American Journal of Biomedicine and Pharmacy
https://biojournals.us/index.php/AJBP
ISSN: 2997-7177

Next-Generation Biomarkers in Clinical Chemistry: Integrating Metabolomics and AI for Early Disease Detection

Fatima Khamis Mahdi Saleh
University of Fallujah College of Applied Sciences Department of Applied Chemistry

Sarah Arshad Mustafa Mohammed
University of Mosul College of Science, Department of Chemistry

Hajar Ahmed Hassan Ali, Luma maky abass babey
University of Basra College of science Department of Biology

Ola Mahmood Abood Mahmood
University of Baghdad College of Science Department of Biotechnology

Received: 2024, 15, Jun
Accepted: 2025, 21, Jul
Published: 2025, 05, Aug

Annotation: This study explores the integration of metabolomics and artificial intelligence (AI) to develop next-generation biomarkers for early disease detection, addressing limitations of current diagnostic methods such as high costs, poor scalability, and interpretive complexity. Despite advances in omics technologies, the translation of metabolomics into clinical practice remains underdeveloped due to data heterogeneity and computational challenges. By employing machine learning algorithms on large-scale metabolomics datasets, the research demonstrates significant improvements in diagnostic accuracy across diseases like cancer, cardiovascular conditions, and neurodegenerative disorders. The Smart Diagnostics platform a microfluidic biosensor system integrated with AI enables real-time, low-cost, and minimally invasive testing with superior sensitivity and specificity compared to traditional methods. Results indicate that AI-augmented metabolomics can outperform conventional biomarkers and facilitate

personalized, point-of-care medicine. This integrated framework presents a scalable solution to enhance early disease prognosis and streamline clinical workflows, especially in resource-limited settings.

Keywords: metabolomics, artificial intelligence, machine learning, biomarkers, disease detection, clinical chemistry, smart diagnostics, personalized medicine, bioinformatics, health technology integration.

1. Introduction to Biomarkers

A biomarker definition adopted by the US National Institutes of Health's Biomarkers Definitions Working Group in 2001 is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. Thus, a biomarker is any hazard-sensitive, measurable or detectable molecular, biochemical or cellular sign of a process, condition or disease. A measurement that reflects a given biological disease state or some other physiological state can be a biomarker. Investigators use biomarkers to detect or confirm the presence of a disease or condition of interest, to identify individuals with a subtype of the disease, or to provide prognostic information. Biomarkers also are used to predict and monitor clinical response to an intervention and to monitor recurrence or disease progression. For effective translation to the clinic, biomarkers should have the following characteristics: [1]

2. Understanding Metabolomics

Metabolomics pertains to the quantification of metabolites and the mapping of their intricate interactions within cells, tissues, organs, and biological fluids [2]. Defined as the comprehensive, quantitative, and time-operated measurement of all metabolites in an organism or in a biological sample, metabolomics offers a direct signature of biochemical activity. These small molecule metabolites provide a snapshot of the physiological state of cells or tissues—the end result of gene expression, enzyme activity, metabolic reactions, and various biochemical processes—and the overall condition of an organism. Disease-induced alterations in metabolite abundance or composition can therefore serve as effective diagnostic markers for understanding complex diseases such as cancer, cardiovascular disease, and diabetes. The metabolome captures both endogenous processes and the effects of exogenous and environmental exposures; it integrates information from the genome, transcriptome, proteome, microbiome, exposome, diet, and other external factors. While increased computational power has facilitated this complex integration, the diversity of data types and scales introduces numerous technical and conceptual challenges. Metabolomics thus complements genomics and provides a more immediate snapshot of organism health due to the dynamic and tightly regulated nature of metabolites [1].

The principal techniques employed in metabolomics are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) coupled to various chromatography methods. These approaches can identify and quantify hundreds to thousands of metabolites in complex biofluids such as plasma, serum, and urine according to their molecular mass, magnetic properties, and concentration. However, the successful translation of metabolomics to the clinic depends heavily on the selection and application of appropriate statistical and computational tools capable of extracting relevant biological features from voluminous spectral data. Multivariate statistics remain the method of choice for biomarker discovery in -omics research, yet distinct computational methods provide varied, complementary information that is not always fully

exploited. To overcome ongoing interpretative challenges, improvements in both algorithms and machine learning techniques are essential. These advancements will facilitate the conversion of burgeoning metabolomics datasets into clinical applications, furnishing novel insights into disease pathophysiology and enhancing the development of predictive diagnostics. [3][4][5]

