Medicine®

OPEN

Big data analytics and machine learning in hematology
Transformative insights, applications and challenges

Emmanuel Ifeanyi Obeagu, PhD[a],*[ID], Christiana Uchenna Ezeanya, PhD[b], Fabian Chukwudi Ogenyi, BEng[c], Deborah Domini Ifu, BMLS[a]

Abstract

The integration of big data analytics and machine learning (ML) into hematology has ushered in a new era of precision medicine, offering transformative insights into disease management. By leveraging vast and diverse datasets, including genomic profiles, clinical laboratory results, and imaging data, these technologies enhance diagnostic accuracy, enable robust prognostic modeling, and support personalized therapeutic interventions. Advanced ML algorithms, such as neural networks and ensemble learning, facilitate the discovery of novel biomarkers and refine risk stratification for hematological disorders, including leukemias, lymphomas, and coagulopathies. Despite these advancements, significant challenges persist, particularly in the realms of data integration, algorithm validation, and ethical concerns. The heterogeneity of hematological datasets and the lack of standardized frameworks complicate their application, while the "black-box" nature of ML models raises issues of reliability and clinical trust. Moreover, safeguarding patient privacy in an era of data-driven medicine remains paramount, necessitating the development of secure and ethical analytical practices. Addressing these challenges is critical to ensuring equitable and effective implementation of these technologies. Collaborative efforts between hematologists, data scientists, and bioinformaticians are pivotal in translating these innovations into real-world clinical practice. Emphasis on developing explainable artificial intelligence models, integrating real-time analytics, and adopting federated learning approaches will further enhance the utility and adoption of these technologies. As big data analytics and ML continue to evolve, their potential to revolutionize hematology and improve patient outcomes remains immense.

Abbreviations: AI = artificial intelligence, CNNs = convolutional neural networks, DL = deep learning, EHR = electronic health record, ML = machine learning, NLP = Natural Language Processing

Keywords: artificial intelligence, big data analytics, gene editing, genomics, hematology, machine learning

1. Introduction

The field of hematology has witnessed a paradigm shift with the advent of big data analytics and machine learning (ML). These technologies leverage vast and complex datasets to derive actionable insights, fundamentally transforming approaches to diagnostics, prognostics, and therapeutic interventions. Hematology, with its reliance on intricate data from diverse sources such as genomic sequencing, clinical laboratory tests, and advanced imaging modalities, is uniquely positioned to benefit from these innovations. As the volume of available data grows exponentially, the ability to analyze and interpret it effectively becomes essential to advancing patient care.[1] Big data analytics encompasses a set of methodologies designed to handle massive, high-dimensional datasets that are beyond the capacity of traditional data-processing techniques. In hematology, these datasets are derived from electronic health records (EHRs), high-throughput genomic data, proteomics, and imaging studies, among other sources. The integration of these datasets enables a comprehensive understanding of disease mechanisms, facilitates the discovery of novel biomarkers, and informs the development of targeted therapies. The power of big data lies not only in its size but also in its ability to reveal patterns and correlations that may otherwise go unnoticed.[2] Machine learning, a subset of artificial intelligence, has further enhanced the analytical capabilities of big data. By using algorithms capable of learning from data and making predictions, ML has emerged as a powerful tool in hematology. Techniques such as supervised learning, unsupervised learning, and reinforcement learning are applied to classify diseases, predict patient outcomes, and optimize treatment plans. For instance, convolutional neural networks (CNNs) have shown remarkable success in analyzing peripheral blood smears and

bone marrow aspirates for the detection of hematological malignancies.[3]

The integration of big data analytics and ML in hematology has already demonstrated its potential in various domains. For example, predictive analytics is increasingly being used to stratify patients into risk categories for diseases such as leukemia and thrombosis, enabling timely interventions. Similarly, unsupervised clustering algorithms are uncovering hidden subtypes of diseases, thereby refining diagnostic criteria and enhancing the personalization of therapies. These developments have laid the foundation for precision hematology, a field that aims to tailor medical care to the individual characteristics of each patient.[4] The promise of these technologies lies in their ability to bridge the gap between bench and bedside, transforming basic research findings into clinical applications. This requires interdisciplinary collaboration between hematologists, data scientists, and bioinformaticians to develop robust algorithms and analytical frameworks. Moreover, regulatory bodies must establish guidelines to ensure that ML models are validated and meet safety and efficacy standards before their widespread implementation in clinical settings.[5] Emerging trends in explainable artificial intelligence (XAI) and federated learning offer potential solutions to some of these challenges. XAI aims to make ML models more transparent and interpretable, thereby increasing their acceptance among clinicians. Federated learning, on the other hand, enables collaborative data analysis across institutions while preserving patient privacy, thus addressing issues related to data sharing and security.[6]

2. Aim

The aim of this review is to explore the transformative role of big data analytics and ML in the field of hematology.

3. Rationale

The rapidly growing complexity and volume of data in hematology, fueled by advances in genomics, proteomics, and imaging technologies, demand innovative analytical approaches for effective utilization. Traditional methods of data analysis are increasingly inadequate for managing these datasets and extracting clinically actionable insights. Big data analytics and ML have emerged as transformative tools capable of addressing these challenges by offering powerful solutions for pattern recognition, predictive modeling, and decision support. Hematological disorders, ranging from malignancies such as leukemia to inherited conditions like sickle cell anemia, often require intricate diagnostic and therapeutic approaches. Early detection, precise risk stratification, and personalized treatment are crucial for improving patient outcomes, yet these goals remain difficult to achieve using conventional techniques. Artificial intelligence (AI)-driven methodologies can bridge this gap by enabling more accurate diagnostics, refining treatment protocols, and identifying novel therapeutic targets. Despite their promise, the application of big data analytics and ML in hematology is still in its infancy, facing barriers such as data heterogeneity, lack of standardization, and concerns over model interpretability and ethical implications. This review is warranted to provide a comprehensive overview of how these advancements can be leveraged to revolutionize hematological practice, research, and patient care.

