

Review

Artificial Intelligence in Yeast Biotechnology: Applications, Opportunities, and Challenges

Hossein Zakariapour Bahnamiri ¹, Alica Navrátilová ², Marek Kovár ³, Lucia Klongová ¹ and Miroslava Požgajová ^{1,*}

¹ AgroBioTech Research Center, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 01 Nitra, Slovakia; hossein.bahnamiri@uniag.sk (H.Z.B.); lucia.klongova@uniag.sk (L.K.)

² Institute of Nutrition and Genomics, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 01 Nitra, Slovakia; alica.navratilova@uniag.sk

³ Institute of Plant and Environmental Sciences, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 01 Nitra, Slovakia; marek.kovar@uniag.sk

* Correspondence: miroslava.pozgajova@uniag.sk; Tel.: +421-37-641-4922

Abstract

The adoption of artificial intelligence (AI) has been steadily growing across various fields in recent years. AI can have significant advantages for industries, including automation, reliable prediction, enhanced process control, and improved efficiency. Particularly, the field of academic research has embraced AI as a tool to address global challenges. Yeast strains, especially *Saccharomyces cerevisiae*, serve as key model organisms and industrial cell factories, enabling fundamental biological discoveries and advanced biotechnological applications through conserved eukaryotic pathways and modern metabolic engineering tools. Yeast cell biology can benefit from AI through its optimization of the fermentation process via an advanced control system, accelerated identification and classification of yeast strains through deep learning, predictive analysis of multi-omics biological data, strain engineering, and as a powerful approach in the evolution of the yeast genotype–phenotype map to explore novel biology. This review evaluates the benefits of using AI in biological studies leveraging yeast strains, discusses the challenges and limitations, and outlines the prospects.

Keywords: artificial intelligence; machine learning; yeast synthetic biology; phenotype prediction; metabolic engineering; fermentation optimization

Academic Editors: Dirk Tischler and Giovanni Barone

Received: 19 June 2026

Revised: 19 July 2026

Accepted: 23 July 2026

Published: 27 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Yeast cells are widely used as a model organism for studying eukaryotic cell biology and investigating the molecular mechanisms underlying a wide range of human diseases, as many fundamental eukaryotic genes, proteins, and cellular pathways are evolutionarily conserved between yeast and humans [1]. Moreover, yeast strains, especially *Saccharomyces cerevisiae*, are widely used as cell factories for producing fuels, beverages, foods, food ingredients, chemicals, pharmaceuticals, and vaccines [2,3] for use in the environmental, agricultural, medicine, energy, and materials sectors. Advances in metabolic engineering, aimed at improving cellular metabolism to enhance metabolite production, together with developments in synthetic biology, have led to the creation of microbial strains with enhanced capabilities to produce chemicals from renewable

feedstocks. These microbes, known as microbial cell factories (MCFs), were engineered to produce multiple specialized compounds, including pharmaceuticals, biofuels, food additives, and platform chemicals [2]. In most cases, the bacterium *E. coli* and the yeast *S. cerevisiae* are used as platform cell factories [4]. Despite success in developing many processes, the complete development of MCFs is usually time- and budget-consuming, requiring substantial effort to advance a proof-of-concept strain toward commercial production [5].

Historically, scientific discoveries were solely dependent on humans to generate hypotheses, which then underwent experimental validation and analysis. With advances in scientific research, even comparatively simple model organisms, such as *S. cerevisiae*, exhibit complexity far beyond what humans can reliably interpret or experimentally validate within a reasonable timeframe. To achieve reliable and scalable scientific outcomes under these conditions, acquiring integrated automated experimental and computational methods is essential. The time- and resource-intensive process of MCF development can be mitigated using genome-scale metabolic models (GEMs) and computational algorithms to identify non-intuitive gene-engineering targets that enhance production efficiency [6]. Methods for MCF design, including MOMA (Minimization of Metabolic Adjustment), which predicts metabolic responses to gene deletions; FSEOF (Flux Scanning based on Enforced Objective Flux), which identifies overexpression targets; optKnock (Optimization of Knockouts), which determines gene deletions that couple growth with product formation; and optForce (Optimization-based Force Identification), which identifies the set of genetic interventions required to achieve a desired production phenotype, can predict extensive lists of gene target candidates and generate ranked lists and prioritize suggestions; however, they lack integration with additional analysis [2].

The Design–Build–Test–Learn (DBTL) cycle is the foundation of synthetic biology methodology, which adopts engineering principles to ensure reproducibility and efficiency in biological research. Iterative DBTL biological engineering cycles enable researchers to test large-scale genetic designs and apply AI/Machine Learning (ML) to improve the design process [7]. All phases of the DBTL cycle can benefit from AI and ML applications. In the design phase, AI has the potential to support hypothesis generation, host organism selection, and the design of sgRNAs, metabolic pathways, and experiments. In the build phase, AI contributes to chromosome synthesis, genome editing, pathway engineering, and DNA assembly. During the test phase, AI facilitates sequencing, automation, screening, and experimental manipulation, while in the learn phase, it enhances omics data analysis, phenotype identification, and classification. Foundational pre-trained AI models, such as Large Language Models (LLMs), can serve as valuable assistants in creating training materials for new employees, optimizing protocols, and gathering biological information. Moreover, ML applications in synthetic biology can overcome the DBTL cycle's "learning" bottleneck by processing vast volumes of data and providing predictive models by choosing appropriate features to represent the phenomenon of interest and identify unseen patterns [8].