3. Role of AI in Clinical Chemistry

AI is designed to mimic human thinking and learning abilities, using large clinical datasets to support clinical decision-making, uncover disease subtypes, and formulate scientific hypotheses. AI applications have been demonstrated in radiology, pathology, ophthalmology, cardiology, and surgery. Enabled by advances in statistical algorithms and abundant computational power, machine learning, a key branch of AI, has the ability to handle the complex, multidimensional data space better than conventional statistical approaches. The development of clinical laboratory automation and high-throughput instruments, laboratory information system management, and electronic medical/health record-keeping, coupled with data standardization initiatives, have resulted in the creation of large medical databases. A systematic analysis, leveraging various clinical testing data in the same database, may unlock new medical knowledge with enhanced diagnostic potential. Clinlabomics analyses the signatures of routine clinical blood and body fluid tests (clinical laboratory data) to identify clinically relevant indexes, develop predictive models, and improve disease diagnosis, risk stratification, and prognosis. It enables the evaluation of the added value of laboratory tests, reduction of unnecessary examination costs, and acceleration of laboratory diagnosis and treatment processes [6].

The main task of the medical laboratory is to measure analytes within the patient´s specimen, so such data can be translated into clinically actionable information. Clinicians are often left responsible for test selection and interpretation, which may lead to over- and underuse of laboratory resources and potential patient risks. Laboratory diagnostic algorithms, such as expert rule systems implementation into the laboratory information system, have been used. The availability of large, structured patient data and increased computational power has led to the development of AI models. AI algorithms can aid in many steps of the laboratory process. Failing to identify the appropriate test contributes significantly to resource misuse. The first AI-based test recommendation system was developed in 2020, using deep learning to analyze laboratory orders from millions of patients with minimal input data. Similar models have been applied to predict impactful laboratory tests and to assess tests used for diagnosing and monitoring conditions like sepsis and renal failure, employing reinforcement learning to forecast patient states and trajectories [7].

4. Integrating Metabolomics with AI

Integrating metabolomics data with artificial intelligence (AI) stands to revolutionize personalized medicine. Large-scale metabolomic datasets, combined with additional layers of patient metadata, empower AI models to predict health risks and disease prognoses [1]. Advanced cloud platforms support storage and processing of metabolomics data on community-wide scales, enabling ensemble computation and epidemiological assessments. Real-time streaming from wearable biosensors generates telemetry data that inform forecasting of complex chronic diseases such as Parkinson’s and rheumatoid arthritis. The US Food and Drug Administration has already cleared an AI-based application for measuring heart ventricle volumes from medical images, accelerating and improving decision-making in heart surgery. Computational methods have evolved in tandem, progressing from univariate to multivariate techniques, linear to nonlinear models, and classical statistics to machine learning algorithms that capture intricate genetic, environmental, and lifestyle interactions underpinning disease [2]. Careful selection of computational approaches is essential, since each can differently impact biomarker-panel performance, thereby influencing clinical outcomes and healthcare costs. A concise overview of both classical and emerging statistical and computational methodologies supports the generation of optimized metabolic-marker panels.

Metabolomics promises detailed new insights and superior biomarkers for complex diseases. Integrating metabolomics data within AI and machine-learning frameworks enables quantitative prediction of health risks and prognosis on an individual basis. Advancing beyond traditional trial-and-error paradigms, this integrated approach paves the way for improved healthcare outcomes.

5. Current Trends in Biomarker Research

The discovery of thousands of potential biomarkers has expanded their definition and application across medical disciplines. Despite substantial progress, the identification of accessible biomarkers that enable earlier disease detection and improved clinical decisions remains a critical research challenge. Biomarkers derived from diverse molecular species provide comprehensive information exceeding the discriminatory power of clinical parameters like blood glucose, cholesterol, or PSA. The advent of metabolomics, the unbiased analysis of the complete repertoire of metabolites, offers new opportunities for biomarker discovery. This approach integrates analytical technologies such as gas- and liquid-chromatography, ^1H nuclear magnetic resonance profiling, and mass spectrometry, thereby capturing the complete molecular composition of a biological system in an unbiased manner. Recent years have witnessed the successful application of artificial intelligence in clinical practice, including diagnoses, translational research, and drug development. Artificial intelligence derives its utility from data analysis algorithms that establish associations between variables, irrespective of the underlying processes. In the context of metabolomics, these techniques focus on metabolite associations without presuming a metabolic or biological connection. This capability opens an additional analytical layer by providing complementary statistical information and intensifying biological interpretation [8]. The stable and individual-specific nature of plasma protein levels justifies the pursuit of individual-based diagnostics over population-based cutoff values, particularly in the presence of prevalent co-morbidities. Comprehensive plasma proteomic profiles can simultaneously screen for multiple pathologies, presenting a more efficient diagnostic strategy. Nonetheless, the breadth of information may lead to incidental findings, supporting the preference for targeted diagnostic approaches in certain cases. The complexity of resultant multi-dimensional data necessitates advanced analytical methods. The consolidation of multiple biomarkers into quantitative panels, augmented by clinical metadata, holds promise for enhancing interpretative efficacy. Progress in deep learning and big data methodologies further augments the potential to uncover potent novel associations. The disciplines of clinical biochemistry and translational research have experienced unprecedented growth, yet the translation of metabolomic biomarkers into clinical practice remains limited. Predominantly, studies to date comprise small-scale, preliminary investigations with inadequate experimental design, undermining the reliability of proposed biomarkers. To facilitate the transition towards point-of-care diagnostic applicability, emphasis must shift towards extensive cohort, multi-center studies for biomarker identification. The recommended discovery trajectory commences with initial comprehensive profiling in modestly sized case-control cohorts (~100 samples), proceeds to validation in independent cohorts, preferentially from divergent geographic locales, and culminates in an analytical development phase aimed at deploying biomarkers via cost-effective technologies. Contemporary investigations target a spectrum of diseases including diabetes, chronic pancreatitis, Alzheimer’s disease, chronic kidney disease, diastolic heart failure, bovine tuberculosis, bladder cancer, and colorectal cancer. Despite the breadth of research, data interpretation often remains convoluted, underscoring the necessity for methodological refinement [1].

