World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(REVIEW ARTICLE)

Check for updates

Machine Learning in Metabolic Syndrome: Toward Precision Diagnosis, Risk Stratification and Personalised Intervention

Shifan Khanday *, Lian wathek, Rehaam Fathima, Fatma Abdulla rahma, umaima syed, Aisha Hassan, Shaikha abulaziz and Zaha zainab manzoor

College of Medicine, Dubai Medical University.

World Journal of Advanced Engineering Technology and Sciences, 2026, 19(01), 271-278

Publication history: Received on 20 March 2026; revised on 25 April 2026; accepted on 28 April 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.19.1.0236

Abstract

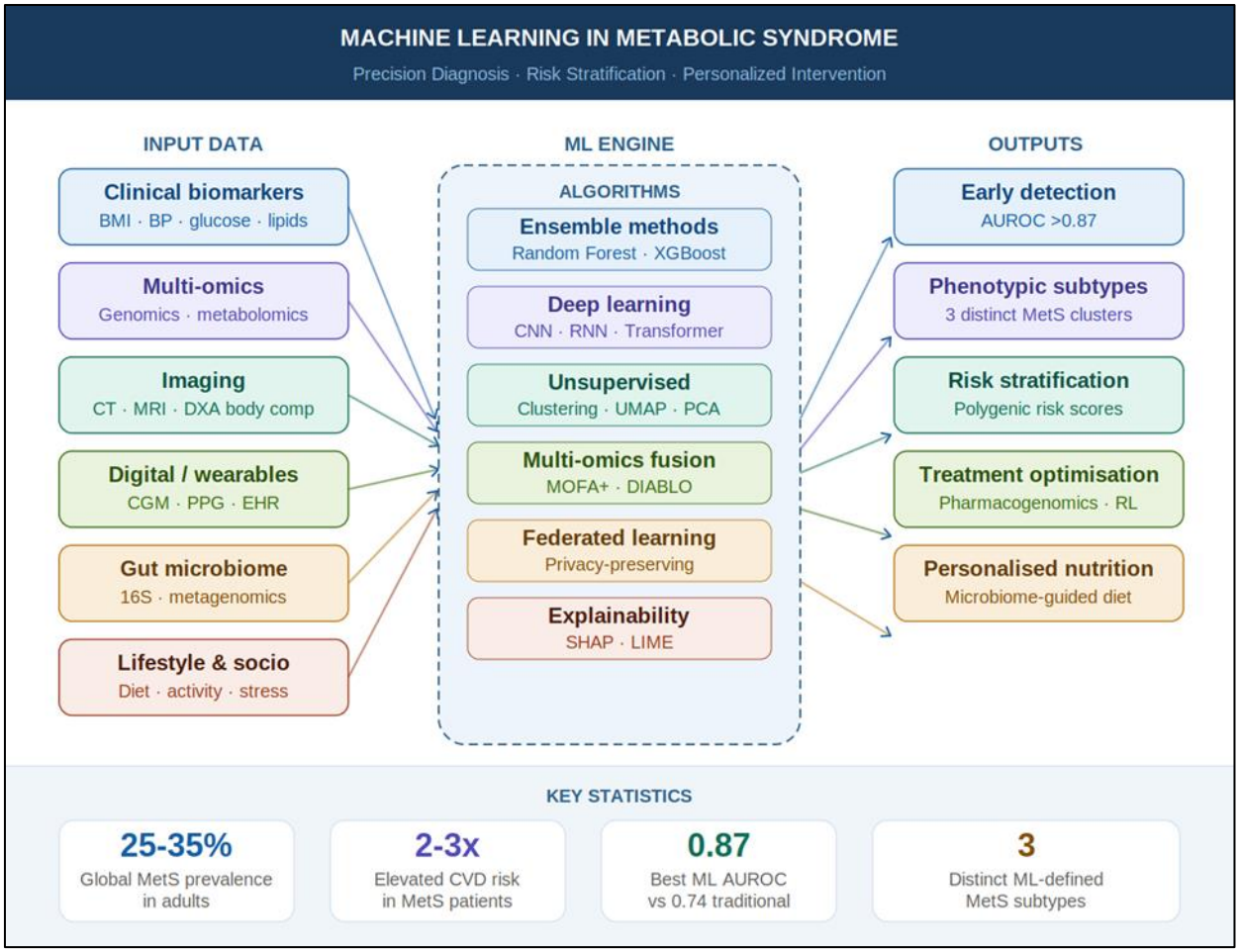
Metabolic syndrome (MetS) — a cluster of interrelated cardiometabolic risk factors including central obesity, insulin resistance, dyslipidemia, and hypertension — affects approximately 25–35% of adults worldwide and confers a 2–3-fold increased risk of cardiovascular disease and type 2 diabetes mellitus. The heterogeneous, multifactorial nature of MetS renders traditional diagnostic and therapeutic frameworks insufficient for capturing its biological complexity. Machine learning (ML) has emerged as a transformative analytical paradigm, offering unprecedented capacity to model non-linear interactions among genomic, metabolomic, clinical, and lifestyle variables. This review synthesises current evidence on ML applications in MetS, spanning early detection and risk prediction to omics integration and digital health optimisation. We critically evaluate algorithmic approaches — including ensemble methods, deep learning, and natural language processing — and discuss key challenges around data quality, model interpretability, algorithmic bias, and clinical translation.