Wet-lab assays impose a high cost, where only a limited number of designs can be empirically tested, while AI-driven in silico exploration of thousands of genetic configurations enables the prioritization of high-confidence variants for validation [9]. In addition, AI can be utilized to evaluate the reliability of high-throughput screening data through the identification of anomalies from robotic errors or instrument drift and enhances standardization through domain-specific 'Construction File' (CF) formats encoding molecular biology operations, leading to automated detection of design flaws and validation of workflows across labs [10,11]. LLMs, PLMs (Protein Language Models), GEMs, BioCAD (Biological Computer-Aided Design), SBOL (Synthetic Biology Open

Language), and design space exploration have the potential to improve designing speed, predictive power, and design diversity. In addition, the acquisition of ML/AI integration (Bayesian optimization, deep learning (DL), active learning), LIMS (Laboratory Information Management System), and cloud-based feedback loops during the learn phase improved predictivity, cycle speed, and hypothesis generation [12].

The optimization of the fermentation process is an undeniable requirement to enhance the production of the desired compounds while mitigating waste, energy consumption, and environmental impact [13]. Effective management and monitoring of the fermentation process are essential in this context [14]. The complexity and dynamic interactions among microorganisms and fermentation variables are obstacles preventing mechanistic models from properly distinguishing the effects of various fermentation conditions and simulating the fermentation process [15,16]. Moreover, these models cannot predict the impact of factors beyond substrate concentration. In ML and AI, empirical data and specific algorithms are used on a computer to develop models with predictive functions. Acquiring ML methods can help in predicting fermentation yield, optimizing the fermentation process, and monitoring the fermentation process [17].

Given the broad applications of yeast as both a model organism for understanding eukaryotic biology and as a platform for the production of valuable bioproducts, improving yeast research and engineering strategies remains a major objective in biotechnology. Although several recent reviews have discussed AI applications in synthetic biology, biotechnology, or individual areas such as metabolic engineering and DBTL workflows, a comprehensive review dedicated specifically to yeast biotechnology has been lacking. This review addresses that gap by providing an integrated yeast-centered perspective on AI applications across multiple levels of biological research, including omics data analysis, microscopy-based phenotyping, phenotype prediction, metabolic engineering, and fermentation optimization (Figure 1). In addition, it discusses the opportunities offered by emerging AI technologies, the current limitations affecting their implementation, and future directions for AI-assisted yeast biotechnology.