6. Technological Advances in Metabolomics

Advances in analytical instrumentation have allowed for high-throughput metabolite identification and quantitation, but further improvements in detection sensitivity, resolution, and dynamic range are still needed to capture subtle metabolomic differences. Emerging tools such as ion mobility coupled with mass spectrometry (IM-MS) provide promising avenues to improve

resolving power and separation efficiency, thus enabling more comprehensive coverage of the metabolome. Additionally, enhanced chromatographic techniques and alternative derivatization methods continue to expand the accessible spectrum of metabolites, including difficult-to-analyze compounds. Another important development in metabolomics is the integration of artificial intelligence (AI) and machine learning for data analysis. A range of classifiers, including partial least squares discriminant analysis (PLS-DA), support vector machines (SVM), random forest (RF), and k-nearest neighbours (kNN), have been adopted to analyze complex metabolomic datasets and discover potential biomarker panels [1] [2]. These advanced computational approaches facilitate more refined interpretation of high-dimensional information, streamlining the biomarker discovery pipeline.

7. AI Algorithms in Biomarker Discovery

Machine-learning approaches facilitate the analysis of complex data sets and the elucidation of relationships between variables, thereby promoting the versatility of metabolomics. Supervised methods have therefore been applied in the identification of metabolomics-based biomarkers. Next-generation biomarkers follow the traditional ones in an evolving sequence based on their potential for delivery via POC platforms. Because they tend to be more complex, integration with bioinformatics and AI is particularly important. Several types of AI-guided biomarkers exist. While targeted profiling examines only a set of pre-identified metabolites for predictive classification, non-targeted profiling requires more data-preprocessing and is better suited to identify previously unknown minimum sets of discriminatory biomarkers. Hybrid approaches combining both methods, supported by AI, can also be used [9]. Artificial intelligence combined with machine learning can also be integrated with conventional biomarkers for a multi-omics approach or even applied directly to raw biosignals without feature extraction.

Various applications of AI in the field illustrate how this approach can contribute to the identification of predictive biomarker relationships. For example, the development of a neural-network ensemble for the diagnosis of cardiovascular diseases (CVDs) entailed a combination of traditional statistics and advanced machine-learning techniques applied to an extensive transcriptomic data set. Additionally, a recursive feature-elimination classifier ranked the features based on their importance. Subsequently, a set of 18 transcriptomic biomarkers was identified that accurately distinguished CVD patients from healthy patients with up to 96% predictive accuracy [10]. The determined biomarkers may thus serve as potential indicators of CVDs or patients at risk of developing these conditions.

Medical-pathology analyses illustrate another promising approach for the integration of AI with biomarkers. Classic tissue biomarkers provide diagnostic, prognostic, and predictive information in personalized medicine and include such factors as tumor histotype, grade, and stage, as well as the expression of hormone receptors and proliferation markers. AI models applied to histological images of breast tumors have demonstrated the ability to distinguish benign from malignant lesions, differentiate in situ from invasive cancer, and support tumor grading, yet are especially appealing for mitosis detection. Similarly, AI has been used to classify histologic subtypes and quantify the expression of common biomarkers such as ER, PR, and HER2, correlating well with manual assessments [11]. A further instance is provided by a metabolomics-based discrimination of patients with colorectal cancer and those exhibiting advanced adenomas using the profiles of urinary volatile organic compounds (VOCs). Seventeen relevant ions were selected via a random forest classifier, and subsequent model training using an artificial neural network yielded predictive accuracies of 87.9% (healthy controls) and 67.8% (advanced adenomas) when validated in blind-sample tests. Machine learning might thus improve the specificity of VOC-based screening and enable the detection of a disease stage preceding malignancy, such as adenomatous polyposis.

