



Artificial Intelligence-Driven Multi-Omics Diagnostic Pipelines for Infectious, Neurodegenerative, and Metabolic Diseases: From Biomarker Discovery to Precision Medicine and Digital Twin Healthcare

Emmanuel Nkansah, Micheal Abimbola Oladosu, Moses Adondua Abah, Olaide Ayokunmi Oladosu

Abstract: The combination of AI and multi-omics has ushered in a new, revolutionary era in disease diagnosis and precision medicine. This review aims to summarize the current status of the artificial intelligence (AI)-based multi-omics diagnostic workflows employed in three disease paradigms: infectious, neurodegenerative, and metabolic diseases, and critically evaluate their translational potential from the discovery of biomarkers to digital twin healthcare systems. Machine Learning (ML) and Deep Learning (DL) algorithms, as well as Explainable Artificial Intelligence (XAI) algorithms, are explored for their application in linking genomics, transcriptomics, proteomics, metabolomics, and microbiomics datasets to enable high-dimensional molecular phenotyping. Harmonisation strategies for multi-omics data, AI-driven feature selection, molecular pathway elucidation using graph neural networks (GNNs) and transformer architectures, and the development of digital twin models for personalised, dynamic health simulation are among the key themes. We delve deeper into the regulatory, ethical, and equity issues arising from the deployment of AI-omics systems across heterogeneous clinical settings. The review ends with a strategy for integrating validated AI-omics pipelines into the next-generation precision therapeutics and global health infrastructure.

Keywords: Artificial Intelligence; Multi-Omics; Biomarker Discovery; Precision Medicine; Digital Twin; Infectious Diseases; Neurodegenerative Diseases; Metabolic Diseases; Machine Learning; Deep Learning.

Nomenclature:

T2DM: Type 2 Diabetes Mellitus
AD: Alzheimer's Disease

GNNs: Graph Neural Networks
DL: Deep Learning
ML: Machine Learning
AI: Artificial Intelligence
XAI: Explainable Artificial Intelligence
DT: Digital Twin
sCCA: Sparse Canonical Correlation Analysis
VAE: Variational Autoencoders
TB: Tuberculosis
CSF: Cerebrospinal Fluid
A β : Amyloid-Beta
PD: Parkinson's Disease
MetS: Metabolic Syndrome
NAFLD: Non-Alcoholic Fatty Liver Disease
CDSS: Clinical Decision Support Systems
HER: Electronic Health Record
GANs: Generative Adversarial Networks
NMF: Non-Negative Matrix Factorisation

I. INTRODUCTION

Today, the burden of non-communicable and infectious diseases is rapidly growing, and unprecedented strain is being put on diagnostic systems, which are largely reactive, disease-specific and data poor [1]. Standard diagnostic methods for single- or two- or three-analyte disease assessments are inherently limited in reflecting the biological complexity of multifactorial diseases like Alzheimer's disease (AD), type 2 diabetes mellitus (T2DM) and emerging infectious diseases. With the introduction of high-throughput 'omics' platforms that include genomics, transcriptomics, proteomics, metabolomics and microbiomics, it is now possible to conduct in-depth molecular phenotyping of patients across multiple regulatory levels [2].

The dimensionality, heterogeneity, and noise of multi-omics data, however, pose formidable analytical challenges for classical bioinformatics [3]. Here, artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) algorithms, has emerged as the analytical cornerstone capable of extracting biologically meaningful signals from these extensive molecular catalogues [4,5]. Ensemble methods, graph neural networks (GNNs), variational autoencoders (VAEs), transformer architectures, and federated learning

Manuscript received on 12 June 2026 | First Revised Manuscript received on 23 June 2026 | Second Revised Manuscript received on 28 June 2026 | Manuscript Accepted on 15 July 2026 | Manuscript published on 30 July 2026.

*Correspondence Author(s)

Emmanuel Nkansah, Department of Accounting, Economics and Finance, School of Business, La Sierra University, Riverside, United States. Email ID: emma.ampofol122@gmail.com, ORCID ID: [0009-0006-6512-644X](https://orcid.org/0009-0006-6512-644X)

Micheal Abimbola Oladosu*, Department of Chemical Sciences, Faculty of Sciences, Anchor University, Ayobo, Ipaja, Lagos, Nigeria. E-mail ID: mikeoladosu@gmail.com, moladosu@aul.edu.ng, ORCID ID: [0009-0000-5098-1247](https://orcid.org/0009-0000-5098-1247)

Moses Adondua Abah, Department of Biochemistry, Faculty of Pure and Applied Sciences, Federal University of Wukari, Wukari, Taraba State, Nigeria. Email ID: m.abah@fuwukari.edu.ng, ORCID ID: [0000-0002-9268-1661](https://orcid.org/0000-0002-9268-1661)

Olaide Ayokunmi Oladosu, Department of Computer Science, Faculty of Science and Technology, Babcock University, Ilishan-Remo, Nigeria Email ID: pinkrosesng@gmail.com, ORCID ID: [0009-0001-8550-5598](https://orcid.org/0009-0001-8550-5598)

© The Authors. Published by Lattice Science Publication (LSP). This is an open-access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>



paradigms for privacy-preserving multi-site data integration are key innovations [6].

