Nutrition & Food Science
International Journal
ISSN 2474-767X

Juniper Publishers
key to the Researchers

Review Article
Volume 15 Issue 4 - July 2026
DOI: 10.19080/NFSIJ.2026.15.555918

Nutri Food Sci Int J

AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics

Songyu Zhang[1,3], Shanbin Chen[1], Jia Zheng[2], Wei Liu[1], Jian Su[2], Jie Meng[1], Deliang Wang[1], and Chunguang Luan[1]*

[1]*China National Research Institute of Food and Fermentation Industries, Beijing 100015, China*

[2]*Wuliangye Yibin Cooperation Limited. Yibin, Sichuan 644000, China*

[3]*China Agricultural University, Beijing 100083, China*

Submission: June 24, 2026; **Published:** July 30, 2026

***Corresponding author:** Chunguang Luan, China National Research Institute of Food and Fermentation Industries, Beijing 100015, China

Abstract

In recent years, the rapid development of artificial intelligence (AI) technology has provided a powerful tool for addressing the bottleneck in correlating microbial phenotypes with metabolites. By integrating multi-omics data, key process parameters, and high-frequency, real-time sensing data, AI can construct high-precision mathematical models to enable in-depth analysis of the operational status of microbial cell factories. Specific models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are particularly adept at analyzing complex datasets. Additionally, algorithms such as random forests and deep learning architectures can discover hidden patterns of phenotype-product correlations in large datasets, enabling dynamic simulation and flux prediction of metabolic pathways. On this basis, AI models can achieve intelligent optimization from "experience-driven" to "data-driven + model-guided". AI has improvement and enhanced fermentation efficiency, reduced costs, accelerated strain improvement, and improved accuracy and control across the industry. However, challenges remain, including obtaining high-quality data, integrating data from multiple sources, ensuring that models perform well in novel settings, and trusting algorithms in real-world settings. In the future, advances such as self-supervised, transfer, and reinforcement learning, along with improved sensors and automation, are likely to further enhance the benefits of AI in microbial fermentation. These changes will facilitate the transition from laboratory predictions to real-world industrial use, leading to more robust and efficient fermentation systems.

Keywords: Phenotype; Metabolites; Artificial; Intelligence; Fermentation; Dynamics; nutritional limitations.

Abbreviations: AI: Artificial intelligent; ARIMA: Autoregressive Integrated Moving Average; CER: Carbon dioxide release rate; CNNs: convolutional neural networks; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; DO: Dissolved oxygen; FP: Fermentation percent; FPPM: Fermentation process prediction model; GCNs: Graph convolutional networks; GEMs: Genome-scale metabolic models; GRU: Gated Recurrent Unit; HPLC: High-Performance Liquid Chromatography; LSTM: Long Short-Term Memory; MS: Mass Spectrometry; NMR: Nuclear Magnetic Resonance; OUR: Oxygen Uptake Rate; PID: Proportional–Integral–Derivative; PLS: Partial Least Squares Regression; RF: Random Forest; PSSMs: Position-Specific Scoring Matrices; RNNs: Recurrent Neural Networks; SVR: Support Vector Regression; SVMs: Support Vector Machines; VAE: Variational Autoencoder.

Introduction

Microbial fermentation is central to sustainable manufacturing, converting organic substrates into enzymes, organic acids, and biofuels while reducing environmental footprints [1]. Its advantages—non food biomass use, mild conditions, and biorefinery integration-lower energy consumption and greenhouse gas emissions by over 50% relative to fossil-based routes [2,3], with synthetic biology extending applications to complex molecules aligned with UN Sustainable Development Goals [4]. However, optimization is hindered by multi scale complexity [5], "black box" metabolic networks [6,7], and numerous coupled variables that drive up costs and uncertainty [8,9]. Suboptimal control can cause losses more than 20%, wasting resources and increasing environmental impact [10].

Addressing these issues requires linking microbial phenotypes—metabolic flux, stress responses, transcriptomes, and proteomes—to product synthesis [11]. Phenotypes act as the "driving engine" for target compounds, requiring specific induction to shift from biomass accumulation to antibiotic or organic acid production [12,13]. Moreover, in biofuels, phenotypic state governs substrate conversion, guiding metabolic engineering

to raise yields [14]. Transcriptomic and proteomic analyses reveal regulatory points, enabling scientific strategies over trial and error [15,16]. High throughput, multi omics approaches allow precise phenotype characterization, with CRISPR editing and feedback loops enhancing reliability and feasibility [17-20]. AI shifts fermentation from empirical "black boxes" to interpretable, predictive models. Machine learning and deep learning handle high dimensional, nonlinear, and time series bioprocess data, with RNNs and LSTMs capturing temporal dynamics beyond ARIMA or PID limits [21-25]. Following successes in drug discovery and genomics [26,27], the field adopts deep learning for cell morphology monitoring and reinforcement learning for adaptive control [28]. Integrating multi omics with AI driven digital twins overcomes traditional bottlenecks, enabling dynamic regulation for next generation sustainable biomanufacturing [17,29] (Figure 1).

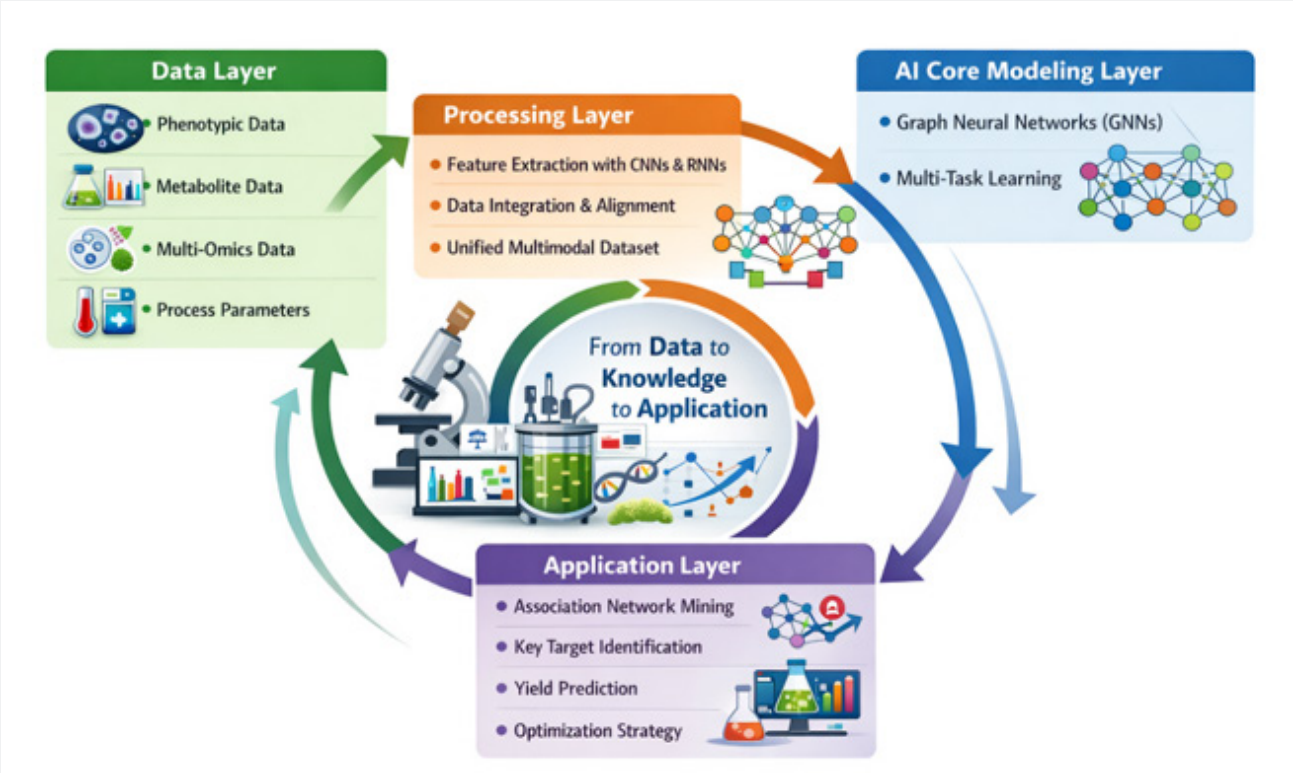

Figure 1: Framework illustrating AI-driven analysis of correlations between microbial phenotypes and metabolite profiles.

Data Foundation: Acquisition and Integration of Multi-source Heterogeneous Data

Process parameter data

Process parameter data plays a fundamental role in AI-driven microbial fermentation, providing critical inputs for analyzing the dynamic correlation between microbial phenotypes and metabolites. These data are typically divided into online sensor data and offline detection data, with the former enabling real-time monitoring and the latter supplementing detailed biochemical information. Together, they form a multi-source heterogeneous dataset that supports the training and optimization of machine learning models. By integrating these data, researchers can more accurately simulate fermentation kinetics and predict metabolic pathway changes, thereby optimizing process control strategies [30].

Online sensor data provides real-time dynamic monitoring capabilities: Online sensor data, including pH, temperature, dissolved oxygen (DO), tank pressure, and exhaust gas analysis parameters such as oxygen uptake rate (OER) and carbon dioxide release rate (CER), are the core of real-time monitoring in fermentation processes. Embedded sensors continuously collect these parameters and reflect the instantaneous state of microbial metabolic activity. The pH changes can indicate substrate consumption or product accumulation, while DO levels correlate with the intensity of cellular respiration. In AI applications, these real-time data are used to train dynamic models, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, to predict phenotypic evolution and metabolite fluctuations. However, sensor data are often affected by noise interference and calibration deviations, which may lead to model overfitting or prediction errors. Therefore, data preprocessing

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

techniques such as filtering and outlier detection are necessary to improve data quality [22]. Thus, online sensor data serve as the foundational input for AI-driven fermentation processes, with their high-frequency acquisition enabling real-time feedback control. However, algorithmic optimization is required to address data heterogeneity.