8. Case Studies in Disease Detection

Detecting cancers in their early stages can be difficult because initial symptoms often overlap

with those of a cold or flu. More-accurate tools for early disease detection could give patients a better chance of surviving illness [12]. Biomarkers are widely used in screening for various diseases. For example, prostate-specific antigen (PSA) is one of the most common screening approaches for prostate cancer [13]. Metabolic biomarkers can also play an important role in the early diagnosis of disease because changes in metabolite concentrations often precede symptoms. Many of these metabolites may be present at low concentrations below the detection limits of conventional biochemical analysis methods [14]. Parallel measurements of multiple metabolites coupled with sophisticated machine-learning algorithms enable greater insight and accuracy and also can identify the best predictive subset of biomarkers. Parallel mass spectrometry can achieve the widespread detection of metabolites necessary for machine learning but requires sophisticated instruments with long acquisition times. A great deal of sample preparation is also necessary to detect low-concentration metabolites in complex samples such as blood plasma.

A miniaturized metabolomics platform can provide rapid and inexpensive initial cancer screening with a single drop of blood followed by more-reliable mass spectrometry for positive cases. The approach combines metabolomic analysis with machine-learning algorithms in a miniaturized 1-cm2 CMOS device integrating highly multiplexed sensors, integrated data processing, and on-chip classification. Four metabolites compose the provisional biomarker panel: l-histidine, proline, arginine, and lysine. The sensor platform targets lipoate-protein ligase A (LplA), which plays an important role in prostate cancer metabolism. The random-forest machine-learning model computed from mass-spectrometry measurements of these amino acid metabolites on 72- mL plasma volumes discriminates disease-free patients from those with malignant prostate cancer with an area-under-the-curve (AUC) value of 0.78, exceeding PSA (AUC ¼ 0.68) by approximately 10%. The new test demonstrated a sensitivity of 32% at a high specificity of 87%, but the sensitivity increased further to 94% for a more-typical specificity level of 70%. Additional examination of the model's size-versus-sensitivity dependence showed that the performance of the four-metabolite test could be further improved through the inclusion of a larger number of metabolites; larger clinical trials are needed to validate the platform before clinical adoption. [15][16][17]

8.1. Cancer Detection

More than 1,400 cancer-related genes have been identified to date, with their numbers continuing to increase [14]. However, many of these genes remain difficult to track for diagnosis through metabolomics. Nonetheless, multiple studies have recognized metabolic pathways and specific metabolites that regulate tumor progression [13] , enabling distinctions among tumor stage, histological type and drug response. Changes in metabolite patterns have been used to evaluate the clinical characteristics of colorectal, ovarian, renal, oral and pancreatic tumors. More specific and sensitive biomarkers are still sought in the case of lung cancer, the leading cause of cancer mortality worldwide. Artificial intelligencee (AI), in particular machine learning, can manage very large data sets efficiently. It has been applied to develop survival and prognostic prediction models for multiple cancer types employing different algorithms. Machine learning models can learn from existing data, iteratively improve performance and identify the most effective variables, sometimes with accuracy comparable to human experts. The increasing adoption of AI in lung cancer research holds great promise in addressing the limitations of current diagnostic and prognostic approaches.

8.2. Cardiovascular Diseases

Cardiovascular diseases (CVDs) encompass stroke, congenital heart disease, rhythm disorders, atherosclerosis, coronary heart disease, heart failure, valvular heart disease, venous disease, and peripheral artery disease, collectively representing the leading cause of death worldwide and imposing substantial health and economic burdens. Prevention requires maintenance of healthful lifestyles and monitoring of key risk factors such as inflammation and diabetes. Early diagnosis

is pivotal. Acute coronary syndrome (ACS) must be diagnosed at presentation using a combination of clinical history, physical examination, ECG changes, and cardiac biomarkers. Biomarker detection in body fluids—blood, urine, saliva, sweat, cerebrospinal fluid, breath, tears, interstitial fluid—can enhance diagnosis, inform prognosis, and facilitate therapy selection. Ideal biomarkers exhibit high sensitivity, specificity, and cost-effectiveness. Detection strategies fall into two categories: knowledge-based, which leverage biological understanding to improve existing assays (ELISA, immunofluorescence, Raman scattering, SPR, SERS) and develop antibody-based sensors; and unbiased, which simultaneously profile large assemblies of molecules to uncover new diagnostic fingerprints. Optical sensing platforms can service both approaches [18]. Recent advances in genomics and bioinformatics have deepened insights into the complex etiology of CVDs, yet AI/machine learning (ML) models employing genetic biomarkers remain nascent. Prior models, trained on genomic and exomic data, successfully predict incident heart failure and atrial fibrillation but primarily interrogate genes already associated with CVDs. A new AI/ML framework addresses this limitation by training on transcriptome-based gene expression and clinical data to identify a set of 238 statistically significant biomarkers, both known and novel. Algorithms such as Recursive Feature Elimination, Pearson Correlation, Chi-Square Test, and ANOVA guide marker selection. The resulting model performs 10-class classification encompassing seven distinct CVD types (heart failure, aortic valve stenosis, atrial fibrillation, myocardial infarction, atherosclerosis, mitral valve disease, pulmonary hypertension), bypassing conventional binary-comparison frameworks. Transcriptomic biomarkers can support accurate, high-resolution prediction of incident CVD cases, enabling early detection of high-risk individuals prior to symptom onset and guiding timely medical intervention [10].