4. Review methodology

4.1. Literature search

A broad literature search was conducted using databases such as PubMed, Scopus, Web of Science, and Google Scholar. Keywords used in the search included:

- "big data analytics in hematology"
- "machine learning in hematology"
- "AI in blood disorders"
- "diagnostics in hematology using AI"
- "predictive analytics in hematology"
- "explainable AI in hematology"
- "federated learning in healthcare"

The search was limited to peer-reviewed articles published in English, with an emphasis on studies and reviews from the past decade to ensure relevance and inclusion of recent advancements.

4.2. Inclusion and exclusion criteria

- Inclusion criteria: Articles were included if they focused on the application of big data analytics or ML in hematology, highlighted advancements in diagnostics, therapeutics, or research, and addressed challenges or emerging solutions in this field.
- Exclusion criteria: Studies were excluded if they lacked specific focus on hematology, provided only theoretical discussions without practical implications, or were inaccessible due to paywalls.

4.3. Evolution of big data analytics in hematology

The evolution of big data analytics in hematology represents a significant milestone in the understanding and management of hematological disorders.[7] Initially, hematology relied predominantly on manual observations, limited datasets, and traditional diagnostic methods.[8] With the advent of computers and electronic data storage, the groundwork for data-driven approaches in hematology was laid.[9] The digitization of patient records, advancements in diagnostic technologies, and the proliferation of genomic data repositories led to the accumulation of vast datasets in hematology.[10] This influx of diverse data types laid the foundation for big data analytics. The deciphering of the human genome and subsequent advancements in genomic sequencing technologies triggered a revolution in hematology.[11] Genomic profiling of hematological disorders became feasible, revealing intricate genetic alterations and disease-associated mutations.[12] Beyond genomics, the integration of other "omics" data (such as transcriptomics, proteomics, and metabolomics) expanded the scope of understanding hematological disorders.[13] This multi-omics approach provided a holistic view of disease biology. Data mining techniques and computational algorithms started gaining prominence, enabling the extraction of meaningful insights from large datasets.[14] These approaches aided in pattern recognition, identifying disease signatures, and predicting treatment responses. The advent of ML and AI propelled hematology into a new era. ML algorithms enabled the development of predictive models, risk stratification tools, and diagnostic decision support systems, revolutionizing disease management.[15]

Digital imaging technologies facilitated the digitization of blood smears, bone marrow biopsies, and histopathological slides.[16] Machine learning algorithms applied to digital pathology revolutionized diagnostic accuracy and automated disease classification.[17] The integration of wearable devices and real-time monitoring tools contributed to continuous data collection.[18] These technologies provided clinicians with dynamic patient data, enhancing personalized treatment strategies. Collaborative research initiatives, consortia, and data-sharing platforms emerged, fostering collaborative efforts among researchers, clinicians, and institutions. This sharing of data and knowledge accelerated research and improved data quality.[19] With the burgeoning use of big data in hematology, ethical considerations, patient privacy, and regulatory frameworks became crucial.[20] Guidelines and ethical standards were developed to

govern the ethical use and protection of patient data.[21] The evolution of big data analytics in hematology has been characterized by a shift from conventional diagnostic and treatment paradigms to data-driven, precision medicine approaches.[22] As technologies continue to advance, the integration of big data analytics and AI-driven solutions is poised to redefine hematological research, diagnosis, and therapeutic interventions, promising more precise and personalized care for patients with hematological disorders.[23]

4.3.1. AI in blood disorders.
AI is significantly advancing the diagnosis, management, and understanding of various blood disorders. By leveraging ML and deep learning (DL) techniques, AI has proven to be a transformative tool in analyzing complex data sets, enhancing diagnostic accuracy, predicting outcomes, and personalizing treatments.

4.3.1.1. AI in hematological malignancies.
AI has shown remarkable potential in diagnosing and monitoring hematological cancers such as leukemia, lymphoma, and multiple myeloma. Machine learning models, including CNNs, have been utilized to analyze microscopic images of blood smears and bone marrow aspirates. These models can identify abnormal cell morphologies and classify subtypes of leukemia with accuracy comparable to, and sometimes exceeding, that of expert pathologists. Furthermore, predictive algorithms have been developed to assess patient-specific risks of progression or relapse, aiding in personalized treatment planning.[8]

4.3.1.2. AI in hemoglobinopathies.
Conditions such as sickle cell anemia and thalassemia have benefited from AI-powered tools designed to optimize disease management. AI algorithms are employed to detect abnormal hemoglobin patterns in electrophoresis data, enabling faster and more reliable diagnoses. In sickle cell anemia, AI models are being used to predict vaso-occlusive crisis episodes by analyzing biomarkers, clinical data, and patient-reported symptoms. Such predictive capabilities allow for timely intervention, reducing hospitalizations and improving patient outcomes.[9]

4.3.1.3. AI in coagulation disorders.
AI has revolutionized the management of bleeding and clotting disorders, including hemophilia and thrombophilia. Predictive models are used to assess bleeding risks, optimize clotting factor replacement therapy, and personalize treatment schedules. Deep learning algorithms are also being integrated into coagulation assays, improving the interpretation of test results and aiding in the diagnosis of conditions such as disseminated intravascular coagulation and antiphospholipid syndrome.[10]

4.3.1.4. AI in bone marrow failure syndromes.
In diseases like aplastic anemia and myelodysplastic syndromes, AI is enabling earlier detection through the analysis of genomic, transcriptomic, and proteomic data. AI algorithms can identify genetic mutations and epigenetic changes associated with these disorders, facilitating targeted therapy development. Additionally, predictive models are being used to assess the likelihood of disease progression to acute leukemia, supporting timely therapeutic interventions.[11]

4.3.1.5. AI in rare blood disorders.
For rare disorders such as paroxysmal nocturnal hemoglobinuria and hemophagocytic lymphohistiocytosis, AI offers new hope. By analyzing EHR data and laboratory results, AI models can identify patterns indicative of these conditions, leading to faster and more accurate diagnoses. These capabilities are especially valuable in settings with limited access to specialized expertise.[12]

4.3.2. Diagnostics in hematology using AI.
AI has revolutionized hematological diagnostics by enhancing accuracy, speed, and precision in detecting and classifying blood disorders. Through the integration of ML and DL techniques, AI-driven diagnostic tools are now capable of analyzing complex hematological data, offering clinicians invaluable insights and reducing the diagnostic burden.