Offline detection data supplements key biochemical information: Offline detection data, such as measurements of substrate concentration, biomass, and product concentration, are obtained through periodic sampling and laboratory analysis, providing high-precision, low-frequency biochemical parameters for the fermentation process. These data are typically acquired using methods such as high-performance liquid chromatography (HPLC), mass spectrometry, or cell counting, enabling precise quantification of metabolite levels and cell growth dynamics and thereby revealing deep correlations between phenotypes and product formation. In AI models, offline data is often used to calibrate online sensor outputs and serve as labels for supervised learning, enabling the training of predictive models to infer metabolic states at unmeasured time points. The challenge lies in the delayed and destructive nature of offline data acquisition, which may lead to discontinuous time series and degrade model real-time performance; additionally, integrating multi-source data requires addressing scale inconsistencies, such as aligning offline and online data through fusion algorithms [31]. Thus, offline detection data is an indispensable component of fermentation process optimization, enhancing the reliability and interpretability of AI models by providing ground-truth values. However, it requires advanced data assimilation techniques to integrate seamlessly with real-time data.

Overall, the multi-source, heterogeneous nature of process parameter data provides a rich information layer for AI-driven fermentation research. However, effective integration through intelligent algorithms is required to analyze the complex dynamics of microbial phenotypes and metabolites. In the future, combining edge computing and IoT technology could further enhance the real-time and accuracy of data acquisition [30].

Microbial Phenotype Data ("Multiomics" Revolution)

In AI-driven microbial fermentation processes, analyzing the dynamic correlation between microbial phenotypes and metabolites requires integrating multiple omics data [31]. The multi-omics revolution captures comprehensive information from the genome to metabolic fluxes using high-throughput technologies, providing an unprecedented perspective on the adaptability, metabolic regulation, and product synthesis of microorganisms in fermentation environments. Traditional single-omics methods often fail to capture the complexity of phenotypic changes fully. In contrast, multi-omics integration, combined with AI algorithms, can extract patterns, predict behavior, and optimize fermentation processes using large datasets [32]. For example, machine learning models can integrate genomic, transcriptomic, proteomic, metabolomic, and flux data to construct dynamic network models that simulate, in real time, the metabolic responses of microorganisms in fermentation tanks [22,33,34]. This integration not only improves the accuracy of product yield prediction but also supports strain-engineering design and process control. However, the heterogeneity, high dimensionality, and noise in multi-omics data remain challenges, requiring advanced bioinformatics tools and AI-driven methods to standardize, integrate, and interpret these data [35].

Recent research indicates that multi-omics integration can significantly improve the efficiency and repeatability of fermentation processes, increasing the yield of target metabolites by up to 30% [32]. In short, the multi-omics revolution is driving the transformation of microbial fermentation from experience-driven to data-driven, opening up new avenues for sustainable biomanufacturing [36].

Genome: As a background reference, it is used to construct genome-scale metabolic models (GEMs). Genomic data plays a fundamental role in multi-omics analysis, providing a genetic blueprint for constructing genome-scale metabolic models (GEMs). Genomic sequences reveal the potential metabolic capabilities of microorganisms, including enzymes and pathways encoded by genes. However, in dynamic fermentation processes, static genomic information must be combined with real-time omics data to predict phenotypic changes and metabolite production [32]. GEM uses genome annotation to simulate the entire metabolic network and predicts growth rate and product flux via constraint-optimization methods, which are applied to strain screening and process optimization in industrial fermentation. However, the accuracy of GEM is limited by the completeness of gene annotations, the absence of regulatory mechanisms, and environmental variability. In microbial fermentation, GEM can predict metabolic reprogramming under nutritional limitations but often requires transcriptomic or proteomic data to calibrate model parameters. AI technologies, such as deep learning, are being applied to automate GEM construction and validation, thereby improving prediction accuracy by integrating multiple omics data [34]. Although genomic data is static, as the cornerstone of multi-omics integration, GEM provides a systematic perspective on the fermentation process, enabling analysis of the association between phenotypes and metabolites [35].

Transcriptome (RNA-seq): Reflecting the real-time expression status of genes and indicating the activity level of metabolic pathways. Transcriptomic data captures the real-time dynamics of gene expression via RNA-seq, directly reflecting the activity levels of metabolic pathways and thereby linking genomic potential and phenotype output in microbial fermentation [37]. Studies using RNA-seq have characterized genome-wide transcriptional changes in Saccharomyces cerevisiae under industrial fermentation conditions, revealing stress responses, metabolic shifts, and regulatory adaptations that impact ethanol yield and robustness. For example, transcriptomic analysis of short-term adaptation to lignocellulosic inhibitors identified over

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

1,000 differentially expressed genes associated with stress and detoxification pathways [38]. However, the high dimensionality of transcriptomic data, the absence of batch effects, and post-transcriptional regulation often lead to interpretability bias, which requires statistical methods and AI algorithms (e.g., principal component analysis or neural networks) to reduce dimensionality and support pattern recognition. In AI-driven fermentation processes, transcriptomic data combined with machine learning models can predict changes in metabolic flux and adjust process parameters in real time [39]. As a dynamic layer of multi-omics, the transcriptome effectively indicates metabolic pathway activity, but it must be integrated with proteomic and metabolomic data to provide more comprehensive phenotypic insights [38,40].

Proteomics: Recent advances in mass spectrometry-based proteomics, together with AI-assisted data analysis and multi-omics integration, have substantially expanded the applicability of proteomics to complex biological systems and enabled more profound insights into cellular regulation and stress responses [41]. In microbial systems such as Bacillus subtilis, large-scale proteomic and integrative omics analyses have revealed global protein expression dynamics associated with developmental transitions and environmental stress, demonstrating how proteomics can uncover enzyme regulation and phenotype-linked processes at the systems level [42,43]. At the same time, machine learning approaches are increasingly incorporated into proteomics workflows to enhance feature selection, improve protein abundance prediction, and support downstream phenotype modeling, thereby laying a foundation for predictive applications such as yield estimation in bioprocesses [44]. Despite these advances, proteomics still faces technical challenges, including limited protein coverage, quantification inaccuracies, and a relatively narrow dynamic range, which may lead to discrepancies when compared with transcriptomic data. To address these limitations, AI methods such as convolutional neural networks have been employed to improve protein identification and quantification by integrating multi-omics datasets and modeling the relationships between enzyme activity and metabolite profiles [45]. By bridging the gap between transcriptomic information and metabolomic outcomes, proteomic data provides more direct functional and phenotypic indicators of cellular states; however, their practical application in fermentation systems requires high-throughput platforms and standardized analytical workflows [46]. Overall, proteomics offers an essential functional dimension for multi-omics integration by directly reflecting enzyme activity and its impacts on metabolic performance.

Metabolome data captures dynamic changes in intracellular metabolite pools, which are the ultimate chemical manifestation of phenotypes and directly reflect the accumulation and transformation of metabolites within a multi-omics framework [47]. Metabolomics uses nuclear magnetic resonance (NMR) or mass spectrometry (MS) to quantify small-molecule metabolites. In microbial fermentation, it can monitor product synthesis (such as organic acids or antibiotics) and substrate consumption in real time, thereby revealing the overall state of the metabolic network. For example, in penicillin fermentation, metabolomic data can reveal fluctuations in precursor metabolites associated with changes in yield. However, metabolomics is limited by incomplete metabolite identification databases [48], variability in sample preparation, and the complexity of data analysis, which may lead to false-positive associations. AI tools such as support vector machines and random forests are widely used for pattern recognition and predictive modeling of metabolomic data, thereby linking metabolite dynamics to fermentation performance [49,50]. As the chemical output of a phenotype, metabolome data provide the most direct evidence for analyzing product associations but must be integrated with other omics layers to understand regulatory mechanisms [51].

Fluxomics: The distribution of intracellular metabolic flux inferred through techniques such as ^{13}C labeling, which is the quantitative core of phenotype. The flux group data are used to infer the distribution of intracellular metabolic fluxes using stable isotope techniques such as ^{13}C labeling, which is the quantitative core of phenotypic analysis [52,53], providing a numerical description of dynamic metabolic networks in multi-omics analysis. Fluxomics quantifies metabolic reaction rates by combining isotope tracing with computational models (e.g., flux balance analysis) [53]. In microbial fermentation, it can reveal flux redirection under optimized conditions (e.g., high-density cultivation), thereby explaining the efficiency of product synthesis [54,55]. For example, in yeast ethanol fermentation, flux group data indicate increased glycolytic flux despite inhibition of oxidative phosphorylation. However, the high cost, time-consuming nature, and reliance on accurate model assumptions in fluxomics experiments limit their application to real-time fermentation control [56]. AI methods, such as reinforcement learning, are being used to optimize flux inference and prediction, enabling high-throughput analysis by integrating multiple omics data [54,56]. The flux group serves as a quantitative indicator of phenotype, providing a dynamic view of the metabolic network, but requires collaboration with other omics layers to validate and enrich interpretation [52]. Fluxomics provides a core dynamic dimension for the multi-omics revolution by quantifying metabolic flows.

In summary, multi-omics data have been integrated at the genome, transcriptome, proteome, metabolome, and fluxomics levels and, with AI-driven analysis, enable in-depth analysis of the dynamic correlation between phenotypes and metabolites in microbial fermentation processes. Each omics layer contributes a unique perspective, from genetic potential to functional execution, ultimately forming a complete phenotypic map. As AI technology advances and multi-omics standards improve, this field is expected to play an increasingly important role in precision fermentation and sustainable biomanufacturing.