8.3. Neurodegenerative Disorders

Dementia refers to a progressive cognitive decline affecting multiple domains that interferes with daily life, arising from heterogeneous pathologies, among which neurodegenerative diseases are prevalent. The most common neurodegenerative dementias are Alzheimer's disease (AD), Parkinson's disease with and without dementia (PDD and PD, respectively), Lewy body dementia (LBD), and frontotemporal dementia (FTD). Although the core symptoms vary, the similarities in cognitive changes and phenotypic overlap of symptoms among these conditions complicate diagnosis and treatment.

AD is the most common type of neurodegenerative dementia, characterized by extracellular amyloid-β and intracellular tau pathology. The disease leads to progressive neurodegeneration, manifesting in memory loss, language impairment, apathy, and impaired executive functions [19]. AD has a preclinical phase with brain pathology but no symptoms and a prodromal phase known as mild cognitive impairment (MCI), during which cognitive changes are detectable but do not yet induce everyday life compromises. A definitive AD diagnosis requires histopathological analysis; therefore, biomarkers are essential for diagnosis during life and clinical trials. Cerebrospinal fluid (CSF) concentrations of amyloid-β 42, phosphorylated tau, and total tau, alongside image-based findings, provide such diagnostic tools. However, distinguishing AD from other neurodegenerative diseases at the MCI or early dementia stages remains challenging, and measures to monitor disease progression are limited. Quantified metabolic profiles could enhance diagnostics by improving stage recognition.

9. Challenges in Integrating AI and Metabolomics

Outstanding challenges remain in the integration of artificial intelligence and metabolomics data for medical diagnosis and prognostication. Translating multi-omic data into medical decisions poses a complex task. Determining suitable AI and machine learning approaches for specific diagnostic tasks is challenging. AI methods are required that identify relevant variables for patient stratification, detect latent statistical patterns, and uncover modifiable psychobiological risk factors. These procedures tend to be automated when analyzing clinical data, while

genotype-phenotype associations need careful evaluation to avoid false discoveries. In the context of precision medicine, genomic information remains a poor predictor of disease without extensive data collected during an individual's lifespan. Large metabolic datasets and imaging studies hold promise in predicting individual disease susceptibility. Metabolomics profiling, combined with AI, offers tools for disease prediction through analysis of biochemical pathways and metabolite distributions. When metabolomics for the masses is achieved, data processing could rely on substantial computing resources and secure storage within advanced cloud environments. Measurements from disparate populations may be linked via the internet, facilitating ensemble computation and epidemiological assessments. Wearable biosensors transmit data directly to cloud services, enabling health forecasts based on telemetry analyses. Examples include the mPower Parkinson's Disease study and an app monitoring weather-related pain in rheumatoid arthritis patients. AI applications extend to screening, decision-making, and management; a recent FDA-approved AI-based tool measures heart ventricular volumes to assist clinical decisions. The integration of metabolomics with AI and machine learning presents significant potential to advance personalized medicine by predicting individual health risks, identifying biomarkers, and correlating data with health outcomes. This approach offers possibilities beyond the constraints of traditional evidence-based methods. A future vision entails patient-centered diagnostics in which metabolomics data integrate alongside other omic information, contingent upon overcoming current economic and usability barriers [1] [6] [20].

10. Ethical Considerations in Biomarker Research

Ethical concerns about biomarkers include confidentiality, privacy, disclosure, informed consent, conflicts of interest, and exploitation of vulnerable populations. Researchers must anticipate potential social, legal, and financial consequences for study participants and decide what to share or withhold, while trusting that biomarker data remain confidential [21].

11. Regulatory Framework for Biomarker Development

Biomarker measurements support key decisions throughout drug development, including regulatory approvals. They help document exposure-response relationships, specificity, potency, side effects, and therapeutic applications [22]. The regulatory landscape governing pharmacodynamic assays and instrumental technologies emphasizes that the usefulness of a biomarker is limited by its specificity and sensitivity and also by its inherent biological and method-induced variability. Controlling sources of variability such as reagent purity, pipetting accuracy, and antibody specificity is essential to ensure consistently high-quality results. Quality standards for assays supporting nonclinical safety studies fall under GLP regulations, while those used in support of diagnostics adhere to CLIA regulations and accreditation standards. Although many biomarker applications in research are not regulated, laboratories adopt practices to ensure high-quality data, anticipating the inevitable extension of these demands as scientific and regulatory authorities develop standards and guidelines. The rise of multiplexed assay platforms driven by genomics is expected to shape the future of biomarker laboratories.

