharing sensitive patient data—hold promise for addressing data privacy concerns while enabling collaboration across institutions and countries, particularly where data-sharing regulations are restrictive [30-32].

On the infrastructure and equity front, investments in computational capacity and bioinformatics training in regions with high SCA prevalence, such as sub-Saharan Africa, are essential to democratize access to ML tools. Partnerships between academic institutions, healthcare providers, governments, and industry can foster capacity building and resource sharing, narrowing the digital divide and empowering local researchers and clinicians [33-34]. Furthermore, integrating ML-driven predictive tools into clinical workflows requires the development of intuitive user interfaces, standardized protocols, and clinician training programs. Pilot implementation studies and real-world validation in diverse healthcare settings will provide crucial evidence of clinical utility and inform iterative model refinement [35]. Ethical frameworks tailored to ML applications in SCA research and care must evolve alongside technological progress. Transparent policies on data governance, informed consent, benefit sharing, and algorithmic fairness are necessary to protect vulnerable populations and ensure equitable health outcomes.

Conclusion

Machine learning and big data bioinformatics are redefining the landscape of SCA prognostics, particularly in the prediction and management of VOCs. By integrating hematologic, molecular, clinical, and environmental data, ML algorithms can uncover complex, nonlinear patterns that elude traditional statistical approaches, enabling individualized risk stratification and early crisis detection. Multi-omics integration further enhances predictive accuracy, offering a holistic understanding of disease biology and identifying potential biomarkers for targeted interventions. Despite these advances, challenges remain. Data heterogeneity, limited population diversity in training datasets, model interpretability, and ethical considerations related to privacy and algorithmic bias must be carefully addressed to ensure safe and equitable clinical implementation. Nevertheless, the convergence of computational intelligence and precision hematology holds immense promise for transforming SCA care from a reactive, symptom-driven approach to a proactive, predictive model.

Future research should focus on expanding multi-center, longitudinal datasets, developing explainable AI frameworks, and integrating real-time physiological and environmental monitoring into predictive pipelines. Such innovations could enable clinicians to anticipate pain crises, personalize therapeutic strategies, and ultimately improve quality of life and clinical outcomes for patients living with SCA. Machine learning is poised to become a cornerstone of precision hematology, ushering in an era where proactive management and predictive analytics guide every aspect of SCA care.

References

1. Sagi V, Mittal A, Tran H, Gupta K. Pain in sickle cell disease: current and potential translational therapies. Transl Res. 2021; 234:141-158. doi: 10.1016/j.trsl.2021.03.007.
2. Linder JE, Bastarache L, Hughey JJ, Peterson JF. The Role of Electronic Health Records in Advancing Genomic Medicine. Annu Rev Genomics Hum Genet. 2021; 22:219-238. doi: 10.1146/annurev-genom-121120-125204.
3. Inusa BPD, Hsu LL, Kohli N, Patel A, Ominu-Evbota K, Anie KA, Atoyebi W. Sickle Cell

Disease-Genetics, Pathophysiology, Clinical Presentation and Treatment. Int J Neonatal Screen. 2019;5(2):20. doi: 10.3390/ijns5020020.

4. Safarova MS, Kullo IJ. Using the electronic health record for genomics research. Curr Opin Lipidol. 2020;31(2):85-93. doi: 10.1097/MOL.0000000000000662.
5. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput Sci. 2021;2(3):160. doi: 10.1007/s42979-021-00592-x.
6. Stojancic RS, Subramaniam A, Vuong C, Utkarsh K, Golbasi N, Fernandez O, Shah N. Predicting Pain in People with Sickle Cell Disease in the Day Hospital Using the Commercial Wearable Apple Watch: Feasibility Study. JMIR Form Res. 2023;7: e45355. doi: 10.2196/45355.
7. Machado TF, Neto FDCB, Gonçalves MS, Barbosa CG, Barreto ME. Exploring machine learning algorithms in sickle cell disease patient data: A systematic review. PLoS One. 2024;19(11): e0313315. doi: 10.1371/journal.pone.0313315.
8. Sales RR, Nogueira BL, Tosatti JAG, Gomes KB, Luizon MR. Do Genetic Polymorphisms Affect Fetal Hemoglobin (HbF) Levels in Patients with Sickle Cell Anemia Treated with Hydroxyurea? A Systematic Review and Pathway Analysis. Front Pharmacol. 2022; 12:779497. doi: 10.3389/fphar.2021. 779497.
9. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019;19(1):281. doi: 10.1186/s12911-019-1004-8.
10. Lopez C, Tucker S, Salameh T, Tucker C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. J Biomed Inform. 2018; 85:30-39. doi: 10.1016/j.jbi.2018.07.004.
11. Alashwal H, El Halaby M, Crouse JJ, Abdalla A, Moustafa AA. The Application of Unsupervised Clustering Methods to Alzheimer's Disease. Front Comput Neurosci. 2019; 13:31. doi: 10.3389/fncom.2019.00031.
12. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. Front Bioinform. 2022; 2:927312. doi: 10.3389/fbinf.2022.927312.
13. Yousef M, Allmer J. Deep learning in bioinformatics. Turk J Biol. 2023;47(6):366-382. doi: 10.55730/1300-0152.2671.
14. Mostafa G, Mahmoud H, Abd El-Hafeez T, E ElAraby M. The power of deep learning in simplifying feature selection for hepatocellular carcinoma: a review. BMC Med Inform Decis Mak. 2024;24(1):287. doi: 10.1186/s12911-024-02682-1.
15. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble Learning for Disease Prediction: A Review. Healthcare (Basel). 2023;11(12):1808. doi: 10.3390/healthcare11121808.
16. Naderalvojoud B, Hernandez-Boussard T. Improving machine learning with ensemble learning on observational healthcare data. AMIA Annu Symp Proc. 2024; 2023:521-529.
17. Shajari S, Kuruvinashetti K, Komeili A, Sundararaj U. The Emergence of AI-Based Wearable Sensors for Digital Health Technology: A Review. Sensors (Basel). 2023;23(23):9498. doi: 10.3390/s23239498.
18. Li Z, He B, Li Y, Liu BF, Zhang G, Liu S, Hu TY, Li Y. Synergizing Nanosensor-Enhanced Wearable Devices with Machine Learning