One of the most intriguing applications of AI-omics pipelines is their combination with digital twin (DT) technology, in which virtual patient models dynamically simulate real-time physiological changes, allowing prospective simulation of disease states, aiding treatment personalisation, and predicting outcomes [7]. This review evaluates the underlying mechanisms, clinical applications, and translation of AI-based multi-omics diagnostic pipelines across three disease categories: infectious, neurodegenerative, and metabolic diseases. It discusses the vision of embedding these into precision medicine and digital health systems.

II. MULTI-OMICS DATA INTEGRATION: ARCHITECTURES AND AI METHODOLOGIES

A. The Multi-Omics Landscape

Multi-omics integration is the simultaneous analysis of data from two or more molecular levels, providing an overall picture of disease biology [8]. Germline and somatic variation are identified by genomics; dynamic gene expression is detected by transcriptomics; the functional repertoire of proteins is characterised by proteomics; the biochemical phenotypes are detected by metabolomics; and microbiomics detect interactions between the host and microbiota. The time resolution and mechanisms of each layer differ, and their combination provides synergistic predictive power that cannot be achieved by any one of them alone [9].

Early methods ranged from simple concatenation, which joined feature matrices to create a single large feature matrix, to more advanced multi-view learning architectures. Multi-view models, kernel models, and matrix factorisation techniques, such as non-negative matrix factorisation (NMF) and joint NMF, are used to identify shared latent molecular programmes across omics layers [8]. To explicitly model cross-omics correlations, sparse canonical correlation analysis (sCCA) and its multi-view extensions are introduced, revealing co-regulatory modules that elucidate disease mechanisms. Integrative strategies that combine transcriptomic, proteomic, and metabolomic data layers with machine learning have been proposed to standardise such cross-omics analytical workflows [10].

The importance of integrating omics data. The need for integrating omics data

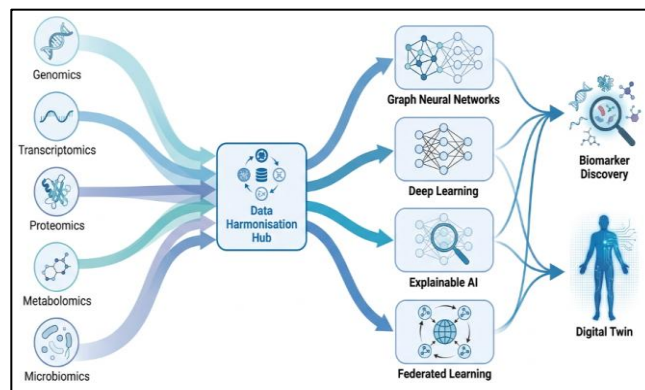
B. Deep Learning Architectures for Omics Integration.

Over the last five years, there has been a surge in deep learning architectures designed to integrate multiple omics [3]. Graph neural networks are designed to represent molecules (such as genes, proteins, metabolites) as nodes and their functional connections as links, enabling the propagation of molecular information across known interaction networks [11]. This topology-oriented learning is quite effective for determining pathway-level dysregulation in complex diseases. The self-attention mechanisms implemented in transformers, first introduced in natural

language processing, are now used to model long-range relationships both within and between omics layers, as well as across biological sequences. Large-scale foundation models built on these transformer architectures are increasingly being adapted for generalist biomedical and multi-omics applications [12].

Variational autoencoders (VAE) are designed to learn compressed representations of high-dimensional omics data, enabling cross-modal data generation, denoising and imputation [4]. Multi-modal fusion architectures fall into three major categories: early, intermediate, and late fusion paradigms, each offering complementary advantages and disadvantages in model flexibility and interpretability. To address data governance challenges, federated learning frameworks have been proposed to train models across different hospital nodes without sharing sensitive patient information while leveraging the diversity of multi-institutional data [6].

III. AI-OMICS PIPELINES IN INFECTIOUS DISEASE DIAGNOSTICS



[Fig.1: Schematic Overview of AI-Driven Multi-Omics Diagnostic Pipeline. The Figure Illustrates Data Acquisition Across Five Omics Layers (Genomics, Transcriptomics, Proteomics, Metabolomics, Microbiomics), Followed by Preprocessing and Harmonisation, AI-Based Feature Extraction and Integration, Biomarker Discovery, and Clinical Translation to Precision Medicine and Digital Twin Platforms. Colour-Coded Arrows Indicate Data Flow and Feedback Loops Between Pipeline Stages]

A. COVID-19 and Respiratory Infections

Multi-omics and AI-based methods for pathogen characterisation and host-response profiling were adopted in unprecedented numbers during the COVID-19 pandemic [13]. A novel integrated transcriptomics and proteomics approach for SARS-CoV-2 patients identified distinct immune endotypes, enabling stratification by the trajectory of their clinical course and guiding adaptive clinical trial designs. Using multi-omics machine learning models with plasma proteomics, metabolomics, and transcriptomics data, high sensitivity and specificity were reported in classifying



mild, moderate, and severe COVID-19 phenotypes [14].

Zhou et al. used single-cell RNA sequencing (scRNA-seq), bulk RNA sequencing (bulk RNA-seq), and machine learning for the analysis of the datasets to find BTD, CFL1, PIGR, and SERPINA3 as key biomarkers related to CD8⁺ T cell dysfunction in COVID-19 patients, candidates for diagnosis and therapeutic intervention [15]. Genomic surveillance with AI-powered variant classification helped promptly identify emerging SARS-CoV-2 variants, enabling vaccine updates within a few weeks rather than months [13].