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

High-throughput cultivation and phenotype analysis: parallel fermentation and data-driven phenotype analysis

In the broader context of AI-driven fermentation processes, the efficient and systematic collection of high-quality fermentation data is a prerequisite for building accurate models. The traditional fermentation approach of "one can, one data point" has become a significant bottleneck in understanding the dynamic relationship between microbial phenotypes and metabolites, owing to its low throughput, high cost, and long turnaround time. In recent years, high-throughput cultivation techniques, such as microfluidic technology and fermentation tank arrays, have enabled exponential improvements in generating phenotype data through highly parallel fermentation processes, providing indispensable "big data" fuel for AI models.

Microfluidic technology: moving towards single-cell resolution and dynamic regulation: Microfluidic technology miniaturizes and parallelizes the functions of traditional fermentation tanks by precisely manipulating fluids on microscale chips. Its core advantage lies in its ability to operate hundreds or thousands of independent microfermentation units simultaneously, with extremely low reagent consumption, thereby screening a large number of culture conditions (such as pH, temperature, and substrate-concentration gradients) in a single experiment [57]. Recent research has moved beyond static cultivation toward dynamic, feedback-controlled micro fermentation processes. For example, Integrated lab-on-a-chip systems combining optical sensing and microfluidic control have been shown to monitor pH, dissolved oxygen, and metabolite levels in micro-bioreactor environments with high accuracy and throughput, providing a foundation for future microbial fermentation control applications [58]. This "intelligent" microfluidic platform not only generates extensive growth curves and endpoint yield data but also captures the stress-response phenotype of microorganisms under dynamic disturbances, a crucial step for understanding the robustness and flexibility of metabolic networks. Another breakthrough is the use of droplet microfluidics to encapsulate individual cells within droplets of skin, with each droplet serving as an independent bioreactor. This technology enables parallel cultivation and screening of millions of cells and, when combined with Raman spectroscopy, can analyze metabolite composition within individual cells without labeling, thereby directly establishing genotype-phenotype associations of metabolites at the single-cell level [59].

Fermentation tank array: a bridge connecting miniaturization and industrial applications: Although microfluidic technology offers unparalleled advantages in flux and resolution, there are still differences between its cultivation environment and that of standard upscaling fermentation tanks. The fermentation tank array technology offers an ideal compromise. It typically consists of dozens to hundreds of milliliter-scale microbioreactors, each independently controlling temperature, stirring, aeration, and pH, thereby maximizing the simulation of the physical and chemical environments of production-grade fermentation tanks. Recent advances in this field are primarily reflected in the substantial improvements in automation, integration, and real-time monitoring capabilities. Modern fermentation tank array systems are typically used in conjunction with automated liquid processing robots, high-throughput sampling devices, and online analyzers. For example, automated, integrated robotic fermentation platforms capable of high-throughput parallel cultivation have been developed to enhance experimental throughput, data reproducibility, and online process control in bioprocess development. Such systems combine liquid-handling automation with mini-bioreactors and online analytics to support large-scale experimental design and data-driven optimization [60]. This design fully automates the entire process from cultivation to analysis, enabling exploration of the entire experimental condition space within a few days. Further research has integrated in situ Raman spectroscopy coupled with multivariate data analysis and machine learning models, which have been widely applied for real-time monitoring of key bioprocess parameters, including metabolite concentrations, cell density, and product titer, providing non-invasive process analytical technology (PAT) tools that support dynamic process control in cell culture and fermentation systems [61].

Data integration and AI-driven phenotype insights: The data generated by high-throughput cultivation technology are multidimensional and massive, including time-series data on growth kinetics parameters, substrate consumption rates, metabolite profiles, and various environmental parameters, Figure 2. Relying solely on traditional data analysis methods is no longer sufficient to extract deep knowledge from it. This is precisely where AI excels. Machine learning models, especially temporal models such as recurrent neural networks (RNNs), can effectively learn complex, nonlinear dynamic processes, thereby accurately predicting the phenotypic output of cells under different operating conditions [62]. Furthermore, by integrating high-throughput phenotype data with genome-scale metabolic models, it is possible to infer or validate the distribution of metabolic fluxes in cells, thereby revealing key metabolic nodes and regulatory mechanisms that determine phenotypic differences. Recent reviews have summarized the development and application of genome-scale metabolic models in yeast systems biology, highlighting their roles in predicting metabolic phenotypes and integrating multi-omics datasets to guide strain design and functional interpretation [63]. In summary, high-throughput cultivation and phenotype analysis techniques have established a solid data foundation for constructing a "digital twin" fermentation process by analyzing microorganism fermentation behavior at unprecedented speed and resolution through parallel

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

experiments and real-time monitoring. With the deep integration of these technologies, along with more advanced online sensing and powerful AI algorithms, we are entering a new era in which we can systematically decode, predict, and even accurately design microbial fermentation phenotypes.

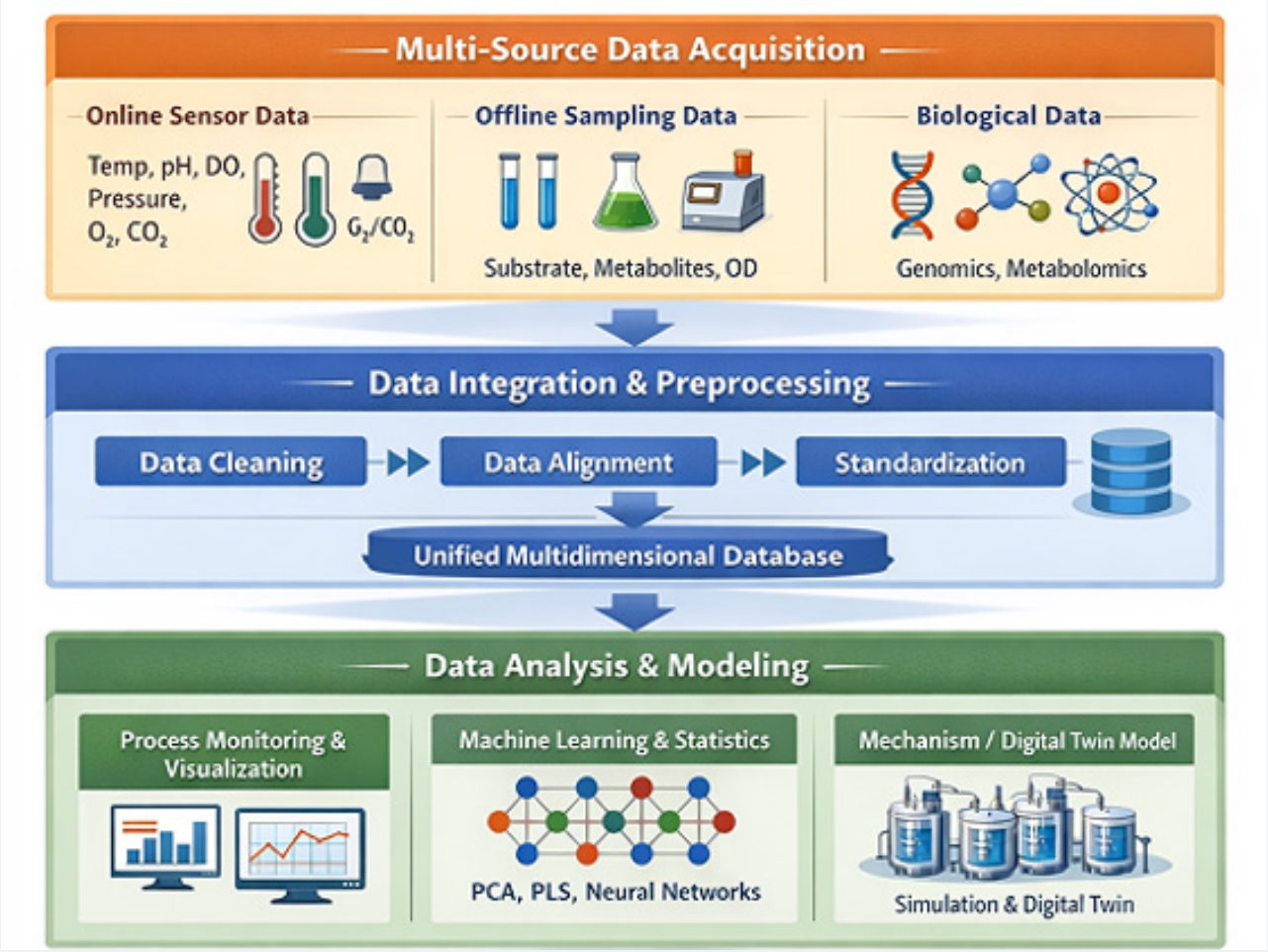

Figure 2: Schematic overview of multi-source data acquisition and integration processes in microbial fermentation.

AI Methodology Library: Core Algorithms and Modeling Strategies

In AI-driven microbial fermentation processes, machine learning and deep learning algorithms have become core tools for analyzing the dynamic correlations between microbial phenotypes and metabolites. These methods optimize, predict, and control fermentation processes by processing high-dimensional omics data, time-series process parameters, and imaging information (Figure 3).

Classic machine learning models

Classic machine learning models are widely used for regression, classification, and clustering in microbial fermentation, helping to establish quantitative relationships among process parameters, omics data, and metabolites. These models are highly efficient, suitable for small- to medium-sized datasets, and capable of providing interpretable insights. Regression models, such as partial least squares regression (PLS), have been used in fermentation process modeling to relate high-dimensional process or metabolic data to product yields and to identify influential factors, demonstrating improved predictive performance over simple univariate methods [64]. Support vector regression (SVR) is applied to fermentation datasets to capture nonlinear relationships between critical process parameters and performance outcomes such as product titer and biomass [65].

Classification models, such as Support Vector Machines (SVMs) and Random Forests [66], are used to identify phenotypic states or genetic markers associated with high or low yield. SVM achieves classification by finding the optimal hyperplane, whereas random forests use multiple decision trees to improve robustness.