Keywords: Metabolic syndrome; Machine learning; Deep learning; Precision medicine; Cardiometabolic risk; Insulin resistance; Artificial intelligence

Graphical abstract

Graphic abstract: From multi-modal patient data through machine learning algorithms to clinical precision medicine outputs in metabolic syndrome. Key statistics panel shows global epidemiological impact

* Corresponding author: Shifan Khanday



1. Introduction

Metabolic syndrome (MetS) is a clinically recognized clustering of physiological abnormalities that collectively amplify cardiometabolic risk beyond what each component confers independently. Current diagnostic criteria — established by the International Diabetes Federation (IDF), the American Heart Association (AHA), and the World Health Organization (WHO) — require the presence of three or more of the following: abdominal obesity, elevated fasting plasma glucose, hypertriglyceridemia, reduced high-density lipoprotein cholesterol (HDL-C), and elevated blood pressure. Despite the apparent simplicity of these criteria, MetS represents a profoundly heterogeneous syndrome with diverse pathophysiological underpinnings, including chronic low-grade inflammation, oxidative stress, gut microbiome dysbiosis, and hypothalamic-pituitary-adrenal axis dysregulation.

The global prevalence of MetS has risen dramatically in parallel with urbanization, dietary shifts, and sedentary lifestyles. Epidemiological estimates suggest that MetS affects between 20% and 35% of the adult population globally, with particularly high rates observed in Gulf Cooperation Council (GCC) countries, the United States, and parts of Southeast Asia. In the GCC region, prevalence rates exceeding 40% have been reported, largely attributable to rapid socioeconomic transitions and high rates of obesity and type 2 diabetes.

Traditional statistical models — such as logistic regression, Cox proportional hazards regression, and Framingham risk scores — while valuable, are inherently limited by assumptions of linearity, inability to handle high-dimensional data, and failure to capture complex variable interactions. Machine learning (ML), a subset of artificial intelligence (AI) encompassing supervised, unsupervised, and reinforcement learning paradigms, addresses many of these limitations. ML algorithms are capable of identifying latent patterns in large, heterogeneous datasets without requiring a priori specification of relationships, making them ideally suited to the complexity of MetS biology.

This review provides a comprehensive synthesis of ML applications across the MetS continuum: from early detection and phenotypic subclassification to multi-omics integration, pharmacogenomic optimization, and remote monitoring via wearable technologies.