B. Tuberculosis and Emerging Pathogens

The multi-omics approaches are also revolutionary for tuberculosis (TB) and other bacterial diseases. Culture-independent drug-resistance profiling of *Mycobacterium tuberculosis* using metagenomics sequencing and ML classifiers can be achieved within hours, drastically reducing

the time for diagnosis from weeks to the same day [13]. Gene expression biosignatures have been identified using an AI-powered host transcriptomic signature, thereby replacing traditional symptom-based screening in high-burden settings. This dual plasma metabolomic-proteomic fingerprinting enables high-performance infection classification with clinically relevant accuracy, differentiating TB from other LRTIs [2].

Multi-omics ML models for *C. difficile* infection have identified candidate metabolic biomarkers which predict recurrence and provide targets for intervention [2]. Such advancements highlight a new opportunity to combine host-response omics with pathogen genomics to create next-generation infectious disease diagnostics systems that can surpass the performance of the traditional culture and immunoassay platforms.

IV. AI-OMICS PIPELINES IN NEURODEGENERATIVE DISEASE DIAGNOSTICS

Table I: Summary of AI-Omics Approaches in Infectious, Neurodegenerative, and Metabolic Disease Diagnostics

Disease Category	Omics Layer(s)	AI/ML Algorithm	Key Biomarkers Identified	Reference
Infectious – COVID-19	Proteomics, Transcriptomics, Metabolomics	Ensemble ML, scRNA-seq	BTB, CFL1, PIGR, SERPINA3	[14,15]
Infectious – Tuberculosis	Metagenomics, Transcriptomics	Deep Learning, Random Forest	Host transcriptomic biosignatures	[13]
Neurodegenerative – Alzheimer's	Proteomics, Genomics, Metabolomics	Multi-task DL, GNN, Network Analysis	GFAP, NPTXR, MEF2C	[16,17,18]
Neurodegenerative – Parkinson's	Proteomics, Transcriptomics	Multi-task DL, UMAP clustering	Lysosomal biology proteins, α -synuclein pathway	[16]
Metabolic – Type 2 Diabetes	Metabolomics, Genomics, Proteomics	XGBoost, LightGBM, Random Forest	Pyruvate, AMP, Tetradecenoic acid	[19,20]
Metabolic – Obesity / MetS	Metabolomics, Microbiomics	XGBoost, AdaBoost	Bile acid metabolites, SCFA profiles	[21]

Abbreviations: DL = Deep Learning; GNN = Graph Neural Network; ML = Machine Learning; SCFA = Short-chain fatty acids; MetS = Metabolic Syndrome

A. Alzheimer's Disease

One of the most difficult diagnostic problems in medicine is Alzheimer's disease because irreversible neuronal loss is happening before the onset of clinical symptoms in a prolonged prodromal phase [16]. Multi-omics approaches using artificial intelligence (AI) technology have dramatically enriched the range of biomarkers from the traditional cerebrospinal fluid (CSF) amyloid-beta ($A\beta$) and tau markers. Ren et al. [16] conducted a comprehensive review of the potential applications of machine learning and deep learning, as well as network-based models that integrate large-scale genomic, transcriptomic, proteomic, metabolomic, and microbiomic datasets to better explain the pathogenesis of AD.

From the NCBI GEO archive, transcriptomic data for AD have been analysed using multi-omics integration and ML approaches to pinpoint MEF2C as a master regulatory node regulating synaptic maintenance under inflammatory stress, which could make it a druggable target in AD [17]. The precision-recall AUC of multi-task deep learning for AD classification was 0.76, and SHAP-based feature importance highlighted GFAP and NPTXR as among the most important diagnostic proteins [18]. These results converge on the focal

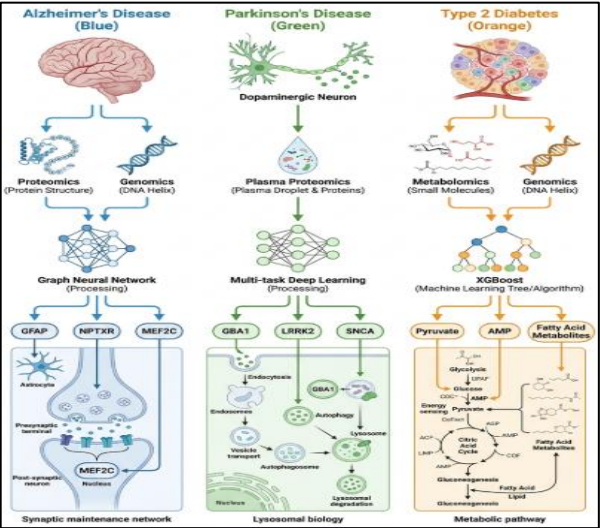
role of neuroinflammatory and synaptic maintenance pathways in the pathobiology of AD.