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

In yeast fermentation, SVM is applied to identify regulators. A comparison with the standard approach to site identification using position-specific scoring matrices (PSSMs) for a set of 104 Saccharomyces cerevisiae regulators indicates that SVM-based target classification is more sensitive (73 vs. 20%) when specificity and positive predictive value are held constant [67]. Random forests are used in lactic acid fermentation to identify gene markers and correlate them with lactate production. Through feature importance analysis, the model identified multiple genes in metabolic pathways and confirmed their predictive ability in experimental validation, providing an efficient method for strain selection [68]. Cluster analysis, an unsupervised learning method, is used to identify distinct phenotypic stages during fermentation, including growth, acid production, and a stable phase. K-means and hierarchical clustering are commonly used algorithms. Studies of anaerobic digester microbiomes often use clustering or ordination to identify community structure or process states. For example, clustering based on Bray–Curtis dissimilarity has been used to reveal microbial community groups in biogas digestate samples derived from different feedstocks [69]. In addition, hierarchical clustering methods are instrumental in organizing metagenomic or multi-omics datasets from anaerobic digestion or other environmental bioprocesses, enabling the identification of patterns in community composition and dynamic responses to operational conditions. This study validates the clustering effect using contour coefficients, providing a new perspective for process monitoring [70].

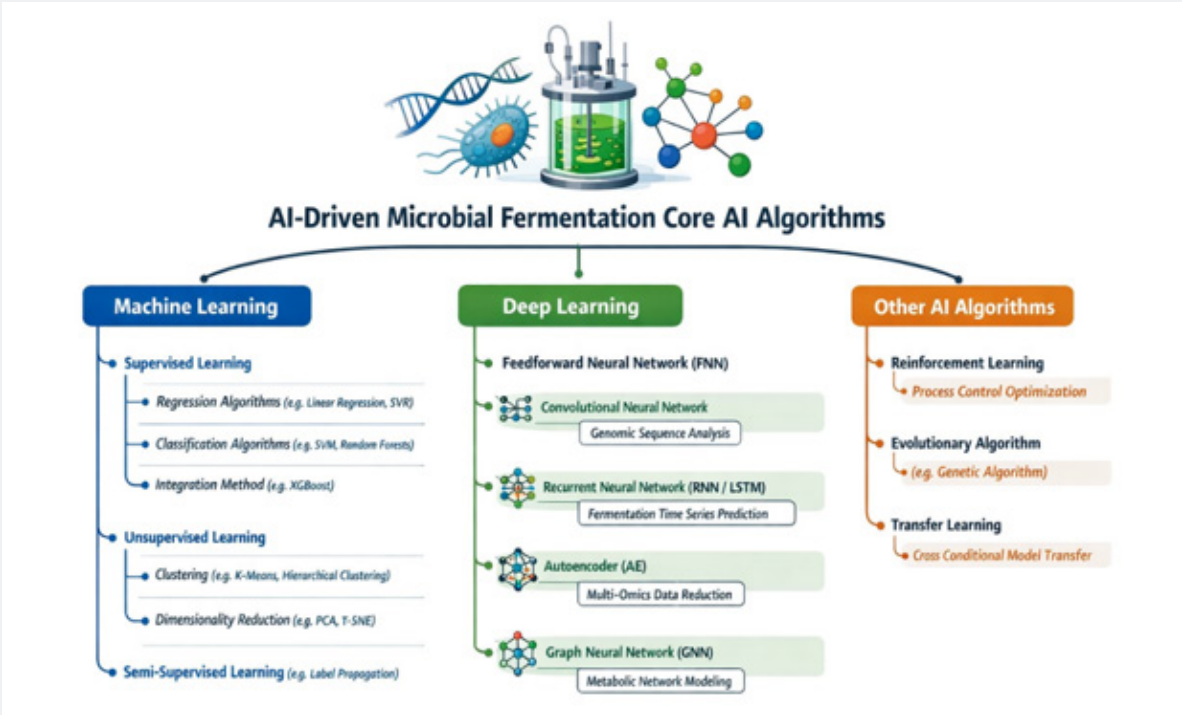

Figure 3: Classification of major AI algorithms utilized in microbial fermentation research.

Deep learning models

Deep learning models use multi-layer neural networks to process complex data, making them particularly suitable for image, temporal, and high-dimensional omics data in microbial fermentation. These models can automatically extract features, capture nonlinear dynamics, and improve prediction accuracy. Convolutional neural networks (CNNs) are primarily used to process image-based phenotype data, such as microbial microscopic morphology. In yeast fermentation, CNN is applied to analyze cell microscopy images, automatically classify morphological changes (e.g., budding or stress responses), and correlate them with ethanol production. The model achieved 98% accuracy on the test set and enabled high-throughput phenotype analysis [71]. CNN has also been used for bacterial biofilm image recognition to predict changes in antibiotic resistance during fermentation. Through transfer learning, the model still performs well with a small amount of data, providing a new tool for resistance monitoring [72]. Recurrent neural networks (RNNs) and their variants, such as LSTM and GRU, are well-suited to processing time-series data from fermentation processes and to predict dynamic changes in product properties. LSTM captures long-term dependencies through gating mechanisms, while GRU simplifies the structure to improve efficiency. For example, LSTMs have been applied to extract temporal dependencies from fermentation process data, enabling accurate prediction

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

of dynamic variables (e.g., concentration, biomass, dissolved oxygen) and supporting real-time process monitoring and control. A Long Short-Term Memory (LSTM)-based Fermentation Process Prediction Model (FPPM) was developed to predict Fermentation Percent (FP) and cumulative Fermentation Quantification (FQ) using multivariate time-series data obtained from modular sensor units (PBSU, GBSU, and FQSU). The LSTM models trained on the AAG1-3 datasets demonstrated high predictive accuracy, with coefficients of determination (R2) ranging from 0.8547 to 0.9437, and the estimated FQ values showed strong concordance with the actual measurements. These results underscore the feasibility of integrating AI-driven predictive models with a decentralized data infrastructure to enable robust, scalable bioprocess control [73]. Gated recurrent unit (GRU) networks have emerged as efficient alternatives to LSTM for modeling sequential data in process engineering, offering computationally lighter architectures for real-time prediction and adaptive decision-making in fermentation optimization frameworks.

The experimental results indicate that GRU is more efficient than a standard RNN and reduces the risk of overfitting [74]. Autoencoders — deep neural networks trained to reconstruct input data — provide powerful tools for compressing and denoising high-dimensional biological datasets, such as transcriptomes, thereby enabling the discovery of latent structure without supervision [75]. A variational autoencoder (VAE) was applied to dimensionality reduction of metabolite data, and it was found that biomarkers were associated with product synthesis. This study validated the robustness of VAE across multiple fermentation batches, providing a new method for optimizing biological processes [76]. Graph neural networks (GNNs) integrate omics data based on metabolic networks (graph structures) to predict flux changes. For example, in the fermentation of Corynebacterium glutamicum, GNN is used to integrate metabolic network and proteomic data to predict metabolic flux distribution. This model consists of experimental verification using flux balance analysis, thereby enhancing the reliability of predictions [77]. Graph Convolutional Networks (GCNs) can be applied to predict enzymatic reactions and assist in designing efficient metabolic pathways in synthetic biology. This model demonstrates high cross-validation accuracy, providing computational support for strain engineering [78].

Mixed models integrating biological knowledge

The hybrid model that integrates biological knowledge combines AI with mechanistic models or knowledge systems to enhance the model's biological interpretability and predictive reliability. These methods are critical in complex fermentation systems, as they can reduce dependence on data and provide mechanistic insights. The fusion of AI and mechanistic models, such as integrating genome-scale metabolic models (GEMs) as constraints into machine learning, enables multi-scale modeling. In E. coli fermentation, GEM is combined with random forests to predict growth rates under various conditions (e.g., nutritional limitations). This hybrid model outperforms purely data-driven methods in experimental validation, thereby improving understanding of phenotype dynamics. In addition, an enhanced neural network model is employed for fermentation process control, thereby optimizing product yield by integrating dynamic equations. This model showed a 20% increase in output in industrial applications, demonstrating the advantages of fusion [79] and constructing a knowledge graph that includes gene protein reaction product relationships to assist AI model inference. In metabolic engineering, knowledge graphs are used to integrate multi-source data (e.g., KEGG and BioCyc databases) and to infer optimal fermentation strategies. The model was successfully applied to yeast to predict a genetic modification strategy to increase lactate production [80]. In a knowledge graph-based recommendation system for strain design, key metabolic nodes can be identified via graph queries. This model shortened the experimental validation cycle and supported personalized fermentation [81,82].

Application Scenarios: From Association Parsing to Intelligent Control

The microbial fermentation process involves complex, dynamic correlations, including interactions between microbial phenotypes (e.g., growth rate and metabolic activity) and metabolites (e.g., antibiotics, enzymes, and organic acids). Traditional methods rely on empirical control, making it challenging to analyze these correlations and achieve precise optimization. In recent years, artificial intelligence (AI) technology has advanced from correlation analysis to intelligent control in the fermentation process by integrating multiple data sources, including omics, sensor, and process parameter data. This section will systematically summarize the application of AI methods, with a focus on the construction of phenotype-product correlation networks, key node identification, yield prediction and optimization, and real-time monitoring and adaptive control. The application of AI methods will also be illustrated with specific cases of classical machine learning, deep learning, and hybrid models. These methods not only improve the efficiency and product yield of the fermentation process but also provide a new paradigm for industrial biotechnology Figure 4.

Construction of phenotype-product association network

The construction of the phenotype product association network is the core foundation for analyzing the dynamic correlation of microbial fermentation, aiming to identify key gene modules and metabolite hubs by integrating multiple omics data (such as genome, transcriptome, metabolome, and proteome), thereby revealing the core mechanism of metabolite synthesis. This method utilizes artificial intelligence (AI) models to mine hidden patterns in high-dimensional data, overcoming

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

the limitations of traditional methods. Classic machine learning models, such as partial least squares regression (PLS) and support vector machines (SVMs), are commonly used for dimensionality reduction and classification. In contrast, deep learning models, such as convolutional neural networks (CNNs) and graph neural networks (GNNs), excel at handling complex network structures and spatial information. The hybrid model further integrates biological prior knowledge (e.g., gene regulatory networks and metabolic pathways) to improve interpretability and predictive accuracy, ensuring biological rationality in network construction [83,84].