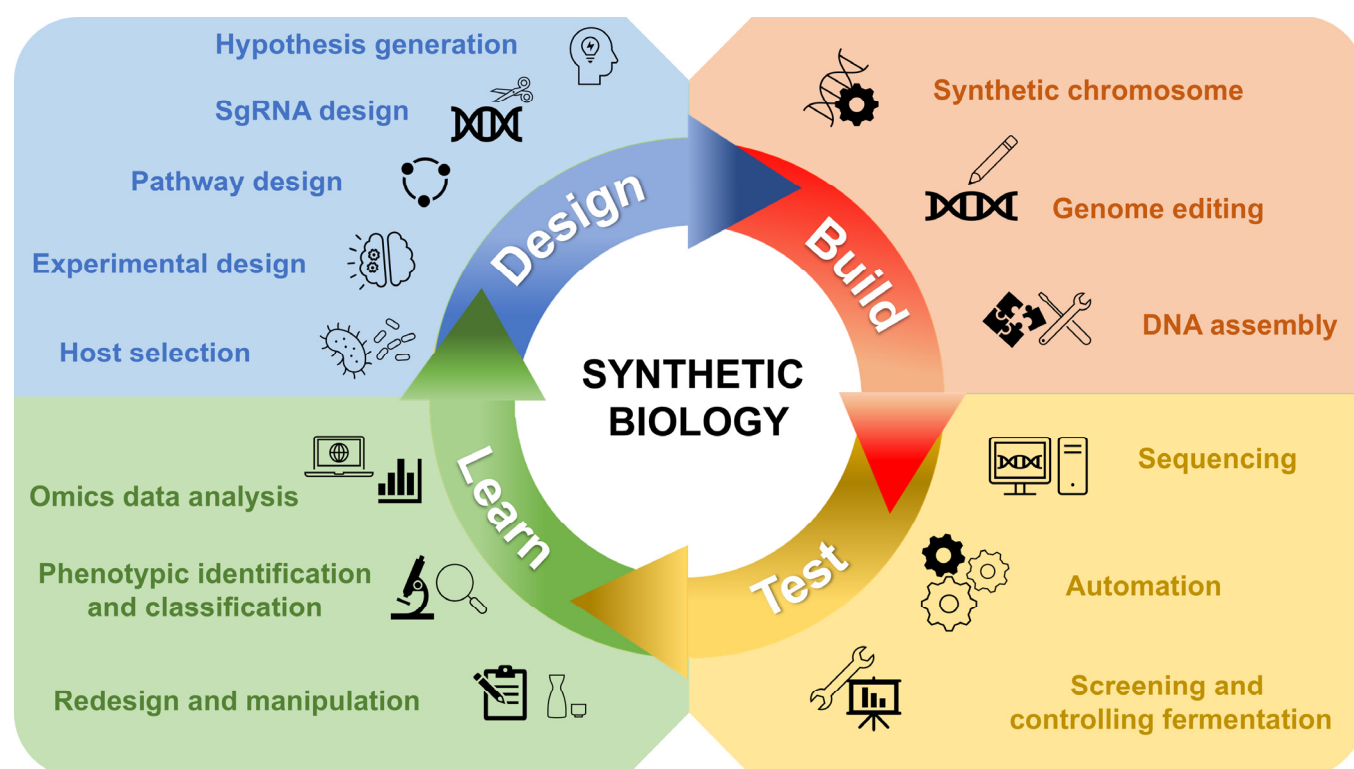


Figure 1. Integration of AI and ML across the Design–Build–Test–Learn (DBTL) cycle in yeast synthetic biology. AI/ML can support each stage of the workflow, from hypothesis generation, pathway and sgRNA design, and host selection (Design), to genome engineering and DNA assembly (Build), automated screening and sequencing (Test), and omics-driven analysis, phenotype prediction, and iterative strain optimization (Learn).

This narrative review was based on a comprehensive, non-systematic literature search conducted to identify relevant studies on the application of AI in yeast biotechnology and synthetic biology. The literature search was performed using the Web of Science, Scopus, PubMed, and Google Scholar databases, primarily covering publications from 2015 to 2026, while earlier landmark studies were included when necessary to provide historical context. The main search terms included combinations of artificial intelligence, Machine Learning, deep learning, yeast, synthetic biology, metabolic engineering, phenotype prediction, omics, microscopy, and Design–Build–Test–Learn (DBTL). The retrieved articles were screened based on their relevance to the scope of this review, scientific quality, and contribution to understanding AI applications in yeast research. Priority was given to original research articles and recent reviews describing methodological advances, experimental applications, and emerging trends in AI-assisted yeast biotechnology.

2. Applications of AI in Yeast Synthetic Biology and Biotechnology

2.1. AI Facilitates the Analysis and Interpretation of Yeast Genomics and Multi-Omics Data

The considerable advancements in high-throughput sequencing and multi-omics technologies have generated vast volumes of genomic, transcriptomic, proteomic, and metabolomic data in yeast research. While conventional statistical approaches, such as regression analysis and analysis of variance, remain useful for structured, low-dimensional datasets, they are often inadequate for capturing the complex nonlinear relationships embedded in large-scale omics data [18,19]. AI and ML approaches have emerged as powerful tools for extracting biologically meaningful patterns, inferring regulatory relationships, and integrating heterogeneous datasets. The studies that applied AI approaches to genomic and omics data analysis in yeast are summarized in Table 1.

Regarding the reconstruction and interpretation of gene regulatory networks, several studies have demonstrated that DL models can uncover complex regulatory relationships that are difficult to identify using traditional approaches. A distributed large-scale transcriptional regulatory neural network model (DLTRNM) developed for *S. cerevisiae* integrated prior biological knowledge with transcriptomic data to predict transcription factor–target gene interactions and identify key regulatory factors [20]. However, DLTRNM remains limited by the dynamic and condition-dependent nature of transcriptional regulation, the scarcity of comprehensive time-series transcriptomic datasets, and the limited interpretability of the models, highlighting the need for further experimental validation and improved model generalizability across diverse yeast conditions. Similarly, deep autoencoder models have been used to learn hierarchical representations of yeast transcriptional machinery, revealing biologically meaningful structures within gene expression data and partially reconstructing the hierarchical organization of transcriptional regulatory systems [21], suggesting the approach is applicable in more complex organisms. Compared to conventional autoencoders, sparse autoencoders generate more biologically meaningful latent representations by activating only a small subset of hidden units for each sample, which better reflects the sparse nature of transcriptional regulation. Furthermore, unlike NMF, the hierarchical architecture of sparse autoencoders captures regulatory information at multiple levels of abstraction, resulting in more interpretable latent features and improved identification of

transcriptional regulatory patterns. These studies illustrate how representation-learning approaches can move beyond simple correlation analyses to capture latent regulatory structures within high-dimensional transcriptomic datasets.

