REVIEW ARTICLE

Eurasian Journal of Medical and Biological Sciences
AN INTERNATIONAL JOURNAL
Eurasian J Med Biol Sci **6(1): 9-13**

Innovative Approaches to Predicting Metabolic Diseases Using Bioinformatics and Machine Learning Technologies

Bolat Seidanov[1*], Ingkar Okhas[1], Nurzhanyat Ablaikhanova[1]

[1] *Department of Biophysics, Biomedicine and Neuroscience, Farabi University, 050000 Almaty, Kazakhstan*

Received : 10/09/2025 Revised : 16/05/2026 Accepted : 29/05/2026

ABSTRACT: Contemporary approaches to predicting metabolic diseases – such as metabolic syndrome, obesity, and type 2 diabetes – are increasingly grounded in bioinformatics methods, including genomic, metabolomic, and polygenic risk modeling, as well as machine learning and multi-omics data integration. This review highlights key trends in the field: the identification of genetic, transcriptomic, and metabolic biomarkers; the development of polygenic and multi-omics predictive models; the use of cohort and prospective studies; and methodological innovations aimed at improving accuracy and translational relevance. Particular emphasis is placed on machine learning approaches based on genetic and metabolomic data, exploring both their potential and limitations. Recent studies are discussed in which models incorporating genetic variation and dietary factors have achieved high predictive performance. Furthermore, the review examines how integrating "omics" data (genomics and metabolomics) with clinical features enhances prediction, captures inter-population variability, and enables responsiveness to lifestyle factors. Finally, future directions are outlined, including data standardization, increasing ethnic diversity in datasets, development of interpretable and clinically applicable models, and validation of intervention outcomes across independent cohorts.

Keywords: Bioinformatics, metabolomics, metabolic syndrome, machine learning, multi-omics.

INTRODUCTION

Metabolic syndrome (MetS) is a complex of interrelated metabolic disorders that includes abdominal obesity, dyslipidemia (elevated triglyceride levels and decreased high-density lipoprotein levels), elevated blood pressure, and hyperglycemia (Neeland et al., 2024; Saklayen, 2018). This condition represents one of the most significant challenges to global healthcare, as it greatly increases the risk of developing type 2 diabetes (T2D) and cardiovascular diseases (CVD). According to estimates, MetS affects about 30% of the adult population, making it a global epidemic (Mohamed et al., 2023).

The etiology of MetS is complex and determined by the interaction between genetic factors and environmental influences, such as unbalanced nutrition and low physical activity (Islam et al., 2024). In this regard, early diagnosis and prediction of the risk of developing metabolic diseases are crucial for timely intervention and prevention of complications.

In recent years, bioinformatics analysis and "omics" technologies such as genomics, proteomics, and metabolomics have become powerful tools for studying complex diseases (Gharipour et al., 2022). These approaches make it possible not only to identify molecular markers but also to create comprehensive predictive models, opening the era of personalized medicine in managing metabolic risks. This review is devoted to the analysis of modern bioinformatics approaches to predicting metabolic diseases, with a special focus on the integration of genomic and metabolomic data through machine learning methods.

MATERIALS AND METHODS

Contemporary approaches to predicting metabolic diseases – such as metabolic syndrome, obesity, and type 2 diabetes – are increasingly grounded in bioinformatics methods, including genomic, metabolomic, and polygenic risk modeling, as well as machine learning and multi-omics data integration.

The processing of vast and complex "omics" datasets necessitates the use of advanced computational methods. Machine learning (ML) has become an indispensable tool for developing accurate predictive models capable of identifying nonlinear relationships and intricate interactions among hundreds or even thousands of variables (Galal et al., 2022). By enabling the discovery of multidimensional patterns that traditional statistical approaches often fail to capture, ML provides a robust framework for precision risk prediction and data-driven decision-making in biomedical research.

* bolatseidanov@gmail.com

RESULTS

Identification of Biomarkers for Predicting Metabolic Diseases: Hereditary predisposition plays a significant role in the development of MetS, as evidenced by family and twin studies that estimate the heritability of the syndrome at 24% or higher. The genetic risk is primarily based on single nucleotide polymorphisms (SNPs) in genes involved in the regulation of lipid metabolism, insulin sensitivity, blood pressure, and inflammatory processes. Among the key candidate genes, FTO (associated with fat mass and obesity), APOC3 and APOA5 (regulators of triglyceride levels), and PPARG (involved in insulin sensitivity) are particularly notable (Rana et al., 2022). However, the contribution of each individual SNP to the overall risk of developing MetS is relatively small. To achieve a more accurate assessment of cumulative genetic risk, polygenic risk scores (PGS) have been developed, which aggregate the effects of numerous genetic variants across the genome (Abaj et al., 2025). PGS enable the identification of individuals at high genetic risk long before the manifestation of clinical symptoms. An important advantage of PGS is their lifelong stability, as genetic variation remains constant throughout an individual's life. Nevertheless, studies indicate that for most metabolic diseases, the predictive performance of PGS is generally lower than that of models based on metabolomic data, although PGS provide complementary and biologically informative insights that enhance multi-omics predictive modeling (Timasheva et al., 2024).