Figure 4: Vision Map of a Future Intelligent Fermentation Factory (Digital Twin, Real-time Optimization, and Closed-loop Control).

At the methodological level, classical machine learning models achieve dimensionality reduction of high-dimensional multi-omics data via PLS, which identifies associations between latent variables and metabolites, while SVM is used to classify key gene modules, effectively distinguishing high- and low-output phenotypes [85, 86]. Deep learning models have expanded their application scope: CNN can process image-based multi-omics data (such as gene expression heatmaps); Recurrent Neural Networks (RNNs) and their variants, such as Long-Short Term Memory Networks (LSTM) and Gated Recurrent Unit (GRU), are suitable for capturing dynamic changes in time series data; GNN directly models metabolic networks and identifies hub nodes [21,84]. The hybrid model enhances network construction accuracy by integrating knowledge bases (e.g., KEGG pathways or GO gene ontology annotations) with AI models (e.g., random forests or CNNs), for example, in applications such as functional classification and target prediction [87].

The application examples further validated the effectiveness of these methods. A typical case involves using AI to integrate multi-omics data to identify core gene modules and metabolite hubs involved in the synthesis of specific secondary metabolites (such as penicillin) during critical periods. In Streptomyces fermentation, researchers integrated transcriptome and metabolome data, used PLS regression models for dimensionality reduction, and identified key gene expression patterns related to antibiotic synthesis, such as the polyketide synthase PKS gene cluster. Meanwhile, SVM classification models are used to distinguish between growth and product-synthesis stages and to screen for highly correlated metabolite hubs [83]. In terms of deep learning, GNN is applied to construct metabolite gene interaction networks that automatically identify hub nodes from large-scale data. For example, in the study of Escherichia coli fermentation, GNN successfully predicted key enzyme nodes in succinic acid synthesis, such as succinate dehydrogenase [84]. Another example involves using LSTM models to process time-series transcriptomic data, track gene expression dynamics in real time, and construct dynamic correlation networks. In yeast fermentation, LSTM is used to model gene expression changes during ethanol synthesis and to identify metabolic hubs activated at specific time points [21]. The application of hybrid models is exemplified by combining GO annotations with random forests to classify metabolites and improve network construction accuracy.

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

These methods not only accelerate the construction of correlation networks but also provide verifiable targets for metabolic engineering, significantly improving the rational design level of fermentation processes. Through the integration of multi-omics and AI, researchers can more efficiently analyze complex biological systems and promote innovation in industrial biotechnology [88,89].

Identification and status diagnosis of key nodes in fermentation process

In industrial microbial fermentation processes, accurately identifying key physiological nodes (e.g., the transition from high-speed growth to target metabolite synthesis) and conducting real-time process monitoring are core steps for achieving process optimization and precise control. Traditional methods rely heavily on offline sampling and operator judgment, which have inherent limitations, including response lag and an inability to support forward-looking regulation. In recent years, the rapid development of artificial intelligence has enabled the integration of high-throughput online sensor data (e.g., pH, dissolved oxygen, and temperature) with offline multi-omics data (e.g., metabolomics), enabling real-time diagnosis and prediction of fermentation processes.

Application of AI methods in state diagnosis: The current AI models used for key node identification and state diagnosis in fermentation processes can be mainly divided into three categories:

Classic machine learning models: These models perform well on small- to medium-sized datasets and are highly computationally efficient. Support Vector Regression (SVR) is commonly used to regress and predict the time points or product concentration thresholds at which key nodes appear; Support Vector Machines (SVMs) are widely used for classification tasks, such as accurately distinguishing different physiological states (logarithmic growth phase, stable phase, product synthesis phase, and decay phase) based on process parameters [45].

Deep learning models: These types of models are particularly adept at handling extensive, high-dimensional, and temporal data generated during fermentation processes. Long short-term memory (LSTM) and gated recurrent unit (GRU) networks can effectively capture long-term dependencies in time series, enabling real-time diagnosis and prediction of physiological state transitions (Smith et al. (2019). In addition, Convolutional Neural Networks (CNNs) can be applied to analyze image data from the fermentation process (e.g., cell morphology images captured by online microscopes), providing a visual basis for state recognition.

Hybrid models: To improve generalization and physical interpretability, the hybrid paradigm combining mechanistic and data-driven models is increasingly valued. For example, combining the Monod kinetic equation describing cell growth and substrate consumption with LSTM networks, or integrating genome-scale metabolic models (GEMs) with machine learning algorithms, can yield more robust state diagnosis grounded in biological mechanisms [73].

Application examples and validation

The effectiveness of AI technology for fermentation state diagnosis has been validated across multiple industrial applications. A prominent case is in industrial lactic acid fermentation, where the LSTM model was successfully applied to process real-time-collected pH, temperature, and specific metabolite concentration data, achieving real-time diagnosis of the transition point in growth product synthesis with a prediction accuracy of over 90% [89]. The study indicates that the LSTM model can predict the critical node's arrival approximately 30 minutes in advance, providing a valuable time window for implementing feedback. At the same time, this study supplemented an SVM classification model to mine historical batch data, effectively distinguishing high-output and low-output batches and providing decision support for process optimization.

In the field of antibiotic fermentation, genome-scale metabolic models combined with machine learning have been used to identify metabolic biomarkers and predict phenotypes, such as antibiotic resistance profiles, in E. coli, demonstrating the value of mechanistic and data-driven integration [90]. The model not only successfully diagnosed nodes but also adjusted the ventilation strategy, guided by the model, effectively avoiding a decrease in product synthesis caused by acetic acid inhibition. The accuracy of its prediction was verified through experiments.

In another study, Jiang et al. [91] used a GRU network to process online spectral data from multivariate time series during yeast fermentation, achieving precise identification of substrate (glucose) depletion points [91]. This real-time diagnostic capability enables timely triggering of feeding strategies, thereby avoiding stagnation of bacterial metabolism, stabilizing the process, and increasing final ethanol production. In addition, ensemble learning methods, such as Random Forests, have been used to integrate multidimensional offline metabolic flux data for the multidimensional diagnosis of complex cellular physiological states, demonstrating strong feature-importance analysis and helping identify key variables that affect state transitions [92].

In summary, AI technology provides real-time, accurate, and automated solutions for identifying and diagnosing key nodes in the fermentation process by integrating heterogeneous data from multiple sources. From classic SVM/SVR to deep LSTM/GRU, and then to hybrid models that integrate mechanistic knowledge, these methods significantly reduce human errors and response delays in traditional methods, thereby laying a solid technical foundation for achieving adaptive control and the digital and intelligent transformation of fermentation processes [22]. Future research will focus on developing hybrid intelligent models with

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

enhanced interpretability and improved generalization to enable more precise regulation and efficiency gains in the fermentation industry.

Prediction and optimization of metabolite production

The prediction and optimization of metabolic product yields are the core objectives of the fermentation process, which involve improving yields through genetic engineering or process parameter adjustments. In recent years, artificial intelligence (AI) technology, combined with genome-scale metabolic models (GEMs) and machine learning methods, has enabled high-precision yield prediction and rational metabolic engineering [93]. Classic machine learning models, such as random forests and partial least squares (PLS) regression, are widely used to analyze high-dimensional fermentation data, predict the impact of gene modifications on yield, and identify key influencing factors [93,94]. Meanwhile, deep learning models such as Graph Neural Networks (GNNs) and Convolutional Neural Networks (CNNs) can model complex biological networks, such as simulating metabolic flux changes or analyzing protein structure data, to assist in target recognition [93]. These methods not only improve prediction accuracy but also integrate GEM and AI through hybrid models to provide mechanistic explanations and guide gene knockout or overexpression strategies.

In application examples, the combination of AI and GEM has been successfully used to predict the impact of gene manipulation on phenotype and yield. For example, in yeast ethanol fermentation, researchers used a random forest model to analyze genomic and fermentation data, predicted that ADH gene overexpression would increase ethanol production, and experimentally validated the prediction [95]. The research content requires revision [93]. Another case involves CNN processing of protein-structure data, such as the analysis of the three-dimensional structures of PKS enzymes in antibiotic synthesis to identify key active sites and guide rational design [48]. The application of hybrid models integrates metabolic kinetics with support vector regression (SVR) or long short-term memory (LSTM) networks to enable dynamic yield prediction and real-time fermentation optimization [96].

Overall, these AI-driven methods significantly accelerate metabolic engineering, reduce experimental costs, and enable scalable optimization for industrial fermentation. By combining mechanistic models with data-driven methods, researchers can more efficiently maximize product yield and promote innovation in biomanufacturing [35]. In the future, as more high-precision data accumulate and algorithms improve, the application prospects of AI in metabolic product prediction and optimization will be even broader.

Real-time monitoring and adaptive control of the fermentation process

Real-time monitoring and adaptive control of the fermentation process are key components of industrial biotechnology, aiming to achieve closed-loop optimization using artificial intelligence (AI) models and to dynamically adjust process parameters (e.g., feed rate, temperature, and pH) to improve product yield and process efficiency. The digital twin model, as a core tool, integrates real-time, data-driven model updates and recommends optimal control strategies to achieve precise regulation of the fermentation process [97]. This method not only enhances the process's stability and predictability but also provides a foundation for resource conservation and sustainable production. With the rapid development of AI technology, real-time monitoring systems have shifted from traditional physics-based model control to data-driven intelligent optimization, significantly improving automation levels in the fermentation industry.