Recent advances have also expanded the use of AI from regulatory network inference toward sequence-based prediction and interpretation. The emergence of protein and DNA language models has enabled the direct extraction of biological information from sequence data without extensive feature engineering. Sakharova and Lareau [22] combined a Protein Language Model with large-scale experimental screening to investigate evolutionary constraints on synonymous codon usage, demonstrating that codon preferences can be predicted from protein sequence information and are strongly influenced by translational accuracy and cotranslational localization; however, it relies exclusively on protein sequence representations and therefore does not explicitly model nucleotide-level factors, such as mRNA structure or other sequence-specific regulatory constraints, which may also influence synonymous codon selection. Likewise, Dash and Bornelöv [23] trained deep neural networks on millions of synthetic and natural yeast promoter sequences to predict gene expression and uncover cis-regulatory grammar. Their prediction capability was shown to be affected by distribution shifts between synthetic training sequences and biologically relevant promoter sequences, highlighting the need for training datasets that better capture the diversity and complexity of natural regulatory landscapes. Together, these studies highlight a broader shift in computational biology from manually engineered sequence features to foundation model-based approaches that learn regulatory principles directly from biological sequences.

Another important sector is the application of AI for multi-omics integration. Biological phenotypes are rarely determined by a single molecular layer, making the integration of transcriptomic, proteomic, metabolomic, and regulatory information essential for system-level understanding. BRANEnet, a network embedding framework, demonstrated that biologically meaningful latent representations can be learned from heterogeneous multilayer biological networks, outperforming existing methods in transcription factor–target prediction, integrated network inference, and module detection [24]; however, their evaluation was limited to a yeast heat-shock multi-omics dataset, and further validation on diverse organisms and additional omics layers is required to assess its generalizability. Similarly, the AI-based framework iAMD integrated transcriptomic, proteomic, and phosphoproteomic data to identify novel regulators of autophagy in *S. cerevisiae* [25]. The performance of iAMD was compared with conventional ML algorithms, including logistic regression, Support Vector Machine (SVM), Gaussian naïve Bayes (GNB), and RF. iAMD outperformed these approaches in predicting Atg1-interacting partners and substrates, achieving higher AUC values and highlighting the advantage of its hybrid meta-learning framework for multi-omics-based prediction. Earlier studies, such as Mendes et al. [26], relied primarily on integrative analyses of transcriptomic and metabolomic data to investigate the molecular basis of aroma compound production across diverse yeast strains. Compared with these earlier approaches, recent AI-driven frameworks are increasingly capable of learning latent biological relationships directly from heterogeneous datasets rather than merely identifying statistical associations between molecular features.

The studies discussed above demonstrate that AI approaches have substantially expanded the capability of yeast genomics and multi-omics analysis beyond conventional statistical methods by enabling the identification of nonlinear relationships, hidden regulatory structures, and cross-layer molecular interactions. However, the suitability of different AI architectures depends strongly on the characteristics of the biological data and the research objective. DL and graph-based approaches are particularly valuable for reconstructing complex gene regulatory networks and integrating heterogeneous

biological interactions, whereas sequence-based foundation models, including protein and DNA language models, are more appropriate for extracting regulatory principles directly from large-scale sequence datasets. For multi-omics integration, network embedding and multimodal learning frameworks provide advantages by capturing latent relationships across molecular layers, while interpretable ML approaches remain important when datasets are limited or experimental validation is required. Despite these advances, most current AI models are trained and evaluated using a limited number of yeast strains and specific environmental conditions, which limits their generalizability to diverse industrial and biological contexts. Future efforts should therefore focus on developing larger, more standardized multi-omics datasets, improving model interpretability, and integrating AI predictions with systematic experimental validation to enable reliable transfer of computational models across yeast species, strains, and cultivation conditions.

Table 1. The studies used AI and ML in yeast strain genomics and omics data analysis alongside the reported benefits.

Organism	Tool/Model	Type	Conventional Approach	DBTL Role	Function/Application	Yeast Eng. Application	Performance Gain	Reference
<i>S. cerevisiae</i>	DLTRNM	Mechanism-informed DNN with transfer learning	Conventional TRN inference models (Boolean, ODE-based, and conventional ML approaches)	L	Reconstruction and prediction of TRNs by integrating curated TF-target interactions with pan-transcriptomic and time-series expression data	Regulatory network inference, identification of key transcription factors, systems-level modeling of yeast gene regulation, and support for synthetic biology circuit engineering	+ Prediction of regulatory correlations, + interpretability, – computational burden, + ability to identify key TFs and complement missing subnetworks	[20]
<i>S. cerevisiae</i>	PLM + codon prediction framework	DL for sequence-to-sequence evolutionary modeling	Conventional codon optimization approaches based on codon usage metrics and nucleotide-level features	L	Prediction of synonymous codon usage constraints from protein sequence to uncover evolutionary and translational rules governing codon choice	Codon optimization for heterologous expression, synthetic gene design, improved protein production strategies	54.5% overall prediction accuracy compared with 50% random baseline, + discovery of translational constraints, + generalization across genes and proteins	[22]
<i>S. cerevisiae</i>	Camformer	DL/ML model for promoter sequence-to-expression prediction	Motif-based and sequence feature-based promoter prediction approaches	L	Prediction of gene expression from promoter sequences to infer cis-regulatory logic and regulatory grammar	Promoter engineering, synthetic biology design of gene expression systems, and regulatory circuit optimization	+ Accuracy in predicting expression from sequence ($r^2 = 0.914 \pm 0.003$, $Q = 0.962 \pm 0.002$), + discovery of cis-regulatory rules, + scalable promoter activity modeling	[23]
<i>S. cerevisiae</i>	Sparse autoencoder	DL autoencoder	Conventional transcriptomic analysis using dimensionality reduction models	L	Learning hierarchical latent representations from microarray data; uncovering transcriptional regulatory structure	Functional genomics, transcription factor activity inference, gene regulatory network inference, systems biology analysis	+ Discovery of TF-associated latent variables, + representation of transcriptional programs vs. linear methods, + hierarchical biological interpretability of latent representations	[21]