2. Pathophysiology of Metabolic Syndrome: A Systems Perspective

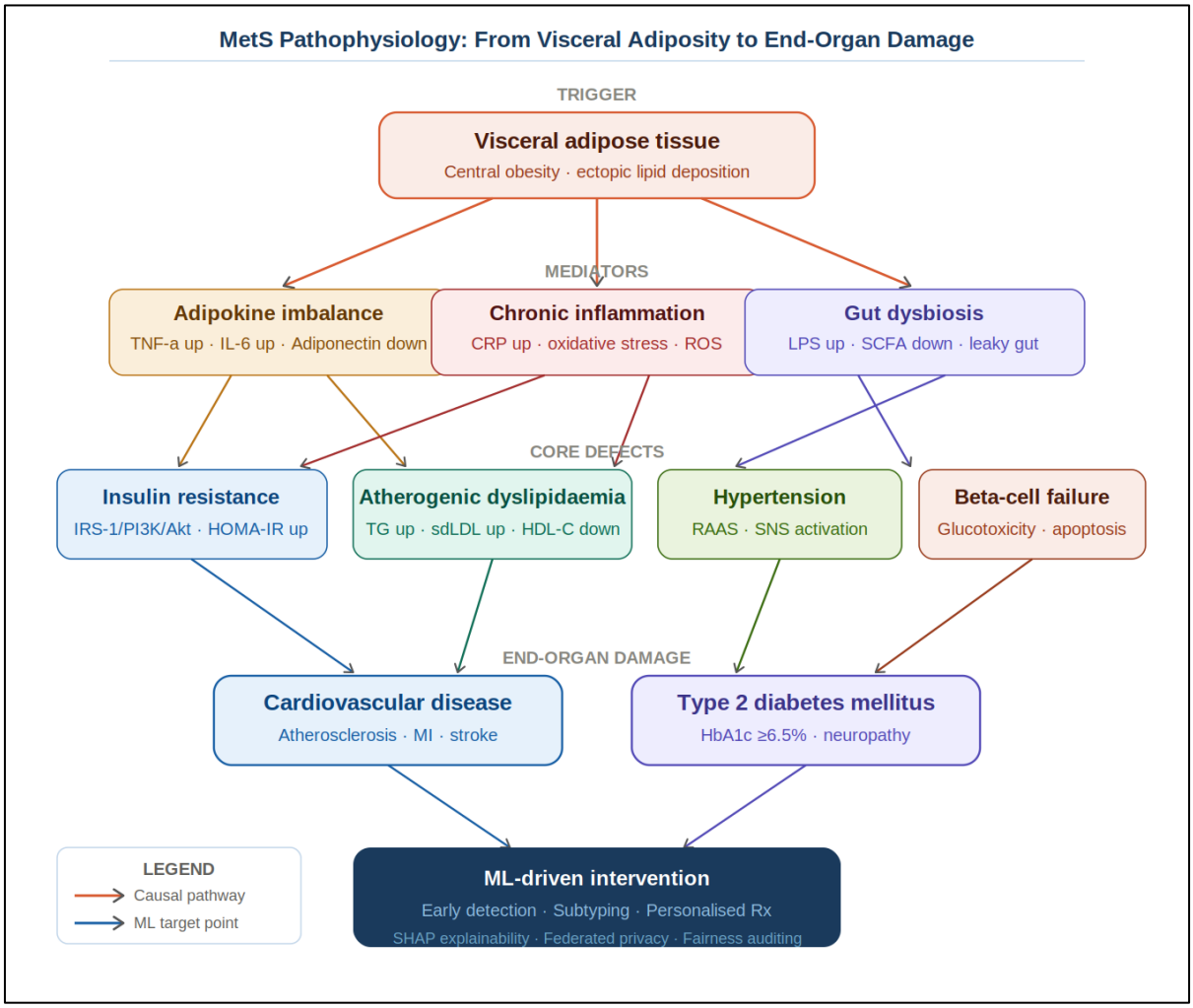


Figure 1 Pathophysiological cascade in metabolic syndrome. Visceral adiposity initiates a multi-tier cascade of adipokine imbalance, chronic inflammation, and gut dysbiosis, converging on insulin resistance, atherogenic dyslipidaemia, hypertension, and beta-cell failure, with ultimate end-organ damage as cardiovascular disease and type 2 diabetes. ML-driven interventions are indicated at the convergence point

2.1. Central Adiposity and Insulin Resistance

The pathological accumulation of visceral adipose tissue (VAT) is central to MetS pathogenesis. Unlike subcutaneous adipose tissue, VAT is metabolically active, secreting pro-inflammatory adipokines — including tumor necrosis factor-alpha (TNF-α), interleukin-6 (IL-6), and resistin — while paradoxically suppressing the secretion of anti-inflammatory adiponectin. This adipokine imbalance promotes hepatic insulin resistance through ectopic lipid deposition and ceramide-mediated interference with the insulin receptor substrate-1 (IRS-1)/phosphatidylinositol-3-kinase (PI3K)/Akt signaling cascade.

Compensatory hyperinsulinemia further drives dyslipidemia through enhanced hepatic very-low-density lipoprotein (VLDL) synthesis and impaired lipoprotein lipase activity, resulting in the characteristic atherogenic dyslipidemia of MetS: elevated triglycerides, small dense low-density lipoprotein (sdLDL) particles, and reduced HDL-C.