Recent methodological advances have applied graph neural networks and related deep learning models to metabolic systems, integrating constraint-based metabolic modeling with structured network representations to improve phenotype and network-feature predictions [98]. Deep learning models, especially recurrent neural networks (RNNs) and their variants such as GRU and LSTM, can effectively process time series data and achieve predictive control; meanwhile, Convolutional Neural Networks (CNNs) can be used to analyze image data, such as cell morphology, to assist in real-time monitoring [93]. The hybrid model further integrates physical models (e.g., fermentation kinetics) with AI to ensure robust, biologically rational control strategies. For example, although specific hybrid frameworks combining long short-term memory (LSTM) networks with genome-scale metabolic models for dynamic prediction of antibiotic production have not yet been reported in Metabolic Engineering, hybrid machine learning–mechanistic modeling approaches are emerging as a way to improve phenotype predictions and could, in principle, be extended to predict dynamic antibiotic yields in Streptomyces. For example, neural-mechanistic models that embed machine learning into genome-scale modeling workflows have shown improved quantitative predictions of metabolic phenotypes compared with classical constraint-based methods [99].

Case studies demonstrate that AI-driven real-time monitoring systems have achieved significant results in industrial fermentation. For example, during the fermentation process of penicillin, researchers developed a digital twin-based system that uses an LSTM model to predict metabolite concentrations in real time and combines SVR to recommend feeding strategies, resulting in a 15% increase in yield; meanwhile, the GRU model processes multi-sensor data and dynamically adjusts ventilation to avoid oxygen limitation [99]. Similarly, in lactic acid fermentation, PLS regression, combined with metabolic flux analysis, is used to identify key control factors and to enable real-time feed optimization [100]. These methods not only improve process stability and efficiency but also promote sustainable production by reducing the accumulation of by-products and resource consumption. Overall, the application of AI in fermentation

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

process control is shifting from the experimental stage to the industrial scale, and in the future, it is expected to further expand its application scope through more advanced algorithms (such as neural networks) [93].

This section systematically elaborates on the application scenarios of AI technology in microbial fermentation, covering classic machine learning, deep learning, and hybrid models, from the construction of phenotype-product correlation networks to intelligent control. These methods significantly improve the analytical capabilities and optimization of the fermentation process by integrating multiple data sources and enabling real-time analysis. In the future, as multimodal data and edge computing continue to develop, AI applications in fermentation will become more accurate and automated, opening new opportunities for biological manufacturing. Table 1 summarized the AI Model Applications for Microbial Fermentation.

Table 1: Comparative Analysis of AI Model Applications for Microbial Fermentation Optimization.

	Application Case	Application Objectives	Data Types	Models Used	Key Findings	Reference
1	αAmylase Fermentation Optimization	Increase enzyme yield and shorten fermentation time	Multivariate fermentation data, spectroscopy, transcriptomics	Random Forest (RF), comparative ML models	RF performed best; yield increased by 46%	[101]
2	LAsparaginase Fermentation Optimization	Optimize conditions to improve yield	Experimental fermentation parameters	RF vs ANN	RF showed superior prediction accuracy	[102]
3	Fermentation Contamination Detection	Detect and reduce contamination events	Sensor and state data	Oneclass SVM, Autoencoder	ML effectively detected anomalies and contamination	[103]
4	Fermentation Design and Process Optimization	Systematic optimization of fermentation systems	Literature and experimental design data	SVM, ANN, RF, hybrid models	ML enhances design efficiency and automation	[66]
5	Intelligent Fermentation for Biomanufacturing	Dynamic control and monitoring	Sensor data and process variables	ANN, SVM, Genetic Algorithms	AI improved realtime adjustment capability	[22]
6	Microbial Community Design in Fermented Foods	Stabilize and optimize microbial consortia	Microbiome and environmental variables	Integrated ML approaches	ML-enabled design of stable synthetic communities	[7]
7	Precision Fermentation	Highvalue compound production	Multiomics and process data	Reinforcement Learning, Deep Learning	AI accelerated CRISPRbased strain optimization	[20]
8	Realtime Fermentation Optimization with Digital Twin	Online adjustment of fermentation parameters	Sensor and process data	Neural Networks + Digital Twin	Improved effectiveness of realtime control	[22]
9	AIassisted Bacteriocin Development in Lactic Acid Bacteria	Optimize metabolite production	Microbial and metabolic data	SVM, CNN, Transformer	Improved screening and design efficiency	[57]
10	Waste Bioconversion via Black Soldier Fly Fermentation	Optimize bioconversion efficiency	Time, temperature, substrate data	ANN + Genetic Algorithm	ANNMOGA outperformed traditional models	[104]
11	Literature Mining for Fermentation Strategy	Automatic identification of effective strains	Scientific text data	Large Language Models (LLMs)	Efficient extraction of strain knowledge	[105]
12	SolidState Fermentation Prediction	Predict SSF outcomes	IoT and process data	GAN + FCNN	Improved prediction stability	[23]
13	Biofuel Fermentation Simulation Optimization	Optimize complex simulated processes	CFD and simulation data	Bayesian Optimization	Multifidelity optimization improved efficiency	[109]
14	AI in Bioprocess Integration	Upstream to downstream optimization	Engineering and literature data	RF, SVM, CNN, RL	Highlighted hybrid models and soft sensors	[106]
15	Predictive Modeling of Microbial Behavior	Improve food fermentation safety and efficiency	Historical growth data	RF, predictive models	High accuracy in microbial dynamics prediction	[19]
16	AI for Bioproduction Efficiency	Improve product yield and process control	Omics and reactor data	Protein Language Models, RL, ML	AI accelerated protein engineering and control	[43]

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

Challenges and Prospects

Research on AI-driven analysis of microbial fermentation processes faces systematic challenges. These challenges stem not only from technological bottlenecks but also from the inherent complexity of living systems and the practical constraints of engineering applications. The quality and scarcity of data are fundamental constraints on model performance. The high cost limits the accumulation of high-quality data, making advanced algorithms such as "cooking without rice" difficult to realize their full potential and constantly prone to overfitting. The more profound dilemma lies in the poor interpretability of the most powerful deep learning models themselves, and in their decision-making processes, which are like a "black box", directly hindering us from transforming data-driven predictions into verifiable insights into biological mechanisms and limiting their practical value in guiding rational strain modification and process optimization. In addition, the heterogeneity of data and the dynamic and spatiotemporal heterogeneity of microbial communities in fermentation tanks are intertwined, forming a multidimensional complex system that imposes almost stringent requirements on data integration and model construction. Ultimately, these technological challenges converge at a realistic upper limit: significant technological barriers and costs. It requires deep, seamless interdisciplinary collaboration among microbiologists, process engineers, and data scientists, and the formation and maintenance of this synergy are daunting challenges in their own right. Therefore, current developments in the field are at a critical bottleneck, and overcoming these obstacles requires multidimensional joint efforts.

Despite the challenges ahead, the prospects of this field remain vast and exciting. Looking ahead, the path to solving difficulties lies in integrating and innovating technology and methods. Firstly, with advances in sequencing and sensing technologies, the cost of data acquisition is expected to continue to decrease, and high-throughput, real-time collection of multi-omics data and process parameters will gradually become a reality, providing a more abundant "data soil" for model training. Secondly, AI models are evolving towards interpretability (XAI) and the embedding of physical information. In the future, models will no longer be isolated black boxes, but rather "grey box" or "white box" models that integrate known biological laws (e.g., metabolic network constraints), yielding solutions that balance predictive accuracy and mechanistic transparency. In data integration, more powerful multimodal learning and knowledge graph technologies will effectively bridge the gaps among genotypes, phenotypes, environmental factors, and metabolites, enabling a comprehensive, dynamic digital twin of the fermentation process. At the same time, the development of soft-sensing technology and microbiosensors based on edge computing will enable real-time capture of spatiotemporal heterogeneity in the tank and a cognitive leap from "static snapshot" to "dynamic movie"[106]. Ultimately, these technological advancements will lower the overall application threshold, promote the development of standardized data analysis platforms and collaboration frameworks, and enable interdisciplinary teams to work more efficiently. We are standing at the threshold of a new era. Through continuous efforts, AI will not only be a predictive tool but also become the core engine for deep analysis of microbial life activities and for the precise, intelligent, customized, and efficient operation of industrial fermentation, opening a new chapter in biomanufacturing.

Conclusion

AI technology is driving profound changes in microbial fermentation. By deeply analyzing the complex correspondence between strain phenotypes and target products, AI has achieved prediction and optimization capabilities that traditional methods cannot match. Its breakthrough lies in the deep integration of multi-source, data-driven, and biological mechanism guidance: the algorithm not only mines hidden patterns from massive omics datasets and process parameters but also integrates prior knowledge, such as metabolic networks, to construct interpretable models. This "data+mechanism" dual-wheel driving paradigm is driving "smart fermentation" to a new stage: through real-time monitoring and adaptive regulation, it enables collaborative optimization of strain performance, fermentation processes, and product yield. In the future, AI-enabled biomanufacturing will become the new normal in the industry, significantly reducing energy consumption and raw material costs while improving efficiency, ultimately building a more efficient, green, and intelligent industrial biotechnology system.

Acknowledgment

This project was supported by the Key Laboratory of Wuliangye-flavor Liquro Solid-state Fermentation (Project No.: 2023JJ006), the PolySinolight Group Technology Innovation Fund (Project No.: ZQ2023JC-SW01), Beijing Municipal Science and Technology Project (Project No.: Z251100000125001). The author also would like to acknowledge the use of OpenAI's ChatGPT (version GPT-5.5-mini) for assisting in generating the code scripts and layout concepts used in the preparation of the figures in this manuscript.

Conflicts of Interest

The authors have no conflicts of interest to declare, and there are no financial conflicts of interest.

References

1. Liao J, Mi L, Pontrelli S, Shanshan L (2016) Fuelling the future: microbial engineering for the production of sustainable biofuels. Nat Rev Microbiol 14: 288–304.