<i>S. cerevisiae</i>	BRANEnet	Graph representation learning (random-walk + matrix factorization + SVD embedding)	Existing network embedding and multi-omics integration methods (OhmNet, Multi-Net, deepNF, BRANE-Exp, and MOSS)	L	Multi-omics integration and embedding of heterogeneous biological networks	TF-target prediction, integrated omics network inference, identification of heat-stress regulatory modules	+ Accuracy of TF-target prediction, Average AUPR improvement was ~10% compared with the best baseline methods for average and weighted L2 operators; weighted L2 AUPR was ~20% higher than BRANE-Exp, + module detection in multilayer omics networks, + scalable integration of heterogeneous omics data, + interpretability	[24]
<i>S. cerevisiae</i> (wine, sake, cachaça, laboratory strains)	PCA and MFA	Multi-omics integration + multivariate statistical modeling	–	L	Identification of gene expression–metabolite associations	Understanding aroma formation in fermentation; identification of candidate genes influencing ester and higher alcohol production	+ Discovery of genes linked to aroma compound synthesis, + identification of aroma-associated pathways	[26]
<i>S. cerevisiae</i>	iAMD framework	AI framework to infer autophagy regulators from multi-omic data	Conventional ML approaches (PLR, SVM, Gaussian Naïve Bayes, RF) and basic DNN models	L	Integration of transcriptomic, proteomic, and phosphoproteomic data to infer autophagy regulatory networks and protein interactions	Systems biology analysis of autophagy regulation, discovery of autophagy-associated genes/proteins	+ Identification of unknown autophagy regulators (AUC = 0.874; 95% CI = 0.820–0.922), improved AUC by >11.6% for Atg1-interacting partner prediction, + integration of heterogeneous omics datasets	[25]

DBTL, Design–Build–Test–Learn cycle; D, Design; B, Build; T, Test; L, Learn. Assignment to individual DBTL stages was based on the primary function of the AI application within the workflow. Applications primarily involving data interpretation, pattern recognition, or predictive modeling were assigned to the Learn stage, whereas those supporting experimental evaluation or phenotype measurement were assigned to the Test stage. DLTRNM, Distributed Large-Scale Transcriptional Regulatory Neural Network Model; PLM, Protein Language Model; BRANEnet, A multi-omics integration framework for multilayer

heterogeneous network; PCA, Principal Component Analysis; MFA, Multiple Factor Analysis; iAMD, inference of autophagy regulators from multi-omic data; DL, Deep learning; ML, Machine Learning; SVD, Singular Value Decomposition; TRN, transcriptional regulatory network; PLR, Penalized Logistic Regression; SVM, Support Vector Machine; RF, Random Forest; DNN, deep neural network; AUC, area under the curve; CI, confidence interval. Plus symbol (+) indicates enhancement or improvement of the evaluated characteristic. Minus symbol (−) indicates a reduction of the evaluated characteristic, such as labor, costs, processing time, or resource consumption.

2.2. Machine Learning Increases the Efficiency of Yeast Phenotype Prediction by Saving Time and Labor

The majority of biological traits are complex and regulated by multiple genetic loci, environmental factors, and their interactions. In organisms with large-scale genomic and phenotypic datasets available, genome-wide association studies (GWASs) have been widely used to identify genetic markers associated with phenotypic traits [27]. However, GWASs often have limited power to detect rare variants, small-effect loci, and higher-order interactions among genetic factors, resulting in an incomplete representation of the genetic architecture underlying complex traits [28]. To address these limitations, ML approaches have increasingly been employed to model complex genotype–phenotype relationships and to prioritize causal variants by integrating genomic, transcriptomic, proteomic, metabolomic, and environmental information [16–18]. Among ML methods, Random Forests (RF) have emerged as one of the most widely applied algorithms for biological data analysis due to their ability to handle high-dimensional datasets, capture nonlinear relationships, and identify informative features that contribute to phenotypic variation. Such capabilities make ML particularly suitable for predicting complex phenotypes in yeast, where traits often arise from interactions among numerous molecular and environmental factors [29]. Studies applying AI and ML for phenotype prediction, along with their advantages, are presented in Table 2.