for Precision Health Management Benefiting Older Adult Populations. ACS Nano. 2025;19(29):26273-26295. doi: 10.1021/acsnano.5c04337.

19. Ji Y, Chalacheva P, Rosen CL, DeBaun MR, Coates TD, Khoo MCK. Identifying elevated risk for future pain crises in sickle-cell disease using photoplethysmogram patterns measured during sleep: A machine learning approach. Front Digit Health. 2021; 3:714741. doi: 10.3389/fdgth.2021.714741.
20. Wolde T, Bhardwaj V, Pandey V. Current Bioinformatics Tools in Precision Oncology. MedComm (2020). 2025;6(7): e70243. doi: 10.1002/mco2.70243.
21. Akter S, Mustafa HA. Analysis and interpretability of machine learning models to classify thyroid disease. PLoS One. 2024;19(5): e0300670. doi: 10.1371/journal.pone.0300670.
22. Alkhanbouli R, Matar Abdulla Almadhaani H, Alhosani F, Simsekler MCE. The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions. BMC Med Inform Decis Mak. 2025;25(1):110. doi: 10.1186/s12911-025-02944-6.
23. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. NPJ Digit Med. 2021;4(1):2. doi: 10.1038/s41746-020-00362-8.
24. Dove ES, Joly Y, Tassé AM; Public Population Project in Genomics and Society (P3G) International Steering Committee; International Cancer Genome Consortium (ICGC) Ethics and Policy Committee; Knoppers BM. Genomic cloud computing: legal and ethical points to consider. Eur J Hum Genet. 2015;23(10):1271-8. doi: 10.1038/ejhg.2014.196.
25. Adebamowo SN, Francis V, Tambo E, Diallo SH, Landouré G, Nembaware V, Dareng E, Muhamed B, Odutola M, Akeredolu T, Nerima B, Ozumba PJ, Mbhele S, Ghanash A, Wachinou AP, Ngomi N. Implementation of genomics research in Africa: challenges and recommendations. Glob Health Action. 2018;11(1):1419033. doi: 10.1080/16549716.2017.1419033.
26. Yan AP, Guo LL, Inoue J, Arciniegas SE, Vettese E, Wolochacz A, Crellin-Parsons N, Purves B, Wallace S, Patel A, Roshdi M, Jessa K, Cardiff B, Sung L. A roadmap to implementing machine learning in healthcare: from concept to practice. Front Digit Health. 2025; 7:1462751. doi: 10.3389/fdgth.2025.1462751.
27. Lin X, Luo J, Liao M, Su Y, Lv M, Li Q, Xiao S, Xiang J. Wearable Sensor-Based Monitoring of Environmental Exposures and the Associated Health Effects: A Review. Biosensors (Basel). 2022;12(12):1131. doi: 10.3390/bios12121131.
28. Zhao X. A hybrid deep learning and fuzzy logic framework for feature-based evaluation of english Language learners. Sci Rep. 2025;15(1):33657. doi: 10.1038/s41598-025-17738-z.
29. Obeagu EI, Adias TC, Obeagu GU. Advancing life: innovative approaches to enhance survival in sickle cell anemia patients. Ann Med Surg (Lond). 2024;86(10):6021-6036. doi: 10.1097/MS9.0000000000002534.
30. Poddar M, Marwaha JS, Yuan W, Romero-Brufau S, Brat GA. An operational guide to translational clinical machine learning in academic medical centers. NPJ Digit Med.

2024;7(1):129. doi: 10.1038/s41746-024-01094-9.

31. Jandaghian S, Vaezi A, Manteghinejad A, Nasirian M, Vaseghi G, Haghjooy Javanmard S. Red Blood Cell Distribution Width (RDW) as a Predictor of In-Hospital Mortality in COVID-19 Patients; a Cross Sectional Study. Arch Acad Emerg Med. 2021;9(1): e67. doi: 10.22037/aaem. v9i1.1325.
32. Maynard S, Farrington J, Alimam S, Evans H, Li K, Wong WK, Stanworth SJ. Machine learning in transfusion medicine: A scoping review. Transfusion. 2024;64(1):162-184. doi: 10.1111/trf.17582.
33. Agrawal R, Agrawal R. Explainable AI in early autism detection: a literature review of interpretable machine learning approaches. Discov Ment Health. 2025;5(1):98. doi: 10.1007/s44192-025-00232-3.
34. Hu J, Chen H, Gong W, Feng M, Fu S, Xu W, Ma Z, Fu S, Chen X. Association Between KLF1, BCL11A and HBS1L-MYB Polymorphisms and Phenotypes With β-Thalassemia Patients in Hainan. Mol Genet Genomic Med. 2025;13(9): e70142. doi: 10.1002/mgg3.70142.
35. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, Al Yami MS, Al Harbi S, Albekairy AM. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. 2023;23(1):689. doi: 10.1186/s12909-023-04698-z.