2.2. Inflammation and Oxidative Stress

Chronic, low-grade systemic inflammation is a defining feature of MetS. Elevated circulating levels of C-reactive protein (CRP), IL-6, and TNF-α reflect ongoing immune activation driven by visceral adiposity, endotoxemia from gut microbiome dysbiosis, and mitochondrial dysfunction. Reactive oxygen species (ROS) generated by dysfunctional

mitochondria and NADPH oxidase further impair insulin signaling, endothelial nitric oxide bioavailability, and beta-cell function, creating a self-reinforcing cycle of metabolic deterioration.

2.3. The Gut-Liver-Brain Axis

Emerging evidence implicates gut microbiome composition as a key modulator of MetS risk. Dysbiosis — characterized by reduced microbial diversity and depletion of short-chain fatty acid (SCFA)-producing genera such as Faecalibacterium and Akkermansia — impairs intestinal barrier integrity, promoting bacterial lipopolysaccharide (LPS) translocation and hepatic toll-like receptor 4 (TLR-4) activation. This endotoxemia amplifies hepatic inflammation and steatosis, mechanistically linking gut ecology to MetS pathogenesis.

3. Machine Learning Algorithms: Technical Foundations

3.1. Supervised Learning

Supervised ML algorithms learn a mapping function from labeled input features to a defined output variable. In the context of MetS, inputs may include anthropometric measurements, biochemical assays, imaging data, or genomic variants, while outputs may include binary MetS status, continuous cardiometabolic risk scores, or time-to-event endpoints. Widely applied supervised algorithms include Random Forests (RF), Gradient Boosting Machines (GBM), Support Vector Machines (SVM), and deep neural network architectures.

XGBoost and LightGBM represent state-of-the-art gradient boosting implementations with strong performance across tabular clinical data. Convolutional neural networks (CNNs) are particularly suited to imaging-based phenotyping, while recurrent neural networks (RNNs) and transformer architectures excel at longitudinal clinical data and electronic health record (EHR) mining.

3.2. Unsupervised Learning and Disease Subtyping

Unsupervised methods do not require labeled outputs, instead identifying intrinsic structure within data. K-means clustering, hierarchical clustering, and latent Dirichlet allocation (LDA) have been applied to delineate biologically meaningful MetS subtypes. Dimensionality reduction techniques — principal component analysis (PCA), uniform manifold approximation and projection (UMAP), and t-distributed stochastic neighbor embedding (t-SNE) — facilitate visualization and feature compression in high-dimensional omics datasets.

3.3. Federated and Transfer Learning

Federated learning enables model training across decentralized datasets without centralizing patient data, addressing privacy regulations such as GDPR and HIPAA. Transfer learning leverages representations learned in one domain to accelerate training in data-sparse clinical settings, offering practical advantages in metabolic imaging analysis.