As biomarker methods become more sensitive, flexible, and capable of high-content readout, the increased emphasis on bioimaging and laboratory measurements is transforming pharmaceutical research and regulatory expectations. Advanced biomarker data support improved decision-making, reduce the risk of project failure, provide a richer body of safety and efficacy data, and promote a more adaptive approach to emerging market and product needs.

12. Future Directions in Clinical Chemistry

The clinical laboratory, which has experienced over a century of operational oversight, is currently poised to exploit the information contained within the vast arrays of laboratory test results generated daily. Data-mining and machine learning approaches have been applied successfully to the laboratory test result databases from different hospital laboratory information systems (LIS), both to explore the relationship between laboratory tests and diseases and to

develop decision support tools that narrow diagnostic possibilities for emerging illnesses [6]. The extension of these approaches to include multiomics techniques, proteomics, and metabolomics is particularly exciting, as these may yield unique early-warning signs of altered metabolism that could presage disease.

12.1. Personalized Medicine

A plethora of clinically relevant data are generated in the biomedical area from scanning and imaging techniques, patient and literature databases, accumulation of 'omics' data, and the monitoring of patients through wireless devices, offering new opportunities for how this rich clinical data could be collated, shared, modelled, analysed, and mined to redefine the diagnosis, prognosis, and prevention of diseases. However, the availability of this mass of dispersed data creates a significant opportunity yet also a bottleneck for the clinical progression trend towards individualized care; the medical community, as well as clinical and pharmaceutical organisations, recognise that the rate of data acquisition far outweighs the development of new computational tools for their management, especially at a systems level [20]. Artificial Intelligence (AI), which has been an active field of research for more than 60 years, can play an important role in this systems-level aspect. Indeed, the ability of AI to automatically identify meaningful and diagnostic patterns that provide insights about a patient's specific condition offers a great promise for improved patient care. Pattern recognition of multi-omics data in conjunction with a large collection of sample-matched clinical data, and the integration of diverse sources of data into one interoperable infrastructure, has the potential to significantly impact patient care from routine clinical to pharmaceutical applications and more generally facilitate the excellent research already underway.

The ability to perform a complete characterization of the patient's molecular abundance profile underpins the concept of personalised medicine, which is not confined to the treatment of the symptom but also to an individual molecular risk assessment, thereby addressing the preventative medicine approach [1]. Metabolomics already shows considerable promise in providing new early stage disease signatures, and the advent of wearable chemical sensors will in the future facilitate frequent monitoring of individuals, generating large volumes of telemetry data. Augmented with appropriate contextual data, such frequent monitoring by analysis of biomarkers or chemical signatures in biofluids facilitates improved differentiation between disease states through anomaly detection; and from a personalised metabolomics perspective such datasets enable an individual molecular profile to be constructed. These spectral 'data-fingerprints' can be interpreted through statistical analysis and pattern recognition targeted towards classification or discovery of unique patterns of interest, or through spectral deconvolution enabling mapping onto biochemical networks.

12.2. Population Health

Metabolomics for the masses may require large computing power and secure data storage within advanced cloud environments. Data from different populations could be linked via the internet, enabling epidemiological assessments of disease progression and spread. Wearable biosensors that sync data to cloud services can produce telemetry data useful for predicting health risks, such as studies on Parkinson's disease and rheumatoid arthritis. AI has been approved for clinical use, such as measuring ventricle volume to aid decision-making in heart surgeries. There is significant potential for integrating metabolomics with AI and machine learning to improve personalized medicine, predict health risks, and link information to dynamic patient metadata for better disease prognosis. This approach aims to surpass traditional evidence-based medicine, focusing on tailored healthcare outcomes [1].

13. Collaboration Between Disciplines

Collaboration across disciplinary boundaries has become essential for the growth of knowledge and the development of new information and technologies across all fields. For example,

chemists, biologists, and computer scientists must work together to develop electronic sensors for chemical detection [1]. Each needs a working knowledge of what the others can and cannot do so that a project can be directed along successful lines immediately, minimizing the amount of trial and error along the way.

Although metabolomics relies heavily on analytical chemistry—exactly what the Organic Chemistry and Analytical Chemistry courses already deliver—it also requires a much wider knowledge of biology, biochemistry, data handling, multivariate statistics, and chemometric techniques, something that the present-day courses in the first-year chemistry curricula do not provide [6]. A knowledge of the other "Omic" approaches and their strengths and weaknesses is also helpful for metabolomic applications.