4.3.2.1. Automated blood smear analysis.
Traditional blood smear examination requires significant expertise and time. AI-based systems, particularly CNNs, have been developed to automate this process. These systems can identify and classify different types of blood cells, including red blood cells, white blood cells, and platelets, as well as detect abnormalities such as anisocytosis, poikilocytosis, and inclusion bodies. For example, AI models can differentiate between subtypes of leukemia by identifying specific morphological changes in leukocytes with high accuracy.[12]

4.3.2.2. Hematological malignancy diagnostics.
AI is playing a pivotal role in diagnosing hematological cancers such as leukemia, lymphoma, and myeloma. By analyzing cytogenetic data, next-generation sequencing results, and flow cytometry outputs, ML algorithms can detect genetic mutations and chromosomal abnormalities indicative of these malignancies. Additionally, deep learning models have demonstrated the ability to classify leukemias into subtypes (e.g., acute myeloid leukemia [AML], ALL) based on bone marrow aspirates and peripheral blood smears, facilitating timely and accurate treatment decisions.[13]

4.3.2.3. Coagulation disorders.
The diagnosis of bleeding and clotting disorders, including hemophilia and thrombophilia, is greatly enhanced by AI. Advanced algorithms analyze coagulation profiles and laboratory test results (e.g., PT, aPTT, and D-dimer levels) to detect patterns associated with these conditions. For instance, AI-powered systems are being used to assess thrombotic risks in patients with antiphospholipid syndrome or disseminated intravascular coagulation, enabling early intervention.[14]

4.3.2.4. Molecular diagnostics and genomics.
AI excels in analyzing the vast and complex data generated by genomic and transcriptomic studies in hematology. Machine learning models can identify mutations, fusion genes, and expression patterns associated with blood disorders, such as mutations in the JAK2 or FLT3 genes linked to myeloproliferative neoplasms and AML, respectively. AI also facilitates the interpretation of multi-omics data, which integrates genomics, epigenomics, and proteomics, for a more comprehensive understanding of disease etiology and progression.[15]

4.3.2.5. Rare and inherited blood disorders.
AI has been instrumental in diagnosing rare and inherited blood disorders such as sickle cell anemia, thalassemia, and paroxysmal nocturnal hemoglobinuria. By analyzing patterns in hemoglobin electrophoresis data, flow cytometry results, and patient phenotypes, AI systems can provide rapid and accurate diagnoses. Moreover, predictive models can assess the likelihood of disease complications, such as vaso-occlusive crises in sickle cell anemia patients, allowing for preventive measures.[16]

4.3.2.6. Point-of-care diagnostics.
AI is also transforming point-of-care diagnostics in hematology by enabling portable devices to deliver reliable results in resource-limited settings. AI algorithms embedded in mobile apps or handheld devices analyze data from miniaturized sensors to diagnose anemia, malaria, or other hematological conditions with minimal infrastructure. This advancement is particularly beneficial for underserved regions, improving access to timely healthcare.[17]

4.4. Applications in disease understanding and diagnosis

The applications of big data analytics and ML in disease understanding and diagnosis within the field of hematology have been

transformative, providing comprehensive insights and enhancing diagnostic precision.[24] Big data analytics enable comprehensive genomic profiling, identifying genetic alterations, mutations, and aberrations associated with various hematological disorders.[25] ML algorithms aid in deciphering complex genomic data, revealing disease-specific molecular signatures and subclassifications.[26] ML-based algorithms assist in disease subclassification based on molecular profiles, enabling more accurate and refined categorization of hematological malignancies.[27] This approach helps in distinguishing subtypes with varying prognoses and treatment responses. ML-driven decision support systems analyze patient data, including clinical parameters and genomic profiles, to aid clinicians in making accurate and timely diagnoses.[28] These systems provide insights, flag potential abnormalities, and assist in differential diagnoses. Big data analytics combined with ML algorithms enhance the analysis of digital images from blood smears, bone marrow biopsies, and histopathological slides.[29] Automated image analysis improves diagnostic accuracy, identifies morphological abnormalities, and assists in disease classification. ML models identify predictive biomarkers or genetic signatures associated with disease progression or treatment responses.[30] These biomarkers aid in prognostication and treatment decision-making, guiding personalized therapeutic strategies. ML-based predictive models analyze patient data to identify individuals at higher risk of developing hematological disorders.[31] This early risk assessment facilitates proactive interventions and personalized screening programs for at-risk populations.

ML algorithms predict drug responses and identify potential drug-resistant patterns in hematological disorders based on genomic and clinical data.[32] This aids in selecting optimal therapies and mitigating treatment resistance. Big data analytics coupled with ML techniques enable precise monitoring of MRD levels post-treatment.[33] ML algorithms predict disease relapse risk by analyzing MRD data, facilitating early intervention and treatment modification if needed.[34] ML algorithms integrate diverse "omics" data (genomics, proteomics, transcriptomics) to generate comprehensive molecular profiles.[35] This integrated approach enhances disease understanding by uncovering complex disease mechanisms and interactions. Continuous monitoring through wearable devices coupled with real-time data analysis assists in tracking hematological parameters and disease progression.[36] ML models analyze streaming data, providing clinicians with dynamic patient information for real-time decision-making.[37] These applications highlight how big data analytics and ML techniques are revolutionizing disease understanding and diagnostic capabilities in hematology, fostering more precise, data-driven, and personalized approaches to patient care.

4.5. Predictive models for treatment response and prognostication

Predictive models powered by big data analytics and ML are instrumental in predicting treatment responses and prognoses in hematological disorders.[38] ML-based models analyze patient-specific data, including genomic profiles, clinical parameters, and treatment history, to forecast individual responses to specific therapies.[15] These models identify patterns correlating with treatment efficacy, aiding clinicians in selecting the most effective treatments for individual patients. ML algorithms predict drug responses and anticipate potential resistance mechanisms by integrating genomic data with treatment outcomes.[39] This allows for personalized treatment selection, avoiding ineffective therapies and anticipating the need for alternative strategies.