The impact of Parkinson's disease and other neurodegenerative disorders

Parkinson's disease (PD) also has shared proteomics signatures with AD. Still, it is characterised by dysregulation of lysosomal biology, as revealed by multi-task deep-learning patient clustering of plasma proteomics [18]. The use of integrative multi-omics and systems bioinformatics approaches is beginning to uncover the molecular mechanisms driving clinical heterogeneity in PD and to provide insights that can inform patient stratification relevant to trial design [16]. A combination of multi-omics and bioinformatics has been identified as one of the main research frontiers over the last 3 years of AI-neuroscience research (2022-2024), consistent with the bibliometric analysis of papers [22].

Precision medicine approaches for neurodegenerative diseases are becoming more comprehensive, with panels of CSF and blood markers, digital cognitive assessments, and neuroimaging increasingly combined in a single ML model to yield individualised risk trajectories and proposed times for preventive intervention [23].

V. AI-OMICS PIPELINES IN METABOLIC DISEASE DIAGNOSTICS



[Fig.2. Comparative Overview of Multi-Omics Biomarker Discovery Workflows for Alzheimer's Disease, Parkinson's Disease, and Type 2 Diabetes Mellitus. The Figure Displays Disease-Specific Omics Input Layers (left), AI Integration Methods (Centre), and key Biomarkers Identified and Mapped to Molecular Pathways (Right), with Colour-Coded Disease Categories (blue = Alzheimer's, Green = Parkinson's, Orange = Type 2 Diabetes)]

Table II: Comparative Analysis of AI/ML Algorithms Used in Multi-Omics Disease Diagnostic Pipelines (2020–2025)

AI/ML Algorithm	Algorithm Class	Omics Compatibility	Strengths	Limitations
XGBoost / LightGBM	Ensemble (Boosting)	Metabolomics, Genomics	High accuracy; handles missing data; feature importance via SHAP	Limited cross-omics integration capacity
Graph Neural Network (GNN)	Deep Learning	All omics (network-based)	Captures molecular interactions; topology-aware; pathway-level insight	Requires high-quality interaction networks; computationally intensive
Variational Autoencoder (VAE)	Deep Learning (Generative)	Genomics, Proteomics, Multi-omics	Dimensionality reduction; data imputation; cross-modal generation	Latent space interpretability challenges
Transformer / Attention	Deep Learning	Genomics, Transcriptomics	Long-range dependencies; pre-training potential; scalable	Data-hungry; high computational cost
Random Forest	Ensemble	Metabolomics, Proteomics, Clinical	Robust; interpretable feature importance; low overfitting risk	Less effective for ultra-high-dimensional data without feature selection
Federated Learning	Privacy-Preserving ML	All omics (distributed)	Privacy preservation; multi-site data utilisation without centralisation	Communication overhead; data heterogeneity across sites

Abbreviations: SHAP = SHapley Additive exPlanations; GNN = Graph Neural Network; VAE = Variational Autoencoder

AI-omics techniques have also proven valuable in characterising the molecular architecture of metabolic syndrome (MetS), non-alcoholic fatty liver disease (NAFLD) and obesity, beyond T2DM. There is growing evidence for an association between gut microbial community patterns and metabolites that discriminate MetS subtypes and therapeutic responses to dietary or pharmacological interventions, through a microbiomics-metabolomics integration using ML classifiers [9]. The results provide new opportunities for precision therapies targeting the microbiome using a patient-specific multi-omics profile.

VI. FROM BIOMARKER DISCOVERY TO PRECISION MEDICINE

A. AI-Informed Drug Delivery Optimization Strategies

A highly stringent multi-stage validation process is needed to translate computationally identified biomarkers into

A. Type 2 Diabetes Mellitus

Metabolomics using AI is a paradigm application for type 2 diabetes mellitus (T2DM), where the disease has a profound mechanistic basis in the human metabolome [19]. However, machine learning methods for metabolomic data were recently shown to be much better at predicting the incidence of T2DM than traditional risk scores and to uncover a novel set of metabolite predictors in addition to the classical glucose and HbA1c markers [19]. Metabolomics-based machine learning models using tree-based algorithms such as XGBoost, LightGBM and AdaBoost have been used to identify pyruvate, D-Rhamnose, AMP, pipercolate, tetradecenoic acid, and tetradecanoic acid as potential diabetic diagnostic biomarkers in multiple independent cohorts [20].

AI-driven network analysis integrates metabolomics, genomics, and proteomics in T2DM, linking metabolite-gene-protein regulatory pathways and identifying converging mechanistic themes, including dysregulated energy metabolism, oxidative stress pathways, and branched-chain amino acid catabolism [21]. Explainable AI (XAI), such as SHAP value-based feature attribution, when combined with metabolomics, can help identify the most diagnostically valuable metabolites, thereby building clinicians' confidence and supporting regulatory validation [20].

B. Obesity, Non-Alcoholic Fatty Liver Disease and Metabolic Syndrome

clinically validated diagnostic tools [24]. National-scale implementations of integrated multi-omics sequencing have already demonstrated accelerated diagnostic yield for rare disease cohorts, illustrating the translational pathway from biomarker discovery to routine clinical diagnostics [25]. Whilst solving the 'dimensionality curse' of high-dimensional omics data, regularisation methods, cross-validation, and independent replication across different cohorts are needed to prevent overfitting and spurious correlations [4]. Mechanistic interpretability methods enable the translation of the outputs of ML models into biologically meaningful insights, a crucial step in clinical adoption, with explainable AI approaches such as SHAP, LIME, and attention-based attribution methods.