One of the major applications of ML in yeast phenotype prediction is the identification of growth-associated traits from genomic and environmental data. Harrison et al. [30] trained an RF model using environmental, metabolic, and genomic features to predict the growth of *Saccharomycotina* species on different carbon sources. The model achieved high predictive performance while simultaneously identifying previously unrecognized features associated with growth on specific substrates. However, the success of these ML models was attributed to the availability of large-scale, well-annotated, and integrative datasets, as RF performance was improved dramatically by the unique combination of genomic, metabolic, and ecological data from the Y1000+ Project, which may limit the application of similar approaches in less-characterized lineages lacking comparable datasets. Moreover, the underlying genetic architecture and evolutionary history of the target trait are influential factors affecting ML models' predictive power. Traits with simple genetic architectures and repeated evolutionary origins may be more accurately predicted, whereas complex polygenic traits or traits shaped by diverse evolutionary mechanisms remain challenging for genotype–phenotype prediction [30]. Similarly, Khaiwal et al. [31] investigated the prediction of 223 phenotypic traits across 1011 genome-sequenced *S. cerevisiae* strains using genomic and transcriptomic information. By comparing four ML algorithms, including Ridge regression, Elastic regression (Elastic Net), Support Vector Regression (SVR), and Gradient-boosted trees (GBM), nonlinear models were superior to linear models in predicting quantitative phenotypes from genomics and gene expression predictors, with GBM as the most effective approach for phenotype prediction. This outperformance of GBM can be attributed to its ability to capture nonlinear relationships and interactions among genomic features underlying complex phenotypes. In addition, the prediction accuracy varied considerably among traits, with stress-resistance phenotypes being predicted more accurately than nutrient-utilization traits, highlighting differences in the genetic complexity of individual phenotypes. This difference may also originate from the fact that complex growth-related phenotypes are regulated by broader metabolic networks and higher-order gene interactions.

In addition to growth-related traits, ML has also been employed to predict metabolic phenotypes and cellular states. Zelezniak et al. [32] demonstrated that metabolite

abundances in *S. cerevisiae* could be accurately inferred from quantitative proteomic data of kinase knockout strains when many-to-many regulatory relationships between enzymes and metabolites were incorporated into the predictive framework and outperformed conventional one-to-one proteome-metabolome correlation analysis. These findings revealed a strong relationship between proteome composition and metabolic states, highlighting the potential of ML for connecting molecular measurements to cellular phenotypes. However, the approach was limited by its dependence on specific proteomic datasets and experimental conditions, which may affect its generalizability. Moreover, the predictive performance depends on the availability and quality of quantitative proteomic measurements, which may limit application when comprehensive proteomic datasets are unavailable.

These studies demonstrate that ML provides a powerful framework for improving yeast phenotype prediction by integrating diverse biological data types and capturing complex genotype–phenotype relationships beyond the capabilities of conventional approaches. The ability of ML models to identify informative features, model nonlinear interactions, and prioritize potential determinants of phenotypic variation highlights their potential for accelerating yeast strain characterization and engineering. However, the predictive success of these approaches remains highly dependent on the availability of comprehensive, high-quality datasets, appropriate representation of biological complexity, and the genetic architecture of the target traits. Therefore, further integration of multi-omics data, improved model interpretability, and validation across diverse yeast lineages will be essential for translating ML-based predictions into reliable biological insights and practical applications in strain development.

Table 2. The studies used AI and ML tools in yeast phenotype prediction and the benefits gained.

Organism	Tool/Model	Type	Conventional Approach	DBTL Role	Function/Application	Yeast Eng. Application	Performance Gain	Reference
<i>Saccharomyces</i>	RF	Supervised ML algorithm	Manual feature-based genotype–phenotype analysis	L	Prediction of carbon-source utilization from genomic, metabolic, and ecological features, identification of determinants of galactose metabolism	Strain engineering, carbon source utilization optimization, evolutionary biology of metabolism, identification of alternative metabolic pathways	+ Accuracy for growth phenotypes prediction, prediction accuracy of 92.2% using curated galactose genes and 93.3% when genomic and metabolic features were combined, + discovery of unknown galactose utilization pathway, + genotype–phenotype mapping	[30]
<i>S. cerevisiae</i>	GBM	Supervised ML algorithm (ensemble tree-based regression model)	Ridge regression, Elastic Net and SVR (benchmark models)	L	Prediction of 223 phenotypic traits from genomic variation across diverse yeast strains	Prediction of complex traits and prioritization of phenotype-associated variants for strain improvement	+ Phenotype prediction from genomic data, + identification of genotype–phenotype associations, + prioritization of causal variants and gene features	[31]
<i>S. cerevisiae</i>	ML regression framework (Elastic Net, RF, SVR, PLSR)	Supervised ML (multi-model regression)	Conventional one-to-one proteome–metabolome correlation analysis	L	Predicts metabolome from quantitative proteome data, inference of regulatory metabolic relationships	Mapping enzyme abundance to metabolic state; systems-level understanding of metabolic regulation	+ Prediction of metabolome changes from proteome (Median cross-validated $R^2 = 0.549$; 10-fold CV), + integration of proteomics and metabolomics datasets, + identification of regulatory metabolic signatures	[32]