The integration of ancillary datasets may eventually become the "standard" in the biomarker field as an avenue for many levels of assessment, including exposure and trace–disease correlations—an immediate relevance to the microbiological and molecular profiling techniques used throughout the agricultural sector concerned with food safety. The accepted approach for simultaneous analysis of large datasets is "data blending," where various data channels (x-blocks) are fused into a single data frame (of x-variables); the samples (individuals) have to be the same in each block. In scenarios where such a prerequisite cannot be met, other integrative methods such as Consensus or Parallel Orthogonalized Partial Least Squares (POPLS) may be adopted. Employing these techniques with Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) based techniques for data fusion, one can then identify the origin of bio-active effects and reveal the underlying trends not evident from the assessment of any one dataset.

14. Data Management and Analysis in Metabolomics

Metabolomics is the study and quantification of metabolites and their interactions, providing a snapshot of an organism's physiological state [2]. By comparing metabolic profiles of diseased and healthy subjects, disease-driven changes can be detected, which assists both diagnosis and the discovery of perturbed metabolic networks. Metabolomics also interprets the signatures from a variety of metabolites to represent multiple factors influencing disease, including food consumption, environmental exposures, and gut microbiota, by which it provides a data-driven approach for understanding gene–environment interactions. With the advent of nuclear magnetic resonance, GC–MS/LC–MS, and other techniques, more than a few hundred metabolites in a large number of biological samples can be measured within a day. Although metabolomics technology has advanced well, the translation of metabolomics data into clinical practice requires effective statistical and computational methods. Consequently, a large fraction of the contemporary metabolomics studies have not been performed with optimal analysis methods. The increasing size and complexity of metabolomics datasets render the aforementioned traditional methods less effective. Powerful heuristic and machine learning approaches—such as ensemble learning, deep learning, and genetic programming—have thus become crucial for making sense of metabolomics profiles.

15. Implementation of AI in Clinical Settings

The integration of artificial intelligence (AI) in clinical laboratory settings is emerging as a means to transform extensive routine laboratory data into critical insights, assisting clinicians in decision-making. The analysis of complex, multi-parameter data by advanced AI algorithms complements expert interpretation for improved early diagnosis [1]. Models that extract informative clinical attributes contribute to diagnostic and prognostic assessment, aiding in the adoption of appropriate clinical guidelines and the control of operational costs [7]. AI approaches also improve the prediction of metabolic patterns associated with specific clinical phenotypes, incorporating parameters related to laboratory measurements, instrumentation, and environmental factors to highlight the potential of the developing field of clinlabomics [6].

16. Patient Perspectives on Biomarker Testing

Access to diagnostic services is a fundamental right; however, the mental strain of long waits for diagnoses can be overwhelming. Near-real-time test results streamline treatment, reducing anxiety [12], yet rapid result delivery is not widespread. Technological hurdles, adoption costs, limited infrastructure, equipment size, and staffing shortages remain significant barriers. In vitro diagnostics combined with artificial intelligence have the potential to advance disease assessment, outperforming conventional methods while offering scalability, high-quality output, and affordability. Smart Diagnostics unifies a universal microfluidic biosensor platform, AI analysis software, and electronic health record integration. The Programmable Bio-Nano-Chip employs a compact analyzer paired with adaptable assay cartridges to translate signals into disease scores for direct consumption by patients or healthcare providers. Applications span oral and ovarian cancers, prostatic malignancy, cardiovascular disease, trauma, substance abuse, and coronavirus severity, enabling diagnostics that are rapid, straightforward, and require minimal operator training. Ease of use facilitates deployment in resource-limited environments and supports longitudinal monitoring of disease progression and therapeutic efficacy. Beyond plasma biomarkers, the approach integrates clinical notes, radiographic imagery, sensor data, and genomic profiles, augmenting the breadth and depth of clinical care available to patients. [23][24][25]

17. Conclusion

The integration of innovative omics-sequencing technologies into clinical and research targets has expanded the scope of precision medicine . Metabolomics, positioned at the downstream end of the omics cascade, provides complementary insight into the pathophysiological processes and responses to health challenges across a comparative timescale . When applied in concert with other omics-sequencing approaches, the complementary insights afforded by metabolomics offer a more detailed picture of molecular perturbations potentially leading to disease development or progression. Current studies increasingly integrate metabolomics with other omics avenues and harness artificial intelligence (AI) for data processing, aiming to improve personalised healthcare outcomes and diagnostics.