Metabolomic Biomarkers and Predictive Models

Metabolomics investigates the comprehensive set of low-molecular-weight metabolites present in biological samples, providing a "snapshot" of the organism's physiological state at a given moment (Gharipour et al., 2022). This characteristic makes metabolomics a powerful tool for identifying biomarkers that reflect ongoing pathophysiological processes in real time.

Large-scale population-based studies have identified a number of metabolites consistently associated with MetS. For example, elevated concentrations of branched-chain amino acids (valine, leucine/isoleucine) and aromatic amino acids (phenylalanine, tyrosine) have been shown to correlate positively with MetS risk. In contrast, levels of glycine and serine, as well as various lipids such as lysophosphatidylcholine (lysoPC a C18:2), demonstrate negative associations. Notably, reduced concentrations of lysoPC a C18:2 are linked not only with MetS itself but also with each of its five clinical components, suggesting a potential protective metabolic role for this compound (Shi et al., 2023).

Analogous to polygenic risk scores (PGS), metabolomic risk scores have been developed to integrate information derived from dozens of individual metabolites. These metabolomic models exhibit high predictive accuracy for a wide range of diseases, including type 2 diabetes, myocardial infarction, and liver disorders. A key advantage of metabolomic profiling lies in its dynamic nature: unlike static genetic information, metabolomic signatures can change in response to lifestyle factors, dietary habits, or therapeutic interventions. This flexibility allows for continuous monitoring of disease risk and provides a framework for assessing the effectiveness of preventive or therapeutic strategies over time (Table 1) (Nightingale Health Biobank Collaborative Group, 2024).

Table 1. Comparison of Genomic and Metabolomic Approaches in Metabolic Disease Prediction

Feature	**Genomic Approaches (e.g., PGS)**	**Metabolomic Approaches (e.g., Metabolomic Risk Scores)**
Nature of Data	Static; remains constant throughout life.	Dynamic; provides a real-time "snapshot" of physiology.
What it Represents	Lifelong hereditary predisposition and cumulative genetic risk.	Current physiological state, reflecting gene-environment interactions.
Key Examples from Text	Single Nucleotide Polymorphisms (SNPs) in genes like FTO, APOC3, PPARG; Polygenic Risk Scores (PGS).	Branched-chain amino acids (BCAAs), aromatic amino acids, glycine, LysoPC a C18:2; Metabolomic Risk Scores.
Response to Intervention	Does not change in response to lifestyle or diet.	Changes in response to lifestyle, diet, or therapeutic interventions.
Primary Utility	Identifying high-risk individuals long before clinical onset; provides complementary biological insights.	High predictive accuracy for current disease risk; monitoring risk over time and effectiveness of interventions.
Relative Performance	Predictive performance is generally lower than metabolomic models for most metabolic diseases (Timasheva et al., 2024).	Often exhibits higher predictive accuracy than PGS for conditions like T2D and myocardial infarction (Nightingale Health, 2024).

Machine Learning as a Key Analytical Tool

A recent 14-year prospective cohort study conducted in a Korean population serves as a compelling example of the effectiveness of this approach. In this study, ML-based models were developed using algorithms such as Random Forest (RF) and AdaBoost (ADB) to predict the onset of MetS. The models incorporated a wide range of features, including demographic and clinical parameters, lifestyle factors (alcohol consumption, smoking), dietary habits (particularly the intake of seaweed), and genomic information represented by a genetic polygenic risk score (gPRS).

The results demonstrated exceptional predictive performance: the area under the receiver operating characteristic curve (AUC) for the RF and ADB models reached 0.994, indicating near-perfect classification accuracy. These findings underscore that integrating genetic, dietary, and clinical variables through ML-based frameworks can yield highly efficient tools for risk stratification and personalized prevention of metabolic syndrome (Shin, 2024).

Integration of Multi-Omics Data to Enhance Predictive Accuracy

One of the most promising contemporary approaches is multi-omics integration — the combination of data from various "omics" layers (genomics, transcriptomics, proteomics, and metabolomics) with detailed clinical information. Genomic data, such as polygenic risk scores (PGS), represent an individual's lifelong genetic predisposition to disease, whereas metabolomic data capture the current physiological state, reflecting the complex interplay among genes, environment, and lifestyle factors (Sanches et al., 2024).