Multi-omics biomarker panels validated in ancestry-diversified cohorts are increasingly becoming integrated into clinical



decision support systems (CDSS). These systems utilise probabilistic reasoning to generate patient-level risk scores by integrating multi-omics data, electronic health record (EHR) phenotypes, imaging data, and longitudinal vital sign trajectories, thereby supporting clinical decision-making [7]. Incorporating pharmacogenomic information alongside transcriptomic and metabolomic data further supports precision prescribing by selecting the right drug, dose, and timing based on the individual's molecular pharmacokinetics and pharmacodynamics.

B. Drug Repurposing through Knowledge Graphs and AI

Drug repurposing is a cost-effective approach to accelerate therapeutic development, especially for rare and neglected diseases. It has become a forefront of AI-driven drug discovery research [26]. Knowledge graph completion approaches using GNNs can model complex drug-gene-disease-pathway relationships and predict new drugs and their potential effectiveness in new indications [27]. Jiménez et al. showed that path-based knowledge graph completion methods can provide drug repurposing predictions with a biological pathway-level justification, which is important for regulatory acceptance [27]. These strategies have been used in other disease settings, including infectious, neurodegenerative, and metabolic diseases, and have led to the identification of candidates that have since entered clinical evaluation [28].

everyday clinical care, although widespread implementation still requires further validation [30].

Mechanistic and data-driven hybrid models are used to integrate multi-omics molecular profiles, wearable sensor data, EHR phenotypes, medical imaging and environmental exposures in the architecture of health DTs [31]. Papachou et al. reviewed the growing clinical use of DT in cardiology, oncology and neurology. They found that its performance in reducing cardiac arrhythmia recurrence by over 13% and in predicting neurodegenerative diseases with up to 97% accuracy was impressive in high-fidelity DT implementations [31]. These early findings provide the basis for the power of DTs as the computational infrastructure for next-generation precision medicine.

B. Digital Twins for Neurodegenerative and Metabolic Disease Management

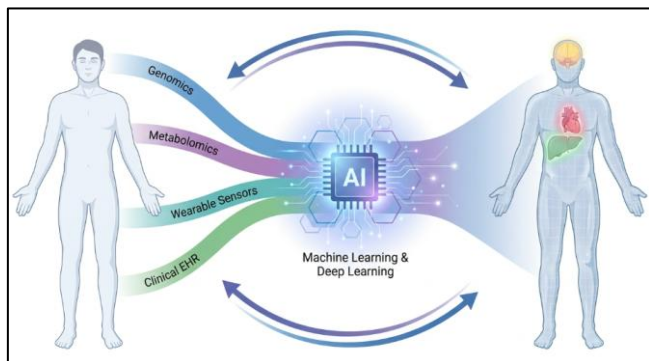
DT frameworks have been suggested in the context of dementia care to dynamically monitor cognitive decline trajectories, tailor NPIs to individuals, and optimise the timing of therapeutic escalation [32]. Wickramasinghe et al. proposed a DT architecture for dementia management that uses clinical, genomic, and sensor data to generate individualised care pathway simulations, enabling care teams to assess the effects of various dementia management approaches on patients' functional outcomes [32]. DT models with generalised metabolic flux modelling can also be applied to metabolic disease management to detect and forecast the risk of chronic kidney disease in T2DM cohorts [33].

The use of generative AI, such as large language models (LLMs) and generative adversarial networks (GANs), is helping to overcome another major limitation in rare disease scenarios: the lack of available clinical data. One of the key challenges in rare disease scenarios is the limited amount of clinical information available, and generative AI, including large language models (LLMs) and generative adversarial networks (GANs), is helping to overcome this hurdle. In drug discovery and clinical trial design, generative AI-driven DTs are being employed to model patient-cohort responses to candidate therapies, thereby minimising early attrition and expediting regulatory processes [34]. Table III presents the Emerging Digital Twin Applications in Precision Medicine Across Disease Domains (2020–2025), while Figure 3 shows the Digital Twin Healthcare Framework integrating AI-driven multi-omics pipelines.

Table III: Emerging Digital Twin Applications in Precision Medicine Across Disease Domains (2020–2025)

Disease Domain	DT Data Inputs	AI Model Integrated	Clinical Outcome Reported	Reference
Cardiovascular Disease	EHR, Imaging, Omics, Wearables	Mechanistic + ML Hybrid	>13% reduction in arrhythmia recurrence	[31]
Neurodegenerative (Dementia)	Clinical, Genomic, Sensor, Cognitive scores	DT Simulation, LSTM, GNN	97% accuracy in disease prediction	[31,32]
Type 2 Diabetes / CKD	Metabolic flux, EHR, Proteomics	Generalised Metabolic Flux Model + ML	Early CKD risk identification in T2DM	[33]
Infectious Disease	Metagenomics, Host transcriptomics, Clinical	Federated ML, Ensemble DL	Variant detection; vaccine response modelling	[13]
Oncology	Multi-omics, Imaging, Immunomics	Generative AI, GAN, Transformer	Sub-millisecond liver response prediction	[34]

Abbreviations: DT = Digital Twin; EHR = Electronic Health Record; LSTM = Long Short-Term Memory; CKD = chronic kidney disease; GAN = Generative Adversarial Network.