ML-driven prognostic models utilize patient data to stratify individuals into risk categories based on disease progression, survival outcomes, and recurrence probabilities.[40] These models incorporate diverse data types to predict patient-specific

prognoses, guiding treatment decisions and surveillance protocols. ML-based analyses of MRD data provide insights into disease persistence post-treatment.[41] Predictive models use MRD measurements to forecast the risk of disease relapse, facilitating timely interventions or treatment modifications. ML algorithms continuously monitor treatment responses by analyzing evolving patient data.[3] These models provide real-time feedback on treatment efficacy, allowing clinicians to adapt therapies based on changing disease dynamics. Predictive models guide the development of tailored therapeutic strategies by identifying patient-specific factors influencing treatment responses.[42] This facilitates personalized treatment plans based on individual characteristics and molecular profiles. ML-driven predictive models support adaptive treatment protocols by analyzing ongoing treatment responses.[43] These models enable clinicians to modify therapies, adjust dosages, or switch treatments based on real-time patient-specific data. ML algorithms integrate diverse datasets, including clinical records, genomic profiles, and treatment outcomes, to generate comprehensive predictive models.[44] This integrated approach improves the accuracy and reliability of treatment response predictions. Continuous validation and refinement of predictive models against new datasets and evolving patient profiles ensure their accuracy and generalizability.[45] Iterative improvements enhance the reliability of these models for clinical application. ML-based predictive models aid in designing more efficient and informative clinical trials.[44] These models help identify patient subgroups likely to respond to investigational therapies, expediting drug development and precision medicine approaches. Predictive models fueled by big data analytics and ML techniques are transforming treatment decision-making in hematological disorders, paving the way for more personalized, effective, and adaptive therapies tailored to individual patient profiles.[46] Continued research and validation are essential to refine these models and translate them into routine clinical practice.

5. ML models in hematology

ML models have shown significant promise in various fields of medicine, including hematology. They are used to assist in diagnostics, treatment prediction, patient risk stratification, and research advancements.[47] ML models can distinguish between different types of blood cells, aiding in the diagnosis of conditions like leukemia, anemia, and other blood disorders by analyzing microscopic images of blood smears. Algorithms can assist in diagnosing and classifying different types and stages of blood cancers, such as leukemia, lymphoma, and myeloma, by analyzing genetic data, cell morphology, and other clinical parameters. ML models can help predict disease progression, relapse, or patient outcomes by analyzing various patient data, such as genetic information, laboratory results, imaging, and treatment history. They aid in predicting a patient's response to specific therapies, optimizing treatment plans, and reducing adverse effects by analyzing historical treatment data and patient characteristics. ML is used in identifying potential drug candidates by analyzing molecular structures, interactions, and genomic data related to hematological disorders. ML models help in analyzing vast amounts of genomic data to identify genetic markers, mutations, and pathways associated with different blood disorders. ML algorithms streamline and automate laboratory processes, such as blood sample analysis, reducing human error and improving efficiency.[47] ML-powered systems aid healthcare professionals in making better-informed decisions by analyzing patient data and suggesting potential diagnoses or treatment options. ML algorithms aid in analyzing complex flow cytometry data to characterize and classify blood cells. ML assists in interpreting vast genomic sequencing data to identify mutations or variations associated with hematological diseases. Development of predictive models to anticipate

disease progression or response to specific treatments in conditions like anemia, hemophilia, or thrombosis. The integration of ML in hematology continues to evolve, showing great potential to improve diagnostics, treatment strategies, and patient care in blood-related disorders. However, continuous validation and refinement are crucial to ensure the accuracy and reliability of these models in clinical practice.

6. Big data techniques in hematology

Big data techniques in hematology involve the utilization of large volumes of diverse data, often including genomic, clinical, imaging, and experimental data, to derive insights, patterns, and correlations that can enhance our understanding of blood-related disorders. These techniques leverage advanced computational methods to analyze and interpret complex datasets in hematology.[48] High-throughput sequencing generates vast amounts of genetic data. Big data techniques help in analyzing this data to identify mutations, genetic markers, and pathways associated with hematological disorders. Identifying genetic variations and mutations in large datasets to understand disease mechanisms and personalize treatment approaches. Extracting and analyzing structured and unstructured data from EHRs to identify patterns, risk factors, and treatment outcomes related to blood disorders. Analyzing data from a large population to identify trends, risk factors, and disease prevalence within specific demographics.[48] Analyzing large collections of digital pathology images of blood smears or bone marrow samples to aid in disease diagnosis and classification. Applying big data analytics to radiological images (such as MRI, CT scans) to extract quantitative features and patterns for disease diagnosis and monitoring in hematology.

7. Machine learning integration

Predictive Analytics: Using machine learning models on large datasets to predict disease progression, treatment response, and patient outcomes based on various parameters.[49]

Clustering and Pattern Recognition: Identifying subtypes or clusters within hematological diseases to personalize treatment strategies or understand disease heterogeneity.[49]

Natural Language Processing (NLP): Analyzing text data from medical records, research papers, or literature to extract valuable insights and facilitate knowledge discovery in hematology.[49]

Big data techniques play a crucial role in advancing hematology research, diagnosis, treatment, and patient care. They enable the extraction of meaningful insights from vast and diverse datasets, leading to improved understanding, more precise diagnostics, and better-tailored treatments for blood-related disorders.[49]

8. The emerging AI in hematology

AI is revolutionizing hematology, offering innovative solutions for diagnostics, treatment, and research. AI technologies, including ML, DL, and NLP, are designed to process vast amounts of data and extract meaningful insights. These capabilities are particularly impactful in hematology, where diverse datasets – ranging from blood cell morphology to genomic data – demand sophisticated analytical tools. AI is poised to address challenges in disease detection, risk assessment, and personalized treatment, fostering a new era of precision medicine in hematology.[7] One of the most notable applications of AI in hematology is its role in disease diagnostics. AI-powered tools, such as CNNs, are increasingly employed for analyzing blood smears and bone marrow aspirates. These models can detect and classify hematological disorders like leukemia, lymphoma, and anemia with accuracy comparable to expert pathologists. Moreover, AI

algorithms have demonstrated the ability to identify subtle morphological changes in blood cells, enabling early diagnosis and monitoring of conditions like myelodysplastic syndromes and sickle cell disease.[8]