A large-scale investigation utilizing data from three national biobanks and encompassing over 700,000 participants demonstrated that genomic and metabolomic scores provide complementary information in disease prediction. For most conditions, including myocardial infarction and type 2 diabetes, a combined multi-omics model integrating both PGS and metabolomic risk scores exhibited the highest predictive performance. Importantly, individuals characterized by both a high genetic predisposition and an "unfavorable" metabolic profile were found to have a markedly elevated risk of developing disease compared with those exhibiting elevation in only one of these dimensions.

Furthermore, the inclusion of multi-omics data into existing clinical risk models—such as QRISK3 for cardiovascular disease—resulted in a statistically significant improvement in predictive accuracy across the majority of studied disorders. These findings emphasize the translational potential of integrative multi-omics frameworks for refining disease risk stratification, guiding personalized prevention, and advancing the paradigm of precision medicine (Nightingale Health Biobank Collaborative Group, 2024).

Synthesis of Predictive Modeling Results

The synthesis of these findings reveals a clear trajectory from single-marker identification to integrated risk modeling. For instance, while the work reviewed by Rana et al. (2022) identified foundational candidate genes like FTO and PPARG, subsequent research, such as that by Timasheva et al. (2024), quantitatively established that polygenic scores (PGS) derived from these markers, while biologically informative, are predictively outperformed by dynamic metabolomic data. This metabolomic layer, as detailed by Shi et al. (2023), provides actionable, real-time physiological data, such as the crucial negative correlation of lysoPC a C18:2 with all five MetS components. The culmination of this trend is evidenced in the multi-omics approach of the Nightingale Health Biobank Collaborative Group (2024), which empirically confirmed that the combination of static genomic risk (PGS) and dynamic metabolic risk scores yields superior predictive accuracy than either modality alone or even established clinical models like QRISK3. This complementarity is the central finding emerging from the current literature, powerfully illustrated by the near-perfect AUC of 0.994 achieved by Shin (2024) when ML models integrated gPRS with dynamic factors like nutrition.

DISCUSSION

Prospects and Challenges

Despite remarkable progress in bioinformatics-driven prediction of metabolic diseases, several significant challenges remain on the path toward broad clinical implementation of these predictive models.

1. **Standardization and Validation**. Ensuring the reliability and reproducibility of results requires strict standardization of data collection, preprocessing, and analytical methods. Predictive models must undergo rigorous validation across large, independent, and demographically diverse cohorts, as demonstrated by the tri-biobank study (Collin et al., 2022). Without such harmonization, discrepancies between datasets may compromise model performance and generalizability.

2. **Ethnic Diversity**. The majority of existing genomic and metabolomic studies are conducted on populations of European ancestry, which limits their applicability to other ethnic groups. Models trained on homogeneous datasets often exhibit

reduced predictive accuracy when transferred across ancestries. Therefore, expanding ethnic and regional diversity in multi-omics research is a critical priority to ensure global clinical relevance and equity in predictive medicine.

3. **Clinical Implementation and Interpretability**. For predictive models to be adopted in clinical settings, they must be both highly accurate and interpretable. Clinicians need to understand which features contribute most significantly to predictions in order to provide evidence-based recommendations to patients. Translating complex, black-box models into clinically actionable, transparent tools remains one of the most pressing challenges in the field (Gharipour et al., 2022). Approaches such as explainable AI (XAI) and model simplification are being explored to bridge this gap.

4. **Evaluation of Intervention Effects**. The dynamic nature of metabolomic biomarkers provides a unique opportunity to monitor the effectiveness of preventive and therapeutic interventions. Longitudinal studies with repeated metabolomic assessments have demonstrated that individuals transitioning from a high-risk to a low-risk metabolic profile—through lifestyle modification or dietary intervention—exhibit a significant reduction in disease incidence. This finding underscores the potential of metabolomics as a tool for tracking real-time health dynamics and guiding personalized prevention strategies (Shin, 2024; Nightingale Health Biobank Collaborative Group, 2024).