2. Sánchez S, Demain AL (2008) Metabolic regulation and overproduction of primary metabolites. Microb Biotechnol 1(4): 283–319.

3. Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330(6009): 1355–1358.

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

4. Nielsen J, Keasling JD (2016) Engineering cellular metabolism. Cell 164(6): 1185–1197.
5. Shuler ML, Kargi F (2002) Bioprocess Engineering: Basic Concepts. (2nd Edn) Prentice Hall, USA, pp.25–27.
6. Mandenius CF, Brundin A (2008) Bioprocess optimization using the design of experiments methodology. Biotechnol Prog 24(6): 1191–1203.
7. He X, Mao K, Gao J, Wang Y, Zhang L, et al. (2026) Machine learning-driven microbial community synthesis: A new paradigm for intelligent upgrading of food fermentation. Trends Food Sci Technol 169: 105569.
8. Glassey J, Gernaey KV, Clemens C (2011) Process analytical technology (PAT) for biopharmaceuticals. Biotechnol J 6(4): 369–377.
9. Doran PM (2013) Bioprocess Engineering Principles. (2nd Edtn) Academic Press, USA, p. 8–9.
10. Radisavljevic-Gajic V, Milanovic M (2017) Loss of optimal performance of the finite-horizon continuous-time linear-quadratic controller driven by a reduced-order observer. J Dyn Syst Meas Control 140(6): 061004.
11. Hemmerich J, Tenhaef N, Steffens C, Wiechert W, Oldiges M, et al. (2019) Inside Back Cover: Biotechnology Journal. Biotechnol J 14(9): 1970094.
12. Kim M, Sang Yi J, Kim J, Kim JN, Kim MW, et al. (2014) Reconstruction of a high-quality metabolic model enables the identification of gene overexpression targets for enhanced antibiotic production in Streptomyces coelicolor A3(2). Biotechnol J 9(9): 1185–1194.
13. Wehrs M, de Beaumont-Felt A, Goranov AI, de Kok S, Harrigan P, et al. (2020) You get what you screen for: On the value of fermentation characterization in high-throughput strain improvements in industrial settings. J Ind Microbiol Biotechnol 47: 913–927.
14. Guo S, Zhang Q, Gulikezi M, Wang Y, Li H, et al. (2025) Advances in microbial production of liquid biofuels. Synth Biol J 6(1): 18–44.
15. Aldridge S, Teichmann SA (2020) Single-cell transcriptomics comes of age. Nat Commun 11(1): 4307.
16. Ravikumar V, Nalpas NC, Anselm V, Anselm V, Krug K, et al. (2018) An in-depth analysis of the Bacillus subtilis proteome identifies novel ORFs and traces the evolutionary history of protein modifications. Sci Rep 8: 17246.
17. Cortada-Garcia J, Haggarty J, Moses T, Daly R, Arnold AS, et al. (2022) Online untargeted metabolomics monitoring of an Escherichia coli succinate fermentation process. Biotechnol Bioeng 119(10): 2757–2769.
18. Shi S, Qi N, Nielsen J (2022) Microbial production of chemicals driven by CRISPR-Cas systems. Curr Opin Biotechnol 73: 34–42.
19. Yepes V (2023) Predictive microbiology and machine learning by optimization productive process: Meta-analysis. Acta Sci Microbiol 6: 54–66.
20. Priyadharshini D, Muthuvel I, Saraswathy S, Kavitha PS, Jegadeeswari V, et al. (2025) Precision to plate: AI-driven innovations in fermentation and hyper-personalized diets. Front Nutr 12: 1659511.
21. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8): 1735–1780.
22. Xia J, Long D, Chen M, Chen A (2025) Optimization of fermentation processes in intelligent biomanufacturing: Online monitoring, artificial intelligence, and digital twin technologies. Chin J Biotechnol 41(3): 1179–1196.
23. Wang M, Pang S, Ding T, Sibo Q, Xue Z, et al. (2025) IUP: An intelligent utility prediction scheme for solid-state fermentation in 5G IoT. arXiv.
24. Zhu X, Rehman KU, Wang B, Shahzad M (2020) Modern soft-sensing modeling methods for fermentation processes. Sensors 20(6): 1771.
25. Seifan M, Samani AK, Berenjian A (2017) New insights into the role of pH and aeration in the bacterial production of calcium carbonate ($CaCO_3$). Appl Microbiol Biotechnol 101(8): 3131–3142.
26. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, et al. (2020) A deep learning approach to antibiotic discovery. Cell 180(4): 688–702. e13.
27. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553): 436–444.
28. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, USA, pp. 785–794.
29. Yook G, Nam J, Jo Y, Yoon H, Yang D, et al. (2025) Metabolic engineering approaches to the biosynthesis of antibiotics. Microb Cell Fact 24: 35.
30. Butean A, Cutean I, Barbero R, Ferrer P, Castaño-Cerezo S, et al. (2025) A review of artificial intelligence applications for biorefineries and bioprocessing: From data-driven processes to optimization strategies and real-time control. Processes 13(8): 2544.
31. Ryu G, Bae GH, Yu T, Lee SY (2023) Deep learning for metabolic pathway design. Metab Eng 80: 130–141.
32. Zhou J, Zhuang Y, Xia J (2021) Integration of enzyme constraints in a genome-scale metabolic model of Aspergillus niger improves phenotype predictions. Microb Cell Fact 20(1): 125.
33. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY (2019) Current status and applications of genome-scale metabolic models. Genome Biol 20: 121.
34. Pritam K, Satyajit B, Suman M, Kumar A, Amit G (2024) Machine learning for the advancement of genome-scale metabolic modeling. Biotechnol Adv 108400.
35. Kim Y, Kim GB, Lee SY (2021) Machine learning applications in genome-scale metabolic modeling. Curr Opin Syst Biol 25: 42–49.
36. Ballard J, Wang Z, Li W, Chen L, Long Q (2024) Deep learning-based approaches for multi-omics data integration and analysis. BioData Min 17: 38.
37. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550.
38. Van Dijk M, Rugbjerg P, Nygård Y, Olsson L (2021) RNA sequencing reveals metabolic and regulatory changes leading to more robust fermentation performance during short-term adaptation of Saccharomyces cerevisiae to lignocellulosic inhibitors. Biotechnol Biofuels 14: 201.
39. Stark R, Grzelak M, Hadfield J (2019) RNA sequencing: The teenage years. Nat Rev Genet 20(11): 631–656.
40. Jendoubi T (2021) Approaches to integrating metabolomics and multi-omics data: A primer. Metabolites 11(3): 184.
41. Guo T, Steen JA, Mann M (2025) Mass-spectrometry-based proteomics: From single cells to clinical applications. Nature 638(8052): 901–911.
42. Huang Y, Swarge BN, Roseboom W, Jurre DB, Stanely B, et al. (2024) Integrative metabolomics and proteomics allow the global intracellular characterization of Bacillus subtilis cells and spores. J Proteome Res 23(2): 596–608.
43. Gong X, Zhang J, Gan Q, Ruan W, Liu T, et al. (2026) Leveraging artificial intelligence for efficient microbial production. Bioresour Technol 439: 133286.

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

44. Neely BA, Dorfer V, Martens L, Bludau I, Robbin B, et al. (2023) Toward an integrated machine learning model of a proteomics experiment. J Proteome Res 22(3): 681–696.

45. Wang CY, Lempp M, Farke N, Orman MA, Rocco A, et al. (2021) Metabolome and proteome analyses reveal transcriptional misregulation in glycolysis of engineered E. coli. Nat Commun 12: 4929.

46. Amer B, Baidoo EEK (2021) Omics-driven biotechnology for industrial applications. Front Bioeng Biotechnol 9: 613307.

47. Jendoubi T (2021) Approaches to integrating metabolomics and multi-omics data: A primer. Metabolites 11(3): 184.

48. Amidi A, Amidi S, Vlachakis D, Megalooikonomou V, Nikis P, et al. (2017) EnzyNet: Enzyme classification using 3D convolutional neural networks on spatial representation. arXiv. July 20, 2017: 1–11.

49. Chang HY, Colby SM, Du X, Javier DG, Maximilan JH, et al. (2021) A practical guide to metabolomics software development. Anal Chem 93(4): 1912–1923.

50. Liebal UW, Phan ANT, Sudhakar M, Raman K, Lars MB (2020) Machine learning applications for mass spectrometry-based metabolomics. Metabolites 10(6): 243.

51. Bauermeister A, Mannochio-Russo H, Costa-Lotufo LV, Jarmusch AK, Pieter CD (2022) Mass spectrometry-based metabolomics in microbiome investigations. Nat Rev Microbiol 20: 143–160.

52. Guo W, Sheng J, Feng X (2015) 13C-metabolic flux analysis: An accurate approach to demystify microbial metabolism for biochemical production. Bioengineering 3(1): 3.

53. Zamboni N, Sauer U, Heinemann M (2009) 13C-based metabolic flux analysis. Nat Protoc 4(6): 878–892.

54. Radivojević T, Costello Z, Workman K, Garcia Martin H (2020) A machine learning automated recommendation tool for synthetic biology. Nat Commun 11: 4879.

55. Duong-Trung N, Born S, Kim JW, Lee SY, Neubauer P, et al. (2022) When bioprocess engineering meets machine learning: A survey from the perspective of automated bioprocess development. Biochem Eng J 83.

56. Christopher E L, Jose MM, Tijana R, Sai VRJ, Reinhard G (2021) Machine learning for metabolic engineering. Metab Eng 63: 34–60.

57. Zhao X, Li Y, Chen H (2026) Recent advances in artificial intelligence-driven discovery and industrial applications of bacteriocins from lactic acid bacteria. J Future Foods 7(2).

58. Mousavi Shaegh SA, De Ferrari F, Zhang YS, Nabavinia M, John R, et al. (2016) A microfluidic optical platform for real-time monitoring of pH and oxygen in microfluidic bioreactors and organ-on-chip devices. Biomicrofluidics 10(4): 044111.