AI is also driving advancements in prognostic modeling and risk stratification. By analyzing patient data, including clinical histories, laboratory results, and genetic profiles, predictive models can forecast disease progression and treatment responses. For example, AI algorithms are used to identify high-risk patients with AML, guiding treatment decisions and improving outcomes. In addition, ML-based tools are enhancing the precision of bone marrow transplant matching, optimizing donor-recipient compatibility to reduce complications.[9] Another frontier for AI in hematology is drug discovery and therapeutic optimization. AI-driven platforms are being utilized to identify novel drug targets and predict therapeutic efficacy. Reinforcement learning algorithms are aiding in the design of personalized treatment regimens, ensuring optimal drug dosages and minimizing adverse effects. In diseases like hemophilia and beta-thalassemia, AI is facilitating the development of gene therapies by analyzing complex genomic datasets and predicting gene-editing outcomes.[10] Beyond diagnostics and therapy, AI is transforming hematology research by accelerating data integration and hypothesis generation. AI tools enable the integration of multi-omics data, including genomics, transcriptomics, and proteomics, to uncover novel biomarkers and pathways. NLP algorithms are mining scientific literature and EHRs to identify trends and correlations, fostering a deeper understanding of hematological disorders. These insights are catalyzing the development of targeted therapies and enhancing disease management strategies.[11]

Despite its promise, the adoption of AI in hematology faces several challenges. The "black-box" nature of many AI models raises concerns about interpretability and trustworthiness. Ensuring the quality and diversity of training datasets is critical to avoid biases that could compromise patient care. Moreover, the integration of AI into clinical workflows requires robust validation, regulatory approval, and clinician training. Addressing these challenges is essential to fully harness the transformative potential of AI in hematology.[12] Emerging trends in XAI and federated learning offer solutions to some of these hurdles. XAI focuses on making AI models more transparent, enabling clinicians to understand and trust the decision-making process. Federated learning, which allows decentralized analysis of data across institutions, ensures patient privacy while facilitating collaborative research. These innovations, coupled with advancements in computational power and data science, are paving the way for the seamless integration of AI into hematology.[13]

8.1. Challenges and limitations

In the realm of employing big data analytics and ML in hematological disorders, several challenges and limitations persist, hindering their seamless integration and optimal utilization.[50] Ensuring the quality, completeness, and harmonization of diverse datasets (genomic, clinical, imaging) poses a significant challenge. Inconsistent or fragmented data, data silos, and difficulties in integrating disparate data sources can affect the accuracy and reliability of predictive models. ML models often operate as "black boxes," lacking interpretability and explainability in their decision-making processes.[45] The inability to elucidate how models arrive at conclusions or treatment recommendations may limit clinician trust and acceptance.

The use of sensitive patient data for predictive analytics raises ethical concerns regarding patient privacy, consent, and data security.[51] Strict adherence to ethical guidelines and regulatory frameworks is crucial to protect patient confidentiality and ensure data security. Biases inherent in training data or underrepresentation of certain populations may lead to

biased predictions or models that do not generalize well to diverse patient populations.[52] Addressing biases in datasets and ensuring model fairness is essential for equitable healthcare. Integrating predictive analytics into existing clinical workflows and EHR systems poses challenges.[53] Resistance to change, lack of interoperability, and additional time demands on clinicians for model interpretation hinder seamless integration. Access to computational resources, expertise in AI/ML development, and data science skills may be limited in some healthcare settings.[54] The need for specialized knowledge in advanced analytics creates barriers to implementation. The evolving landscape of regulations and standards related to AI in healthcare presents challenges. Ensuring compliance with regulatory frameworks, navigating legal complexities, and addressing liability concerns are critical.

Validating predictive models across diverse patient cohorts and clinical settings is essential for their reliability.[55] Transparency in model development, validation, and performance metrics is necessary for gaining clinician and patient trust. Implementing and maintaining AI-driven models and big data analytics solutions may be cost-prohibitive for some healthcare institutions.[56] Assessing the cost-effectiveness and long-term sustainability of these technologies is vital. ML models require continuous updates and refinement to adapt to evolving data and patient characteristics.[45] The need for ongoing improvement and validation adds complexity to their implementation.

8.2. Integration into clinical practice

Integrating big data analytics and ML into clinical practice for managing hematological disorders involves various considerations and steps.[57] Providing clinicians with training and education on the use of data analytics and ML in hematology is essential. Continuing education programs, workshops, and specialized training sessions help in fostering familiarity and competence in utilizing these technologies. Collaboration among hematologists, data scientists, bioinformaticians, and technology experts is vital for successful integration.[58] Multidisciplinary teams enable the interpretation of complex data and the development of clinically relevant applications. Developing user-friendly interfaces and decision support tools that integrate seamlessly into existing EHR systems facilitates clinician engagement.[59] These tools aid in interpreting data and providing actionable insights at the point of care. Validating predictive models and algorithms in real-world clinical settings is crucial for gaining clinician confidence.[60] Robust evidence demonstrating the clinical utility and accuracy of these models is essential for widespread adoption. Educating patients about the role of data analytics and ML in their care fosters transparency and trust.[61] Ensuring patient understanding and obtaining informed consent regarding the use of predictive models and data-driven approaches is imperative.

Adhering to regulatory standards and ethical guidelines regarding patient data privacy, consent, and confidentiality is essential.[62] Compliance with HIPAA and other data protection laws is crucial for ethical implementation.[63] Involving clinicians in the development and validation of predictive models fosters acceptance and adoption. Gathering feedback from clinicians helps in refining models to better suit clinical needs and workflows. Initiating pilot programs within healthcare institutions allows for iterative testing, refinement, and optimization of data-driven tools.[64] Gradual implementation strategies mitigate disruption to clinical workflows. Ensuring ongoing maintenance, updates, and refinement of predictive models based on new data and feedback is vital. Establishing protocols for continuous improvement ensures the relevance and accuracy of these tools. Providing evidence of improved patient outcomes, enhanced diagnostic accuracy, or optimized treatment decisions through data analytics and

ML tools is crucial for demonstrating their clinical value and justifying their integration.[65]

8.3. Future directions and emerging innovations

Looking ahead, several future directions and emerging innovations are poised to shape the landscape of big data analytics and ML in the management of hematological disorders.[66] Advancements in integrating diverse "omics" data (genomics, transcriptomics, proteomics, metabolomics) will provide a more holistic understanding of hematological diseases.[13] Multi-omics approaches will uncover complex disease mechanisms and potential therapeutic targets. Further development of single-cell technologies will enable the characterization of cellular heterogeneity within hematological malignancies.[67] This granular insight will elucidate clonal evolution, identify rare cell populations, and inform personalized treatment strategies. Integrating genomics with immune profiling will advance immunotherapeutic strategies in hematological disorders.[68] Understanding the tumor microenvironment and immune responses will guide the development of more effective immunotherapies.