DBTL, Design–Build–Test–Learn cycle; D, Design; B, Build; T, Test; L, Learn. Assignment to individual DBTL stages was based on the primary function of the AI application within the workflow. Applications primarily involving data interpretation, pattern recognition, or predictive modeling were assigned to the Learn stage, whereas those supporting experimental evaluation or phenotype measurement were assigned to the Test stage. RF, Random Forest; GBM, Gradient Boosting Machine; ML, Machine Learning; SVR, Support Vector Regression; PLSR, Partial Least Squares Regression; CV, Cross-Validation. Plus symbol (+) indicates enhancement or improvement of the evaluated characteristic.

2.3. Deep Learning Enhances the Accuracy and Speed of Yeast Identification and Classification Through Microscopy Imaging, Segmentation, and Automated Image Analysis

Yeasts are frequently present in food products and food-processing environments, where contamination by spoilage strains can lead to substantial economic losses. Since many yeast species can grow under diverse environmental conditions, ranging from ambient temperatures to refrigerated storage, rapid and reliable detection methods are essential for quality control and contamination management [33]. Conventional identification approaches typically rely on culturing visible colonies followed by biochemical or genetic characterization, a process that often requires several days and considerable labor [34]. Recent advances in DL and computer vision have transformed yeast microscopy analysis by enabling automated identification, segmentation, classification, and phenotypic characterization directly from imaging data [35]. The studies that applied AI and DL approaches for yeast imaging and classification are summarized in Table 3.

One of the most direct applications of DL in yeast microscopy is the rapid detection and classification of yeast colonies. Park et al. [35] developed a DL-based classification framework that combined conventional cultivation, white-light microscopy of microcolonies, and deep convolutional neural networks (CNNs) for rapid strain identification. Although the framework still relied on short-term cultivation to generate yeast microcolonies, the whole detection and classification process was reduced to 6 h. However, the classification accuracy varied among different strains, possibly due to a limited training dataset and potential limitations in the morphological differences between their microcolonies, including the overall smaller size of the microcolonies, which makes effective discrimination more challenging [36]. To improve model performance and overcome training data limitations, generative adversarial networks (GANs) were used to generate synthetic microcolony images, thereby enhancing classification accuracy and training efficiency by learning morphological characteristics at the image level and generating synthetic images that closely mimic the real data [37]. Similarly, Aiki et al. [38] employed a YOLO-based classifier for automated detection of *K. humilis* and *S. cerevisiae* colonies in sourdough cultures based on colony morphology and color characteristics with high accuracy and speed. However, missing counts, misclassification, and variable classification were mentioned as limitations of their classifiers. These studies demonstrate how DL can substantially reduce the time required for colony identification while enabling real-time monitoring of microbial population dynamics during fermentation processes.

Before the widespread adoption of DL for yeast microscopy analysis, classical ML approaches were also developed to automate cell morphology classification. One early study combined image processing with SVM, k-Nearest Neighbor (kNN), and Mahalanobis distance (GED) classifiers to identify yeast cell-cycle stages in microfluidic microscopy images based on handcrafted morphological features [39]. The proposed framework achieved a maximum classification accuracy of 82% using SVM while reducing the subjectivity associated with manual microscopy-based cell-cycle analysis. However, its performance relied heavily on handcrafted feature engineering and threshold-based segmentation, making it sensitive to overlapping cells, variations in illumination and focus, and dependent on manually labeled training data. These limitations subsequently motivated the development of DL-based segmentation and feature extraction methods.

Beyond colony recognition, a major focus of AI-assisted microscopy has been the accurate segmentation of individual yeast cells, which represents a critical prerequisite for quantitative image analysis. Traditional segmentation methods often struggle with densely packed cells, complex backgrounds, and imaging artifacts. Hence, several DL-