4. ML Applications in Metabolic Syndrome

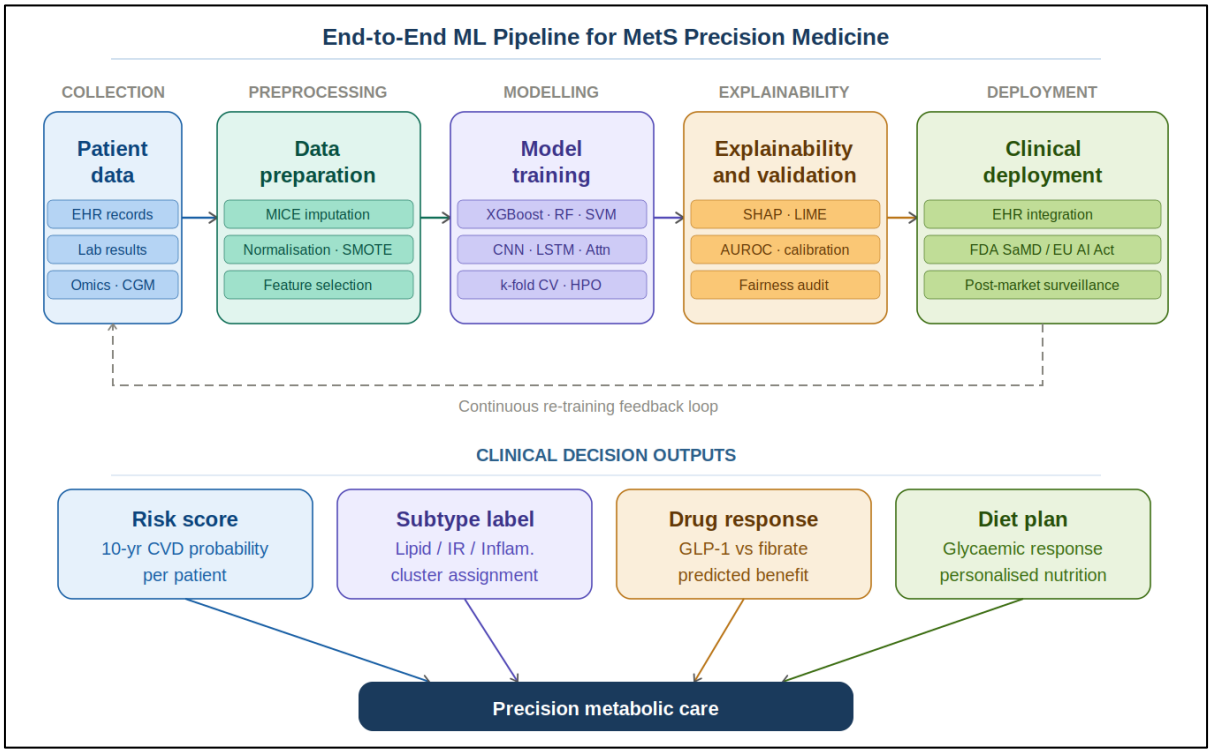


Figure 2 End-to-end machine learning pipeline for metabolic syndrome precision medicine. The pipeline spans data collection through clinical deployment with a continuous re-training feedback loop, and produces four categories of clinical decision outputs: risk scores, phenotypic subtype labels, drug response predictions, and personalised dietary plans

4.1. Early Detection and Risk Prediction

Several studies have demonstrated the superiority of ML over traditional logistic regression in predicting incident MetS. A landmark analysis utilizing the Framingham Heart Study cohort demonstrated that a random forest model incorporating 22 clinical variables achieved an area under the receiver-operating characteristic curve (AUROC) of 0.87, outperforming the conventional Adult Treatment Panel III (ATP III) score (AUROC 0.74). Gradient boosting classifiers applied to UK Biobank data identified novel predictors of MetS onset, including accelerometer-derived sedentary time patterns and dietary inflammatory index scores.

Wearable-derived data has opened new frontiers in continuous MetS risk monitoring. Photoplethysmography (PPG) signals from consumer smartwatches have been leveraged to estimate insulin resistance proxies through deep convolutional networks. Continuous glucose monitoring (CGM) data, analyzed through recurrent neural networks, has demonstrated capacity to predict glycemic deterioration trajectories days to weeks in advance of clinical presentation.

4.2. Phenotypic Subclassification

The clinical heterogeneity of MetS demands subtype-aware therapeutic approaches. Unsupervised clustering of metabolomic profiles has identified at least three biologically distinct MetS subtypes: a lipid-predominant cluster, an insulin-resistance-predominant cluster, and an inflammatory cluster. These subtypes exhibit differential responses to interventions, with the insulin-resistance cluster demonstrating superior glycemic benefit from GLP-1 receptor agonists, while the lipid-predominant subtype responds preferentially to omega-3 supplementation and fibrate therapy.

4.3. Multi-Omics Integration

The integration of genomics, transcriptomics, proteomics, metabolomics, and microbiomics within unified ML frameworks represents a paradigm shift in MetS research.