Focus on developing ML models that are interpretable and explainable will enhance clinician trust and adoption.[45] Explainable AI will provide transparent insights into model predictions and aid in understanding complex algorithms. Further integration of wearable devices and real-time monitoring tools will enable continuous data collection. ML-driven analytics of real-time patient data will facilitate dynamic treatment adjustments and remote patient monitoring. Continued advancements in ML algorithms for digital pathology and imaging analysis will improve diagnostic accuracy.[46] Automated analysis of blood smears, bone marrow biopsies, and histopathological images will aid in disease classification. Research into gene editing technologies, such as CRISPR/Cas9, will pave the way for precise and personalized gene therapies for hematological disorders.[69] Enhanced global collaborations and data-sharing initiatives will foster the creation of larger and more diverse datasets.[70] Shared datasets will enable more robust analyses and discoveries. Further development of ethical guidelines and regulatory frameworks will ensure responsible and ethical use of AI in hematology.[71] Emphasis on patient privacy, data security, and ethical AI practices will guide future implementations. Innovative trial designs using predictive analytics and AI-driven patient stratification will accelerate the development of targeted therapies.[72] Adaptive trials will optimize treatment strategies based on patient-specific characteristics.

The narrative review is summarized in Table 1 showing the Summary of Big Data Analytics and Machine Learning in Hematology: Transformative Insights, Applications, and Challenges (provided by the authors).

9. Conclusion

The integration of big data analytics and AI, particularly ML, has set the stage for a transformative era in hematology. These technologies offer unprecedented capabilities in disease diagnosis, prognostic modeling, and personalized treatment, addressing some of the most complex challenges in the field. From analyzing intricate blood cell morphologies to uncovering novel biomarkers through multi-omics integration, AI is significantly enhancing precision medicine. Furthermore, its applications in drug discovery and optimization of therapeutic regimens underscore its potential to improve patient outcomes and reshape clinical workflows.

However, the path to widespread adoption is not without obstacles. Technical challenges such as data heterogeneity, algorithm validation, and integration into clinical systems must be addressed. Ethical considerations, including bias in training data, patient privacy, and the transparency of AI models, are

Table 1

A summary of big data analytics and machine learning in hematology: transformative insights, applications, and challenges

Section	Key points
Introduction	- Overview of big data analytics and machine learning (ML) in hematology. - Their transformative potential for diagnostics, treatment, and research. - Challenges like data integration, quality, and ethical concerns.
Emerging AI in hematology	- AI technologies, such as machine learning and deep learning, are being integrated into hematological diagnostics and treatments. - AI enhances the accuracy and speed of detecting blood disorders like leukemia and anemia.
Applications of AI in blood disorders	- AI used for diagnosing hematological malignancies, such as leukemia, lymphoma, and multiple myeloma. - AI also assists in rare blood disorder diagnostics like sickle cell anemia and hemophilia. - Personalized treatment decisions are aided by AI-driven predictive models.
AI in hematology diagnostics	- AI-based automated blood smear analysis with CNNs improves the classification of blood cells. - Molecular diagnostics and genomics analyzed by AI assist in the detection of mutations and gene fusions associated with blood cancers.
AI in coagulation disorders	- AI helps diagnose bleeding and clotting disorders like hemophilia and thrombophilia. - Models predict thrombotic risks and optimize treatment plans for disorders such as DIC and antiphospholipid syndrome.
Challenges in AI implementation	- Data quality issues and model interpretability remain significant obstacles. - Ethical concerns regarding AI use, patient privacy, and data bias. - Integration of AI into clinical workflows requires standardization.
Future directions	- Advances in explainable AI (XAI), federated learning, and real-time analytics will improve model transparency and patient privacy. - Automated, AI-driven diagnostic workflows will enhance clinical efficiency and outcomes.
Conclusion	- AI is revolutionizing hematological diagnostics and treatments. - Continued research and overcoming challenges will maximize AI's potential to improve patient care and outcomes in hematology.

AI = artificial intelligence, CNNs = convolutional neural networks, DIC = disseminated intravascular coagulation.

equally critical to ensuring equitable and trustworthy implementations. Collaborative efforts among hematologists, data scientists, and regulatory bodies will be vital in overcoming these barriers and translating AI innovations into practice. As advancements in explainable AI, federated learning, and computational capabilities continue, the future of AI in hematology appears bright. These technologies promise to bridge gaps in research and clinical care, fostering a new era of precision, efficiency, and innovation. By embracing these advancements and addressing the associated challenges, hematology can harness the full potential of AI to improve patient care and advance the field to new horizons.

Author contributions

Conceptualization: Emmanuel Ifeanyi Obeagu.
Methodology: Emmanuel Ifeanyi Obeagu, Deborah Domini Ifu.
Supervision: Emmanuel Ifeanyi Obeagu.
Validation: Emmanuel Ifeanyi Obeagu.
Visualization: Emmanuel Ifeanyi Obeagu, Deborah Domini Ifu.
Writing – original draft: Emmanuel Ifeanyi Obeagu, Christiana Uchenna Ezeanya, Fabian Chukwudi Ogenyi, Deborah Domini Ifu.
Writing – review & editing: Emmanuel Ifeanyi Obeagu, Christiana Uchenna Ezeanya, Fabian Chukwudi Ogenyi, Deborah Domini Ifu.

References

[1] Groopman EE, Povysil G, Goldstein DB, Gharavi AG. Rare genetic causes of complex kidney and urological diseases. Nat Rev Nephrol. 2020;16:641–56.

[2] Banerjee A, Chakraborty C, Kumar A, Biswas D. Emerging trends in IoT and big data analytics for biomedical and health care technologies. In: Chakraborty C, ed. Handbook of Data Science Approaches for Biomedical Engineering. Elsevier; 2020:121–52.