[Fig.3: Digital twin Healthcare Framework Integrating AI-Driven Multi-Omics Pipelines. The Figure Depicts the Bidirectional Data Flow Between a Real Patient (Left) and Their Digital Twin (Right), Mediated by a Central AI Processing Core. Inputs Include Multi-Omics Molecular Profiles, Wearable Sensor Streams, EHR Data, and Medical Imaging. The Digital Twin Simulates Disease Progression, Therapeutic Response, and Adverse Event Risk, Feeding Individualised Recommendations Back to the Clinical Team]

VIII. CHALLENGES AND ETHICAL CONSIDERATIONS

AI-based multi-omics diagnostic pipelines have a significant translational hurdle, yet have transformative potential. The most significant technical challenge to the generalisation of models is data heterogeneity due to variations in sample collection, processing methods, sequencing technologies, and patient ancestry [9]. To overcome the limitations of batch effects and confounding biological variation, robust harmonisation methods, such as ComBat, quantile normalisation, and multi-site reference standardisation, are necessary [3].

Complex DL models have been referred to as ‘black-box’ due to their challenges in regulatory approval and clinician trust [26]. SHAP, LIME, attention visualisation and counterfactual explanations are key XAI methods that enable interpretable model decisions, but have yet to be explored in ultra-high-dimensional omics settings [26]. Regulatory bodies, such as the FDA and EMA, are now adapting their regulations to include AI-powered diagnostic devices. However, there are still no established guidelines for multi-omics-specific validation, prospective monitoring of device performance, or algorithmic auditing [35].

Equity and access are important ethical components. Underrepresented populations, particularly in low-income areas, are underrepresented in training datasets and are vulnerable to further deepening health inequities through AI systems that perform differently across populations [23]. While there are mitigation measures such as federated learning, transfer learning, and synthetic data generation with generative AI, there is a need for proactive policies to ensure inclusive data collection and equitable deployment.

IX. FUTURE DIRECTIONS AND ROADMAP

Implementing the full potential of AI-omics diagnostics and digital twin healthcare involves multiple interrelated priorities on the roadmap. First, training generalisable AI models will require the development of international,

federated, multi-omics reference datasets spanning diverse ancestries, disease stages, and geographies [6]. Second, the standardisation of data reporting formats, metadata schemas, and FAIR (Findable, Accessible, Interoperable, Reusable [8]) omics data will support model replications and cross-study meta-analyses.

Thirdly, the combination of spatial omics technologies, such as spatial transcriptomics and spatial proteomics, with AI will provide important tissue-level contextualisation of molecular profiles, especially in the pathology of neurodegenerative and metabolic diseases [9]. Fourth, there is a need for investments in healthcare IT infrastructure, integration with clinical workflows, training programmes for clinicians, and development of AI health literacy for the clinical use of DTs based on AI-omics [7]. Fifth, regulatory science should be adapted to provide adaptive frameworks for AI-omics diagnostics that are patient-safe and enable rapid innovation [35].

AI, multi-omics, and digital twin technologies, together with wearable biosensor networks, real-world evidence platforms, and patient-generated data ecosystems, will increasingly support a model of continuous, proactive, and personalised health management, moving medicine from episodic, reactive to longitudinal, predictive wellness.

X. CONCLUSION

Multi-omics diagnostic pipelines powered by artificial intelligence are transforming the landscape of disease detection, mechanisms, and therapeutic personalisation in infectious, neurodegenerative, and metabolic disease settings. The computational/analytical tools to extract clinically relevant and biologically meaningful information from the molecular complexity of human diseases will be rooted in machine-learning and deep-learning architectures, such as graph neural networks, variational autoencoders, and transformer models. Combining validated multi-omics biomarker panels with digital-twin healthcare frameworks is a convergent approach to dynamic and individualised precision medicine, predicting disease rather than reacting to it.

Fulfilling this vision requires multi-disciplinary collaboration in discovery science, clinical validation, regulatory science, health informatics, and health equity, with an emphasis on open, diverse, and FAIR data ecosystems. The evidence presented here demonstrates the components of this vision that are in place; the imperative now is the systematic movement from the research world to global clinical practice.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This any organizations or agencies have not funded article. This independence ensures that the research is conducted objectively





and without external influence.

- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396(10258):1204–1222. DOI: [http://doi.org/10.1016/S0140-6736\(20\)30925-9](http://doi.org/10.1016/S0140-6736(20)30925-9)
2. Li Y, Wu X, Fang D, Luo Y. Improving immunotherapy with multi-omics driven machine learning. *NPJ Digit Med*. 2024;7(1):67. DOI: <http://doi.org/10.1038/s41746-024-01043-6>
3. Ballard JL, Wang Z, Li W, Shen L, Long Q. Deep learning-based approaches for multi-omics data integration and analysis. *BioData Min*. 2024;17(1):38. DOI: <http://doi.org/10.1186/s13040-024-00391-z>
4. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med*. 2021;13(1):152. DOI: <http://doi.org/10.1186/s13073-021-00968-x>
5. Yetgin A. Revolutionising multi-omics analysis with artificial intelligence and data processing. *Quant Biol*. 2025;13(3):e70002. DOI: <http://doi.org/10.1002/qub.70002>
6. Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med*. 2020;3(1):119. DOI: <http://doi.org/10.1038/s41746-020-00323-1>
7. Katsoulakis E, Wang Q, Wu H, et al. Digital twins for health: a scoping review. *NPJ Digit Med*. 2024;7(1):77. DOI: <http://doi.org/10.1038/s41746-024-01073-0>
8. Athieniti E, Spyrou GM. A guide to multi-omics data collection and integration for translational medicine. *Comput Struct Biotechnol J*. 2023;21:134–149. DOI: <http://doi.org/10.1016/j.csbj.2022.11.050>
9. Chen C, Wang J, Pan D, et al. Applications of multi-omics analysis in human diseases. *MedComm*. 2023;4(4):e315. DOI: <http://doi.org/10.1002/mco.2.315>
10. Sanches PHG, de Melo NC, Porcari AM, de Carvalho LM. Integrating molecular perspectives: strategies for comprehensive multi-omics integrative data analysis and machine learning applications in transcriptomics, proteomics, and metabolomics. *Biology (Basel)*. 2024;13(11):848. DOI: <http://doi.org/10.3390/biology13110848>
11. Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. *Cell*. 2020;180(4):688–702. DOI: <http://doi.org/10.1016/j.cell.2020.01.021>
12. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259–265. DOI: <http://doi.org/10.1038/s41586-023-05881-4>
13. Molla G, Bitew M. Revolutionising personalised medicine: synergy with multi-omics data generation, main hurdles, and future perspectives. *Biomedicine*. 2024;12(12):2750. DOI: <http://doi.org/10.3390/biomedicine12122750>
14. Zhou Y, Fan P, Zhang H, et al. multi-omics and machine learning identify novel biomarkers and therapeutic targets of COVID-19. *Front Immunol*. 2025;16:1671936. DOI: <http://doi.org/10.3389/fimmu.2025.1671936>
15. Cao S, Dhaliwal HK, Pan M, et al. Extracting a COVID-19 signature from a multi-omic dataset. *Gigascience*. 2025;14: gfa064. DOI: <http://doi.org/10.1093/gigascience/gfa064>
16. Ren F, Wei J, Chen Q, et al. Artificial intelligence-driven multi-omics approaches in Alzheimer's disease: progress, challenges, and future directions. *Acta Pharm Sin B*. 2025;15(8):3985–4009. DOI: <http://doi.org/10.1016/j.apsb.2025.07.030>
17. Eteleeb AM, Novotny BC, Soriano Tarraga C, et al. Brain high-throughput multi-omics data reveal molecular heterogeneity in Alzheimer's disease. *PLoS Biol*. 2024;22(4):e3002607. DOI: <http://doi.org/10.1371/journal.pbio.3002607>
18. Christensen NJ, Lassen FH, Holm LA, et al. Identification of cross-diagnostic protein signatures and subtypes of neurodegenerative diseases using multi-task deep learning on the UK Biobank proteomics dataset. *Alzheimer's Dement*. 2025;21(1):e14381. DOI: <http://doi.org/10.1002/alz.2014381>
19. Leihner A, Muendlein A, Mink S, et al. Machine-learning approach to metabolomic data predicts incidence of type 2 diabetes mellitus. *Int J Mol Sci*. 2024;25(10):5331. DOI: <http://doi.org/10.3390/ijms25105331>
20. Yagin FH, Al-Hashem F, Ahmad I, Ahmad F, Alkhateeb A. Pilot-study to explore metabolic signature of type 2 diabetes: a pipeline of tree-based machine learning and bioinformatics techniques for biomarkers discovery. *Nutrients*. 2024;16(10):1537. DOI: <http://doi.org/10.3390/nu16101537>
21. Arslan AK, Yagin FH, Algarni A, Karaaslan E, Al-Hashem F, Ardigò LP. Enhancing type 2 diabetes mellitus prediction by integrating metabolomics and tree-based boosting approaches. *Front Endocrinol*. 2024; 15:1444282. DOI: <http://doi.org/10.3389/fendo.2024.1444282>
22. O'Connor LM, O'Connor BA, Lim SB, Zeng J, Lo CH. Integrative multi-omics and systems bioinformatics in translational neuroscience: a data mining perspective. *J Proteome Res*. 2023;22(3):660–684. DOI: <http://doi.org/10.1021/acs.jproteome.2c00580>
23. Rossi SL, Subramanian P, Bovenkamp DE. The future is precision medicine-guided diagnoses, preventions and treatments for neurodegenerative diseases. *Front Aging Neurosci*. 2023; 15:1128619. DOI: <http://doi.org/10.3389/fnagi.2023.1128619>
24. Ahmad A, Imran M, Ahsan H. Biomarkers as biomedical bioindicators: approaches and techniques for the detection, analysis, and validation of novel biomarkers of diseases. *Pharmaceutics*. 2023;15(6):1630. DOI: <http://doi.org/10.3390/pharmaceutics15061630>
25. Lunke S, Bouffler SE, Patel CV, et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat Med*. 2023;29(7):1681–1691. DOI: <http://doi.org/10.1038/s41591-023-02401-9>
26. Sabry Z, Arora HS, Chandrasekaran S, Wang Z. AI-driven drug discovery and repurposing using multi-omics for myocardial infarction and heart failure. *Explor Med*. 2025; 6:1001340. DOI: <http://doi.org/10.37349/emed.2025.1001340>
27. Jiménez A, Merino MJ, Parras J, Zazo S. Explainable drug repurposing via path-based knowledge graph completion. *Sci Rep*. 2024;14(1):16596. DOI: <http://doi.org/10.1038/s41598-024-67163-x>
28. Gimeno M, San José-Enériz E, Villar S, et al. Explainable artificial intelligence for precision medicine in acute myeloid leukaemia. *Front Immunol*. 2022; 13:977358. DOI: <http://doi.org/10.3389/fimmu.2022.977358>
29. Vallée A. Envisioning the future of personalised medicine: role and realities of digital twins. *J Med Internet Res*. 2024;26: e50204. DOI: <http://doi.org/10.2196/50204>
30. Gilbert S, Drummond D, Cotte F, Ziemssen T. Digital twins in medicine: transition from theoretical concept to tool used in everyday care. *Front Digit Health*. 2025;7:1573727. DOI: <http://doi.org/10.3389/fdgh.2025.1573727>
31. Papachristou K, Katsakiori PF, Papadimitroulas P, Strigari L, Kagadis GC. Digital twins' advancements and applications in healthcare, towards precision medicine. *J Pers Med*. 2024;14(11):1101. DOI: <http://doi.org/10.3390/jpm14111101>
32. Wickramasinghe N, Ulapane N, Andargoli A, et al. Digital twins to enable better precision and personalized dementia care. *JAMIA Open*. 2022;5(3): ooac072. DOI: <http://doi.org/10.1093/jamiaopen/ooac072>
33. Surian NU, Batagov A, Wu A, et al. A digital twin model incorporating generalized metabolic fluxes to identify and predict chronic kidney disease in type 2 diabetes mellitus. *NPJ Digit Med*. 2024;7(1):140. DOI: <http://doi.org/10.1038/s41746-024-01108-6>
34. Bordukova M, Makarov N, Rodriguez-Esteban R, et al. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. *Expert Opin Drug Discov*. 2024;19(1):33–42. DOI: <http://doi.org/10.1080/17460441.2023.2273839>
35. Rudhari HK, Tseng B, Zhu H, et al. Digital twins in healthcare: a comprehensive review and future directions. *Front Digit Health*. 2025; 7:1633539. DOI: <http://doi.org/10.3389/fdgh.2025.1633539>