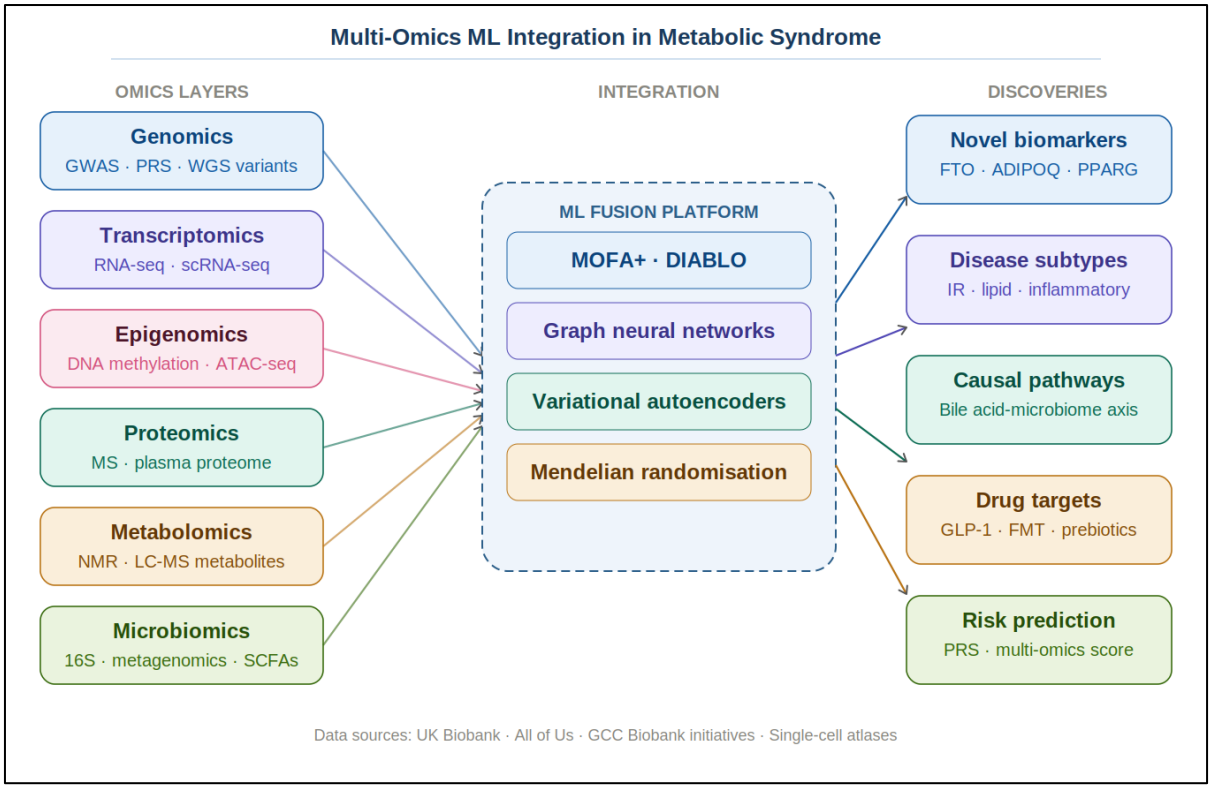


Figure 3 Multi-omics ML integration framework for metabolic syndrome. Six omics data layers converge on an ML fusion platform incorporating MOFA+, graph neural networks, variational autoencoders, and Mendelian randomisation, yielding novel biomarker discoveries, disease subtype definitions, causal pathway identification, drug target nomination, and polygenic risk scores

Multi-omics integration platforms — including MOFA+ (Multi-Omics Factor Analysis), DeepMOCCA, and DIABLO — decompose shared variance across omics layers, identifying trans-omic regulatory modules. A recent multi-omics ML study identified a gut microbiome–plasma metabolome axis mediated by secondary bile acid metabolism as a key determinant of insulin resistance severity. Genome-wide association study (GWAS) data, integrated with transcriptomic and metabolomic profiles, has implicated novel loci including variants in ADIPOQ, FTO, and PPARG.

4.4. Imaging-Based Phenotyping

Deep CNN models trained on abdominal CT and MRI scans achieve near-radiologist accuracy in automated segmentation of visceral and subcutaneous adipose tissue compartments, liver fat quantification, and pancreatic fat estimation. DXA-derived body composition models have demonstrated superior predictive validity for insulin resistance compared to BMI-based indices.

4.5. Treatment Optimization and Pharmacogenomics

ML is transforming pharmacological management of MetS by enabling individualized drug response prediction. Reinforcement learning algorithms applied to EHR data have been used to optimize antihypertensive and lipid-lowering regimens. Pharmacogenomic ML models incorporating cytochrome P450 (CYP) enzyme variants have demonstrated utility in predicting statin and metformin response heterogeneity. Personalized nutrition algorithms utilizing gut microbiome composition and CGM-derived postprandial glycemic profiles have outperformed population-averaged dietary guidelines in randomized controlled trials.

5. Challenges and Limitations

5.1. Data Quality and Harmonization

The performance of ML models is fundamentally constrained by data quality. Clinical datasets are frequently characterized by missing values, measurement error, heterogeneous data collection protocols, and survivorship bias. Imputation strategies — including multiple imputation by chained equations (MICE) and deep learning-based

imputation — partially address missingness but introduce uncertainty that must be propagated through downstream analyses.