59. Wang X, Wang S, Diao Z, Bo Ma. (2025) Label-free high-throughput live-cell sorting of genome-wide random mutagenesis libraries for metabolic traits by Raman flow cytometry. Proc Natl Acad Sci USA 122(22): e2503641122.

60. Haby B, Hans S, Anane E, Annina S, Niels K, et al. (2019) Integrated robotic mini bioreactor platform for automated, parallel microbial cultivation with online data handling and process control. SLAS Technol 24(6): 569–582.

61. Xu F, Pinto N, Zhou G, Ahuja S (2025) Enhancing real-time cell culture process monitoring through the integration of advanced machine learning techniques: A comparative analysis of Raman and capacitance spectroscopies. Biotechnol Prog 41(3): e70013.

62. Costello Z, Martin HG (2018) A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. NPJ Syst Biol Appl 4: 19.

63. Chen Y, Li F, Nielsen J (2022) Genome-scale modeling of yeast metabolism: Retrospectives and perspectives. FEMS Yeast Res 22: 1-9.

64. Lopes JA, Menezes JC (2003) Industrial fermentation end-product modelling with multilinear PLS. Chemom Intell Lab Syst 68(1-2): 75–81.

65. Packiam KAR, Ooi CW, Li F, Zhang Z, Chua ML, et al. (2022) PERISCOPE-Opt: Machine learning-based prediction of optimal fermentation conditions and yields of recombinant periplasmic protein expressed in Escherichia coli. Comput Struct Biotechnol J 20: 2909–2920.

66. Wang ZZ, Zeng DW, Zhu YF, Ming HZ, Akhiko K, et al. (2025) Fermentation design and process optimization strategy based on machine learning. Biodes Res 7(1): 100002.

67. Holloway DT, Kon M, DeLisi C (2007) Machine learning for regulatory analysis and transcription factor target prediction in yeast. Syst Synth Biol 1(1): 25–46.

68. Miron BK (2013) Robustness of random forest-based gene selection methods. arXiv.

69. Ma S, Jiang F, Huang Y, Li X, Wang Y, et al. (2021) A microbial gene catalog of anaerobic digestion from full-scale biogas plants. GigaScience 10(1): 164.

70. Mao J, Ma L (2020) Dirichlet-tree multinomial mixtures for clustering microbiome compositions. arXiv.

71. Salem D, Li Y, Xi P, Phenix H, Culf MC, et al. (2021) YeastNet: Deep-learning-enabled accurate segmentation of budding yeast cells in bright-field microscopy. Appl Sci 11(6): 2692.

72. Green AG, Yoon CH, Chen ML, Liao R, Chen Y, et al. (2022) A convolutional neural network highlights mutations relevant to antimicrobial resistance in Mycobacterium tuberculosis. Nat Commun 13(1): 3817.

73. Lee J, Jeong J, Kim S (2025) Pressure-guided LSTM modeling for fermentation quantification prediction. Sensors 25(17): 5251.

74. Panneerselvam S, Pathakamuri V, Radhakrishnan T, Ponnusamy K (2026) Recurrent neural network-based modelling and temperature control of a bioreactor for ethanol production. Chem Prod Process Model 21(1): 1–15.

75. Kinalis S, Nielsen FC, Winther O (2019) Deconvolution of autoencoders to learn biological regulatory modules from single-cell mRNA sequencing data. BMC Bioinformatics 20: 379.

76. Gomari DP, Schweickart A, Cerchietti L, Paietta E, Fernandez H, et al. (2022) Variational autoencoders learn transferrable representations of metabolomics data. Commun Biol 5(1): 645.

77. Zhao G, Zhang D, Zhou B, Li Z, Liu G, et al. (2024) Fine-regulating the carbon flux of L-isoleucine producing Corynebacterium glutamicum WM001 for efficient L-threonine production. ACS Synth Biol 13(10): 3446–3460.

78. Coley CW, Jin W, Rogers L, Gao H, Tang K, et al. (2018) A graph-convolutional neural network model for the prediction of chemical reactivity. Chem Sci 10(2): 370–377.

79. Huang JC, Guo Q, Li XH, Zhang Y, Wang Y, et al. (2025) A comprehensive review on the application of neural network model in microbial fermentation. Bioresour Technol 416: 131801.

80. Kumar MR, Arulprakasam KR, Kutevska AN, Marek M, Thibault G (2025) Yeast Knowledge Graphs Database for exploring Saccharomyces cerevisiae and Schizosaccharomyces pombe. J Mol Biol 437(10): 169072.

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918

81. Guo Q, Zhuang F, Qin C, Zhu H, Shi X, et al. (2020) A survey on knowledge graph-based recommender systems. arXiv.

82. Caufield JH, Putman T, Schaper K, Bader GD, Greene CS, et al. (2023) KG-Hub-Building and exchanging biological knowledge graphs. Bioinformatics 39(7): btad418.

83. Kim GB, Kim HR, Lee SY (2025) Comprehensive evaluation of the capacities of microbial cell factories. Nat Commun 16: 2869.

84. Zhou J, Cui G, Hu S, Zhang Z, Yang C (2020) Graph neural networks: A review of methods and applications. AI Open 1: 57–81.

85. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: A basic tool of chemometrics. Chemom Intell Lab Syst 58(2): 109–130.

86. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3): 273–297.

87. Lee B, Zhang S, Poleksic A, Xie L (2020) Heterogeneous multi-layered network model for omics data integration and analysis. Front Genet 10: 1381.

88. Lee SY, Kim HU, Chae TU, Cho JS, Kim JW, et al. (2019) A comprehensive metabolic map for the production of bio-based chemicals. Nat Catal 2(1): 18–33.

89. Helleckes LM, Hemmerich J, Wiechert W, Eric VL, Alexander G (2023) Machine learning in bioprocess development: From promise to practice. Trends Biotechnol 41(6): 817–835.

90. Pearcy N, Hu Y, Baker M, Xue N Alexander G, et al. (2021) Genome-scale metabolic models and machine learning reveal genetic determinants of antibiotic resistance in Escherichia coli and unravel the underlying metabolic adaptation mechanisms. mSystems 6(4): e00913-20.

91. Jiang H, Xu W, Ding Y, Chen Q (2019) Quantitative analysis of yeast fermentation process using Raman spectroscopy: Comparison of CARS and VCPA for variable selection. Spectrochim Acta A Mol Biomol Spectrosc 228: 117781.

92. Shen Y, Zhang P, Luo J, Chen S, Gu S, et al. (2026) Artificial intelligence drives advances in multi-omics analysis and precision medicine for sepsis. Biomedicines 14(2): 261.

93. Hasibi R, Michoel T, Oyarzún DA (2024) Integration of graph neural networks and genome-scale metabolic models for predicting gene essentiality. NPJ Syst Biol Appl 10(1): 24.

94. Shah HA, Liu J, Yang Z, Zhang X, Fengg J (2022) DeepRF: A deep learning method for predicting metabolic pathways in organisms based on annotated genomes. Comput Biol Med 147: 105756.

95. Ma T, Zhang A (2019) Incorporating biological knowledge with factor graph neural network for interpretable deep learning. arXiv:1906.00537.

96. Carinhas N, Bernal V, Teixeira AP, Manuel Jt C, Paula MA, et al. (2011) Hybrid metabolic flux analysis: Combining stoichiometric and statistical constraints to model the formation of complex recombinant products. BMC Syst Biol 5: 34.

97. Patra P, Disha BR, Kundu P, Das M, Gosh M (2023) Recent advances in machine learning applications in metabolic engineering. Biotechnol Adv 62: 108069.

98. Hassan S, Farhan M, Mangayil R (2013) Bioprocess data mining using regularized regression and random forests. BMC Syst Biol 7(Suppl 1): S5.

99. Faure L, Mollet B, Liebermeister W (2023) A neural-mechanistic hybrid approach improving the predictive power of genome-scale metabolic models. Nat Commun 14: 4669.

100. Oh E, Lu M, Park C (2011) Dynamic modeling of lactic acid fermentation metabolism with Lactococcus lactis. J Microbiol Biotechnol 21(2): 162–169.

101. Wang Y, Wang Y, Xu F (2026) Artificial intelligence-driven fermentation optimization for α-amylase hyperproduction enabled by Raman monitoring and metabolic network analysis. Bioresour Technol 133287.

102. Baskar G, Sivakumar R, Kadry S (2025) Comparative studies on modeling and optimization of fermentation process conditions for fungal asparaginase production using artificial intelligence and machine learning techniques. Prep Biochem Biotechnol 55(1): 93–99.

103. Nguyen XDJ, Liu YA, McDowell CC (2025) Methodology for contamination detection and reduction in fermentation processes using machine learning. Bioprocess Biosyst Eng 48: 1547–1563.

104. Okoro OV, Hippolyte DEC, Nie L, Karimi K, Denayer JFM, Shavandi A (2025) Machine learning-based predictive modeling and optimization: Artificial neural network-genetic algorithm vs. response surface methodology for black soldier fly (Hermetia illucens) farm waste fermentation. Biochem Eng J 218: 109685.

105. Sun X, Sarmah N, Guo M (2025) Literature mining system for nutraceutical biosynthesis: From AI framework to biological insight. arXiv.

106. Gong X, Zhang J, Gan Q, Teng Y, Hou J, et al. (2024) Advancing microbial production through artificial intelligence-aided biology. Biotechnol Adv 74: 108399.

DOI: 10.19080/NFSIJ.2026.15.555918

How to cite this article: Songyu Z, Shanbin C, Jia Z, Wei L, Jian S, et al. AI-Driven Microbial Fermentation: Phenotype-Metabolite Dynamics. Nutri Food Sci Int J 2026. 15(4): 555918. DOI: 10.19080/NFSIJ.2026.15.555918