[3] Padhi A, Agarwal A, Saxena SK, Katoch CDS. Transforming clinical virology with AI, machine learning and deep learning: a comprehensive review and outlook. Virusdisease. 2023;34:345–55.

[4] Akbulut U, Cifci MA, Aslan Z. Hybrid modeling for stream flow estimation: integrating machine learning and federated learning. Appl Sci. 2023;13:10203.

[5] Afrin LB. Mast cell activation disease and the modern epidemic of chronic inflammatory disease. Transl Res. 2016;174:33–59.

[6] Goldmann N, Skalicky SE, Weinreb RN, et al. Defining functional requirements for a patient-centric computerized glaucoma treatment and care ecosystem. J Med Artif Intell. 2023;6:22.

[7] Engert A, Balduini C, Brand A, et al.; EHA Roadmap for European Hematology Research. The European hematology association roadmap for European hematology research: a consensus document. Haematologica. 2016;101:115–208.

[8] Ghosh A, Parui S, Samanta D, Mukhopadhyay J, Chakravorty N. Computer Aided Diagnosis: Approaches to Automate Hematological Tests: Approaches to Automate Hematological Tests. In: Dutta et al, eds. Modern Techniques in Biosensors: Detection Methods and Commercial Aspects. Springer Nature Singapore Pte Ltd.; 2021:111–34.

[9] Rodríguez Ruiz N, Abd Own S, Ekström Smedby K, et al. Data-driven support to decision-making in molecular tumour boards for lymphoma: A design science approach. Front Oncol. 2022;12:984021.

[10] Naik N, Rallapalli Y, Krishna M, et al. Demystifying the advancements of big data analytics in medical diagnosis: an overview. Eng Sci. 2021;19:42–58.

[11] Badalian-Very G. Personalized medicine in hematology—a landmark from bench to bed. Comput Struct Biotechnol J. 2014;10:70–7.

[12] Duncavage EJ, Bagg A, Hasserjian RP, et al. Genomic profiling for clinical decision making in myeloid neoplasms and acute leukemia. Blood. 2022;140:2228–47.

[13] Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nat Rev Genet. 2018;19:299–310.

[14] Iqbal R, Doctor F, More B, Mahmud S, Yousuf U. Big data analytics: Computational intelligence techniques and application areas. Technol Forecast Soc Change. 2020;153:119253.

[15] Quazi S. Artificial intelligence and machine learning in precision and genomic medicine. Med Oncol. 2022;39:120.

[16] El Achi H, Khoury JD. Artificial intelligence and digital microscopy applications in diagnostic hematopathology. Cancers (Basel). 2020;12:797.

[17] Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. J Intern Med. 2020;288:62–81.

[18] Guk K, Han G, Lim J, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials (Basel). 2019;9:813.

[19] Dey P, Ross JS, Ritchie JD, Desai NR, Bhavnani SP, Krumholz HM. Data sharing and cardiology: platforms and possibilities. J Am Coll Cardiol. 2017;70:3018–25.

[20] Moore JH, Boland MR, Camara PG, et al. Preparing next-generation scientists for biomedical big data: artificial intelligence approaches. Per Med. 2019;16:247–57.

[21] Scheibner J, Ienca M, Kechagia S, et al. Data protection and ethics requirements for multisite research with health data: a comparative examination of legislative governance frameworks and the role of data protection technologies. J Law Biosci. 2020;7:lsaa010.

[22] Nedungadi P, Iyer A, Gutjahr G, Bhaskar J, Pillai AB. Data-driven methods for advancing precision oncology. Curr Pharmacol Rep. 2018;4:145–56.

[23] Ho D, Quake SR, McCabe ER, et al. Enabling technologies for personalized and precision medicine. Trends Biotechnol. 2020;38:497–518.

[24] Walter W, Haferlach C, Nadarajah N, et al. How artificial intelligence might disrupt diagnostics in hematology in the near future. Oncogene. 2021;40:4271–80.

[25] He J, Abdel-Wahab O, Nahas MK, et al. Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. Blood. 2016;127:3004–14.

[26] Glaab E, Rauschenberger A, Banzi R, Gerardi C, Garcia P, Demotes J. Biomarker discovery studies for patient stratification using machine learning analysis of omics data: a scoping review. BMJ Open. 2021;11:e053674.

[27] El Alaoui Y, Elomri A, Qaraqe M, et al. A review of artificial intelligence applications in hematology management: current practices and future prospects. J Med Internet Res. 2022;24:e36490.

[28] Yasmin F, Shah SMI, Naeem A, et al. Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. Rev Cardiovasc Med. 2021;22:1095–113.

[29] Brück OE, Lallukka-Brück SE, Hohtari HR, et al. Machine learning of bone marrow histopathology identifies genetic and clinical determinants in patients with MDS. Blood Cancer Discov. 2021;2:238–49.

[30] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nat Rev Clin Oncol. 2019;16:703–15.

[31] Mahmood N, Shahid S, Bakhshi T, Riaz S, Ghufran H, Yaqoob M. Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. Med Biol Eng Comput. 2020;58:2631–40.

[32] Wu Z, Lawrence PJ, Ma A, Zhu J, Xu D, Ma Q. Single-cell techniques and deep learning in predicting drug response. Trends Pharmacol Sci. 2020;41:1050–65.

[33] Riva G, Nasillo V, Ottomano AM, et al. Multiparametric flow cytometry for MRD monitoring in hematologic malignancies: clinical applications and new challenges. Cancers (Basel). 2021;13:4582.

[34] Pan L, Liu G, Lin F, et al. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. Sci Rep. 2017;7:7402.

[35] Misra BB, Langefeld C, Olivier M, Cox LA. Integrated omics: tools, advances and future approaches. J Mol Endocrinol. 2019;62:R21–45.

[36] Anikwe CV, Nweke HF, Ikegwu AC, et al. Mobile and wearable sensors for data-driven health monitoring system: state-of-the-art and future prospect. Expert Syst Appl. 2022;202:117362.