Implications for Clinical Intervention and Monitoring

Elaborating on the evaluation of intervention effects, the findings from the Nightingale Health Biobank Collaborative Group (2024) and Shin (2024) present a significant paradigm shift for clinical practice. The dynamic nature of the metabolome, which reflects responses to lifestyle and diet, positions metabolomic risk scores as a viable tool for real-time therapeutic monitoring. Unlike static PGS, a metabolomic profile can quantify the physiological impact of an intervention, such as the inclusion of specific dietary factors like seaweed noted by Shin (2024). This offers a pathway to empirically validate the effectiveness of personalized lifestyle modifications, moving beyond general advice to data-driven, individualized health management. Furthermore, the demonstrated ability to shift from a high-risk to a low-risk metabolic profile, and the corresponding reduction in disease incidence (Nightingale Health Biobank Collaborative Group, 2024), provides a powerful motivational tool for patient adherence and engagement in preventive health strategies. This dynamism, captured by metabolomics (Gharipour et al., 2022) and integrated via ML, allows for a shift from disease prediction to active health management, where interventions can be adjusted based on molecular feedback.

CONCLUSION

Modern computational approaches based on the integration of bioinformatics analysis and machine learning methods are fundamentally transforming strategies for predicting metabolic diseases. The shift from analyzing isolated risk factors to integrating multi-omics data through machine learning enables the development of highly accurate and personalized predictive models. The combination of static genetic risk (polygenic scores, PGS) with dynamic metabolic profiles provides a more comprehensive understanding of individual susceptibility to disease. Overcoming current challenges related to data standardization, heterogeneity, and clinical implementation will enable these advanced tools to be effectively applied for the prevention and management of metabolic disorders at the individual level in the near future.

ACKNOWLEDGEMENTS

None

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

Abaj, F., Aali, Y., Najafi, F., Koohdani, F. (2025). The relationship of genetic signature for cardiometabolic risk with biomarkers of inflammatory and oxidative stress in diabetic patients. BMC Endocr. Disord., 25, 148.

Collin, C. B., Gebhardt, T., Golebiewski, M., Karaderi, T., Hillemanns, M., Khan, F. M., Salehzadeh-Yazdi, A., Kirschner, M., Krobitsch, S., Eu-Stands PM Consortium, Kuepfer, L. (2022). Computational models for clinical applications in personalized medicine: Guidelines and recommendations for data integration and model validation. J. Pers. Med., 12(2), 166.

Galal, A., Talal, M., Moustafa, A. (2022). Applications of machine learning in metabolomics: Disease modeling and classification. Front. Genet., 13, 1017340.

Gharipour, M., Nezafati, P., Sadeghian, L., Eftekhari, A., Rothenberg, I., Jahanfar, S. (2022). Precision medicine and metabolic syndrome. ARYA Atheroscler., 18, 2475.

Islam, M. S., Wei, P., Suzauddula, M., Nime, I., Feroz, F., Acharjee, M., Pan, F. (2024). The interplay of factors in metabolic syndrome: Understanding its roots and complexity. Mol. Med., 30(1), 279.

Mohamed, S. M., Shalaby, M. A., El-Shiekh, R. A., El-Banna, H. A., Emam, S. R., Bakr, A. F. (2023). Metabolic syndrome: Risk factors, diagnosis, pathogenesis, and management with natural approaches. Food Chem. Adv., 3, 100335.

Neeland, I. J., Lim, S., Tchernof, A., et al. (2024). Metabolic syndrome. Nat. Rev. Dis. Primers, 10, 77.

Nightingale Health Biobank Collaborative Group. (2024). Metabolomic and genomic prediction of common diseases in 700,217 participants in three national biobanks. Nat. Commun., 15, 10092.

Rana, S., Ali, S., Wani, H. A., Mushtaq, Q. D., Sharma, S., Rehman, M. U. (2022). Metabolic syndrome and underlying genetic determinants: A systematic review. J. Diabetes Metab. Disord., 21, 1095-1104.

Saklayen, M. G. (2018). The global epidemic of the metabolic syndrome. Curr. Hypertens. Rep., 20, 12.

Sanches, P. H. G., de Melo, N. C., Porcari, A. M., de Carvalho, L. M. (2024). Integrating molecular perspectives: Strategies for comprehensive multi-omics integrative data analysis and machine learning applications in transcriptomics, proteomics, and metabolomics. Biology, 13(11), 848.

Shi, M., Han, S., Klier, K., Fobos, G., Montrone, C., Yu, S., et al. (2023). Identification of candidate metabolite biomarkers for metabolic syndrome and its five components in population-based human cohorts. Cardiovasc. Diabetol., 22, 141.

Shin, D. (2024). Prediction of metabolic syndrome using machine learning approaches based on genetic and nutritional factors: A 14-year prospective-based cohort study. BMC Med. Genomics, 17, 224.

Timasheva, Y., Kochetova, O., Balkhiyarova, Z., Korytina, G., Prokopenko, I., Nouwen, A. (2024). Polygenic score approach to predicting risk of metabolic syndrome. Genes, 16(1), 22.