5.2. Model Interpretability

The opacity of complex ML models constitutes a significant barrier to clinical adoption. Explainability frameworks including SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) provide post-hoc feature attribution, enabling clinicians to audit model reasoning. Inherently interpretable models remain valuable benchmarks in settings where regulatory transparency requirements apply.

5.3. Algorithmic Bias and Health Equity

Most published ML models in MetS research have been trained predominantly on European ancestry cohorts, limiting generalizability to populations that bear disproportionate MetS burden. Waist circumference thresholds for abdominal obesity differ across ethnic groups, yet many ML models fail to incorporate ethnicity-specific diagnostic adjustments. Prospective fairness auditing and diversification of training cohorts are ethical imperatives.

5.4. Regulatory and Clinical Translation

Regulatory frameworks — including FDA's Software as a Medical Device (SaMD) guidelines and the EU AI Act — require rigorous prospective validation, predefined performance benchmarks, and transparency in algorithmic decision-making. Successful clinical translation additionally demands clinician engagement and integration with existing clinical workflows.

6. Ethical Considerations

The deployment of ML in MetS management raises important ethical considerations spanning privacy, autonomy, and justice. The aggregation of genomic, microbiome, and continuous physiological data required by precision MetS models generates intimate phenotypic profiles whose secondary use must be strictly regulated. Informed consent processes must evolve to adequately communicate probabilistic risk estimates and model uncertainty to patients.

Algorithmic accountability frameworks must ensure that ML-generated clinical recommendations are auditable and contestable. The risk of automation bias necessitates integration of AI tools as decision support rather than decision replacement, preserving physician judgment and the therapeutic relationship.

7. Future Directions

The convergence of ML with emerging technologies promises to further transform MetS research and care. Spatial transcriptomics and single-cell multi-omics will enable ML-based mapping of cell-type-specific metabolic dysregulation. Foundation models — large, pre-trained transformer architectures — fine-tuned on EHR and omics data offer the prospect of universal metabolic phenotyping models applicable across diverse clinical settings.

Digital twin technology — patient-specific computational simulations informed by ML-trained physiological models — may enable in silico clinical trial design and personalized intervention simulation prior to pharmacological prescription. Large-scale international consortium efforts, including the UK Biobank, All of Us Research Program, and emerging GCC biobank initiatives, will provide the diverse longitudinal datasets necessary to train generalizable, equitable MetS ML models.

8. Conclusions

Machine learning represents a transformative force in the diagnosis, subclassification, and management of metabolic syndrome. By transcending the limitations of conventional statistical paradigms, ML enables the extraction of clinically actionable insights from the biological complexity that defines MetS. Current evidence supports the utility of ensemble methods, deep learning, and multi-omics integration for improved risk prediction, phenotypic stratification, and treatment personalization.

Addressing data quality deficiencies, ensuring model interpretability, mitigating algorithmic bias, and establishing robust regulatory and ethical frameworks are prerequisites for responsible clinical deployment. A collaborative, interdisciplinary commitment — spanning data scientists, clinicians, ethicists, patients, and regulators — is essential to translate the analytical power of ML into measurable improvements in cardiometabolic health at population scale.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare no conflict of interest

References

[1] Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement. Circulation. 2009;120(16):1640–1645.

[2] Rochlani Y, Pothineni NV, Kovelamudi S, Mehta JL. Metabolic syndrome: pathophysiology, management, and modulation by natural compounds. Ther Adv Cardiovasc Dis. 2017;11(8):215–225.

[3] Cichosz SL, Johansen MD, Hejlesen O. Toward big data analytics: review of predictive models in management of diabetes and its complications. J Diabetes Sci Technol. 2016;10(1):27–34.

[4] Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases. PLOS Med. 2015;12(3):e1001779.

[5] Zeevi D, Korem T, Zmora N, et al. Personalized nutrition by prediction of glycemic responses. Cell. 2015;163(5):1079–1094.

[6] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–4774.

[7] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.

[8] Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis. PLOS Med. 2018;15(11):e1002686.

[9] Zou J, Schiebinger L. AI can be sexist and racist — it's time to make it fair. Nature. 2018;559(7714):324–326.

[10] Argelaguet R, Arnol D, Bredikhin D, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020;21(1):111.