[37] Mohedano-Munoz MA, Soguero-Ruiz C, Mora-Jiménez I, Rubio-Sánchez M, Álvarez-Rodríguez J, Sanchez A. A streaming data visualization framework for supporting decision-making in the Intensive Care Unit. Expert Syst Appl. 2023;227:120252.

[38] Sakamoto T, Goto T, Fujiogi M, Kawarai Lefor A. Machine learning in gastrointestinal surgery. Surg Today. 2021;52:995–1007.

[39] Madhukar NS, Elemento O. Bioinformatics approaches to predict drug responses from genomic sequencing. Methods Mol Biol. 2018;1711:277–96.

[40] Iseke S, Zeevi T, Kucukkaya AS, et al. Machine learning models for prediction of posttreatment recurrence in early-stage hepatocellular carcinoma using pretreatment clinical and MRI features: a proof-of-concept study. AJR Am J Roentgenol. 2023;220:245–55.

[41] Allegra A, Tonacci A, Sciaccotta R, et al. Machine learning and deep learning applications in multiple myeloma diagnosis, prognosis, and treatment selection. Cancers (Basel). 2022;14:606.

[42] Liefaard MC, Lips EH, Wesseling J, et al. The way of the future: personalizing treatment plans through technology. Am Soc Clin Oncol Educ Book. 2021;41:12–23.

[43] Hu Y, Huerta J, Cordella N, Mishuris RG, Paschalidis IC. Personalized hypertension treatment recommendations by a data-driven model. BMC Med Inform Decis Mak. 2023;23:1–0.

[44] Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database. 2020;2020:baaa010.

[45] Shafieibavani E, Goudey B, Kiral I, et al. Predictive models for cochlear implant outcomes: Performance, generalizability, and the impact of cohort size. Trends Hear. 2021;25:23312165211066174.

[46] Seyhan AA, Carini C. Are innovation and new technologies in precision medicine paving a new era in patients centric care? J Transl Med. 2019;17:1–28.

[47] Shouval R, Fein JA, Savani B, Mohty M, Nagler A. Machine learning and artificial intelligence in haematology. Br J Haematol. 2021;192:239–50.

[48] Radakovich N, Cortese M, Nazha A. Acute myeloid leukemia and artificial intelligence, algorithms and new scores. Best Pract Res Clin Haematol. 2020;33:101192.

[49] Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. npj Precis Oncol. 2020;4:19.

[50] Amiri Z, Heidari A, Darbandi M, et al. The personal health applications of machine learning techniques in the internet of behaviors. Sustainability. 2023;15:12406.

[51] Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff (Millwood). 2014;33:1139–47.

[52] Rouzrokh P, Khosravi B, Faghani S, et al. Mitigating bias in radiology machine learning: 1. Data handling. Radiol Artif Intell. 2022;4:e210290.

[53] Huang C, Koppel R, McGreevey JD III, Craven CK, Schreiber R. Transitions from one electronic health record to another: challenges, pitfalls, and recommendations. Appl Clin Inform. 2020;11:742–54.

[54] Bernstam EV, Shireman PK, Meric-Bernstam F, et al. Artificial intelligence in clinical and translational science: successes, challenges and opportunities. Clin Transl Sci. 2022;15:309–21.

[55] Khalid S, Yang C, Blacketer C, et al. A standardized analytics pipeline for reliable and rapid development and validation of prediction models using observational health data. Comput Methods Programs Biomed. 2021;211:106394.

[56] Lawry T, Mutkoski S, Leong N. Realizing the potential for AI in precision health. SciTech Lawyer. 2018;15:23–7.

[57] Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019;20:e262–73.

[58] Vamathevan J, Apweiler R, Birney E. Biomolecular data resources: bioinformatics infrastructure for biomedical data science. Ann Rev Biomed Data Sci. 2019;2:199–222.

[59] Abdel-Rahman SM, Gill H, Carpenter SL, et al. Design and usability of an electronic health record – integrated, point-of-care, clinical decision support tool for modeling and simulation of antihemophilic factors. Appl Clin Inform. 2020;11:253–64.

[60] Lazzarini N, Filippoupolitis A, Manzione P, Eleftherohorinou H. A machine learning model on real world data for predicting progression to Acute Respiratory Distress Syndrome (ARDS) among COVID-19 patients. PLoS One. 2022;17:e0271227.

[61] Kelley M, James C, Alessi Kraft S, et al. Patient perspectives on the learning health system: the importance of trust and shared decision making. Am J Bioeth. 2015;15:4–17.

[62] Hudgins C, Rose S, Fifield PY, Arnault S. Navigating the legal and ethical foundations of informed consent and confidentiality in integrated primary care. Fam Syst Health. 2013;31:9–19.

[63] Mbonihankuye S, Nkunzimana A, Ndagijimana A. Healthcare data security technology: HIPAA compliance. Wirel Commun Mob Comput. 2019;2019:1–7.

[64] Furxhi I, Perucca M, Blosi M, et al. ASINA project: towards a methodological data-driven sustainable and safe-by-design approach for the development of nanomaterials. Front Bioeng Biotechnol. 2022;9:805096.

[65] Castaneda C, Nalley K, Mannion C, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinf. 2015;5:1–6.

[66] Kumar P, Chauhan S, Awasthi LK. Artificial intelligence in healthcare: review, ethics, trust challenges & future research directions. Eng Appl Artif Intell. 2023;120:105894.

[67] Shi M, Dong X, Huo L, Wei X, Wang F, Sun K. The potential roles and advantages of single cell sequencing in the diagnosis and treatment of hematological malignancies.Adv Exp Med Biol. 2018;1068:119–33.

[68] Yofe I, Dahan R, Amit I. Single-cell genomic approaches for developing the next generation of immunotherapies. Nat Med. 2020;26:171–7.

[69] Hong A. CRISPR in personalized medicine: Industry perspectives in gene editing. Semin Perinatol. 2018;42:501–7.

[70] Rehm HL, Page AJH, Smith L, et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell Genom. 2021;1:100029.

[71] Geis JR, Brady AP, Wu CC, et al. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. Radiology. 2019;293:436–40.

[72] Boniolo F, Dorigatti E, Ohnmacht AJ, Saur D, Schubert B, Menden MP. Artificial intelligence in early drug discovery enabling precision medicine. Expert Opin Drug Discovery. 2021;16:991–1007.