References:

1. D. K. Trivedi, K. A. Hollywood, and R. Goodacre, "Metabolomics for the masses: The future of metabolomics in a personalized world," 2017. ncbi.nlm.nih.gov
2. M. Y. Lee and T. Hu, "Computational Methods for the Discovery of Metabolic Markers of Complex Traits," 2019. ncbi.nlm.nih.gov
3. R. Tyagi, P. Kumar, and U. Sharma, "Metabolomics techniques: A brief update," Epigenetics and metabolomics, 2021. [HTML]
4. M. P. M. Letertre, P. Giraudeau, and P. De Tullio, "Nuclear magnetic resonance spectroscopy in clinical metabolomics and personalized medicine: current challenges and perspectives," Frontiers in Molecular, vol. XX, no. XX, pp. XX-XX, 2021. frontiersin.org
5. Z. S. A. Zahra, "Nuclear magnetic resonance (NMR): Principle, Applications, types, and uses in Metabolite Identification and Medical Biotechnology," Current Clinical and Medical Education, 2024. visionpublisher.info
6. X. Wen, P. Leng, J. Wang, G. Yang et al., "Clinlabomics: leveraging clinical laboratory data by data mining strategies," 2022. ncbi.nlm.nih.gov
7. J. Cadamuro, "Rise of the Machines: The Inevitable Evolution of Medicine and Medical Laboratories Intertwining with Artificial Intelligence—A Narrative Review," 2021. ncbi.nlm.nih.gov

8. P. E Geyer, L. M Holdt, D. Teupser, and M. Mann, "Revisiting biomarker discovery by plasma proteomics," 2017. ncbi.nlm.nih.gov
9. J. Tan, F. Qin, and J. Yuan, "Current applications of artificial intelligence combined with urine detection in disease diagnosis and treatment," 2021. ncbi.nlm.nih.gov
10. W. DeGroat, H. Abdelhalim, K. Patel, D. Mendhe et al., "Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine," 2024. ncbi.nlm.nih.gov
11. C. Lancellotti, P. Cancian, V. Savevski, S. Rupa Reddy Kotha et al., "Artificial Intelligence & Tissue Biomarkers: Advantages, Risks and Perspectives for Pathology," 2021. ncbi.nlm.nih.gov
12. M. P. McRae, K. S. Rajsri, T. M. Alcorn, and J. T. McDevitt, "Smart Diagnostics: Combining Artificial Intelligence and In Vitro Diagnostics," 2022. ncbi.nlm.nih.gov
13. V. F. Annese, S. B. Patil, C. Hu, C. Giagkoulovits et al., "A monolithic single-chip point-of-care platform for metabolomic prostate cancer detection," 2021. ncbi.nlm.nih.gov
14. Y. Xie, W. Y. Meng, R. Z. Li, Y. W. Wang et al., "Early lung cancer diagnostic biomarker discovery by machine learning methods," 2020. ncbi.nlm.nih.gov
15. A. J. Grooms and B. J. Burris, "Mass spectrometry for metabolomics analysis: Applications in neonatal and cancer screening," Mass spectrometry, vol. 2024, Wiley Online Library. nih.gov
16. F. Danzi, R. Pacchiana, A. Mafficini, M. T. Scupoli, "To metabolomics and beyond: a technological portfolio to investigate cancer metabolism," *Nature Reviews Clinical Oncology*, vol. 2023. nature.com
17. V. F. Annese, S. B. Patil, C. Hu, C. Giagkoulovits, et al., "A monolithic single-chip point-of-care platform for metabolomic prostate cancer detection," Microsystems & Nanoengineering, vol. 7, no. 1, 2021. nature.com
18. R. V. John, T. Devasiya, N. V.R., S. Adigal et al., "Cardiovascular biomarkers in body fluids: progress and prospects in optical sensors," 2022. ncbi.nlm.nih.gov
19. O. Jääskeläinen, A. Hall, M. Tiainen, M. van Gils et al., "Metabolic Profiles Help Discriminate Mild Cognitive Impairment from Dementia Stage in Alzheimer’s Disease," 2020. ncbi.nlm.nih.gov
20. Z. Ahmed, "Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis," 2020. ncbi.nlm.nih.gov
21. C. M. Erickson, L. R. Clark, F. B. Ketchum, N. A. Chin et al., "Implications of preclinical Alzheimer's disease biomarker disclosure for US policy and society," 2022. ncbi.nlm.nih.gov
22. B. N. Swanson, "Delivery of High-Quality Biomarker Assays," 2002. ncbi.nlm.nih.gov
23. M. P. McRae, K. S. Rajsri, T. M. Alcorn, and J. T. McDevitt, "Smart diagnostics: combining artificial intelligence and in vitro diagnostics," Sensors, 2022. mdpi.com
24. N. Ghaffar Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," Discover Artificial Intelligence, 2023. springer.com
25. H. Shi, A. Kowalczewski, D. Vu, X. Liu, A. Salekin, "Organoid intelligence: Integration of organoid technology and artificial intelligence in the new era of in vitro models," Medicine in Novel Technologies and Devices, vol. 2024, Elsevier. sciencedirect.com