AUTHOR'S PROFILE

Emmanuel Nkansah is a seasoned Data Analytics Professional with a Bachelor's degree in Physics and two Master's degrees, including Business Analytics. His academic background includes specialised training in Quantum Physics and Computational Physics, providing a strong foundation in analytical thinking and mathematics, modelling and complex problem-solving. He possesses extensive experience in database management, data collection, data cleansing, data warehousing, SQL development, exploratory data analysis (EDA), machine learning, predictive modelling, and business reporting. Combining technical expertise with strong business acumen, he leverages data-driven

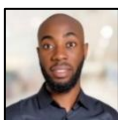


Artificial Intelligence-Driven Multi-Omics Diagnostic Pipelines for Infectious, Neurodegenerative, and Metabolic Diseases: From Biomarker Discovery to Precision Medicine and Digital Twin Healthcare

insights and advanced analytical techniques to develop innovative solutions, optimize organizational performance, support strategic decision-making, and drive sustainable business growth.



Michael Abimbola Oladosu is an interdisciplinary researcher with a strong academic and research background. He is the founder of Research Hub Nexus Institute, a platform dedicated to research publication, academic writing support, and mentorship for emerging scholars. He currently lectures in the Department of Chemical Sciences (Biochemistry Unit), Anchor University, Lagos, Nigeria. His research interests span the intersections of the life sciences, biomedical sciences, and technology, and he has actively collaborated on multidisciplinary research projects that have been published in reputable peer-reviewed journals. He is also a reviewer for several academic journals and serves on the editorial boards of multiple international scientific journals and publications.



Moses Abah Adondua is a dedicated biochemist and lecturer at the Department of Biochemistry, Faculty of Biosciences, Federal University Wukari, Nigeria. With a strong academic foundation and research interests in molecular biology, enzymology, and metabolic disorders, he is actively involved in both teaching and research, mentoring undergraduate students and contributing to innovative research in biochemistry. He is currently the Co-Founder of Research Hub Nexus, where he helps advance research, innovation, and academic collaboration. He also participates in scholarly publishing, peer review, and capacity-building initiatives aimed at strengthening scientific research in Africa and globally, respectively.



Olaide Ayokunmi Oladosu (née Ayansiji) is a dedicated technology professional and researcher. She holds a Bachelor's degree in Computer Science from Babcock University, Ilishan-Remo, Ogun State, Nigeria. She also obtained a Postgraduate Diploma in Computer Science from Teesside University in the United Kingdom. Olaide is a Certified Cisco Network Associate (CCNA). Her passion for continuous learning and innovation in the tech space is reflected in her scholarly contributions. She has co-authored several publications in reputable peer-reviewed journals, demonstrating her commitment to academic and professional excellence. She continues to contribute to research, innovation, and capacity building in the globally respected technology community.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Lattice Science Publication (LSP)/journal and/or the editor(s). The Lattice Science Publication (LSP)/journal and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.