based approaches have been developed to overcome these challenges. Prangemeier et al. [40] applied a U-Net-based DL framework to precisely segment yeast cells in complex microstructured environments, enabling robust large-scale single-cell analysis. Compared with conventional image segmentation methods, the U-Net model achieved more accurate cell boundary detection and improved separation of densely packed cells while reducing the need for manual parameter optimization. However, the model relied on accurately annotated training datasets, and its applicability to different imaging conditions or experimental setups may require retraining or fine-tuning. Furthermore, the model was trained on a relatively narrow dataset representing a single microfluidic configuration, and its performance in more diverse microstructured environments remains to be established. Moreover, another U-Net-based model, YeastNet, was developed by Salem et al. [41], capable of generating accurate segmentation masks from bright-field microscopy images of *S. cerevisiae*. The model showed consistent performance across various focuses, demonstrating resistance to segmentation errors usually caused by changes in focus. YeastSpotter, introduced by Lu et al. [42], further simplified automated cell segmentation through a user-friendly and generalizable web-based platform, eliminating the need for extensive parameter optimization. In addition, Haja and Schomaker [43] proposed a fully automated end-to-end workflow that integrates segmentation, detection, and classification into a unified framework for large-scale fluorescence microscopy analysis with minimized human effort in training and operation. Collectively, these tools demonstrate the transition from manual image processing toward fully automated, high-throughput single-cell analysis pipelines.

Recent developments have expanded the role of DL from image preprocessing to the extraction of biologically meaningful phenotypes and cellular behaviors. Egebjerg et al. [44] combined soft X-ray tomography, fluorescence microscopy, and DL to investigate vacuole fusion and lipophagy in *S. cerevisiae*. Their CNN-based models, DYFNet and LipoSeg, enabled automated classification of vacuole fusion events and three-dimensional segmentation of lipid droplets and vacuoles, providing quantitative insights into intracellular lipid metabolism. However, the vacuole complex phenotype limited the DYFNet model from properly classifying various phenotypes, forcing the complex phenotype into a binary annotation of fully fused versus partially fused phenotypes. In the same way, Szücs et al. [45] developed a DL image-recognition pipeline for automated tetrad detection and classification of meiotic crossover events, enabling high-throughput analysis of chromosome segregation and recombination patterns without the need for tetrad dissection. At the proteome level, Litsios et al. [46] applied CNNs to millions of fluorescence microscopy images to investigate protein localization and abundance dynamics throughout the yeast cell cycle, revealing that a substantial proportion of the proteome exhibits periodic regulation. The networks produced ~20 subcellular localization classes. Although DeepLoc achieved 83.7% accuracy in the cell cycle-classified dataset, misclassification was observed between closely related localization classes, such as vacuole and vacuole periphery, and bud periphery. In combination with the findings of Egebjerg et al. [44], these observations indicate that classification of vacuole-associated features can remain challenging for DL approaches, particularly when phenotypes exhibit subtle morphological or spatial differences. These studies illustrate how AI is increasingly being used not only to identify cells but also to quantify complex biological processes and cellular phenotypes.

Another emerging trend is the use of generative AI to overcome limitations associated with annotated microscopy datasets. Reich [47] developed DL models (Cell-GAN) capable of generating realistic synthetic microscopy images of *S. cerevisiae*. Such synthetic datasets can augment training data, improve model robustness, and facilitate the development of more accurate segmentation and classification algorithms. The

integration of generative models with image analysis pipelines is expected to further accelerate the adoption of AI-driven microscopy in yeast research. Although GAN-based adversarial training can potentially improve future cell behavior prediction by enabling the generation of plausible future frames, these predictions may differ from the true future frames while still representing biologically valid outcomes [47].

The studies reviewed demonstrate that DL has evolved beyond a simple tool for yeast identification into a comprehensive framework for automated microscopy image analysis, encompassing detection, segmentation, classification, and quantitative phenotypic characterization. These approaches significantly reduce analysis time, improve accuracy, minimize manual intervention, and enable high-throughput analysis, highlighting their potential for applications in food quality control and yeast biology research. Nevertheless, the performance and generalizability of current models remain constrained by the availability of diverse, high-quality annotated datasets and by the difficulty of distinguishing subtle morphological and subcellular differences, particularly in complex phenotypes. The integration of generative AI, such as GAN-based synthetic image generation, represents a promising strategy to overcome data limitations and improve model robustness. Future advances are therefore expected to focus on standardized datasets, more generalizable and multimodal DL architectures, and explainable AI frameworks, facilitating the broader adoption of reliable AI-driven microscopy for yeast identification and characterization in both research and industrial settings.

Table 3. The studies used AI and ML in microscopy and imaging to identify, segment, and classify yeast strains.

Organism	Tool/Model	Type	Conventional Approach	DBTL Role	Function/Application	Yeast Eng. Application	Performance Gain	Reference
<i>S. cerevisiae</i>	SVM, kNN, GED	ML-based image classification and automated analysis	Manual identification of yeast cell-cycle phases based on bud presence and size	T	Automated segmentation, feature extraction, and classification of yeast cell-cycle phases	Automated single-cell phenotyping and cell-cycle monitoring	+ Automated yeast morphology classification, – subjectivity, classification accuracy up to 82% (SVM, 2D feature space), + robustness performance under variable illumination conditions, + classifier decision speed (SVM: 0.7 ms/sample)	[39]
<i>D. hansenii</i> , <i>W. anomalous</i>	Deep CNN, GAN augmentation	DL image classification model	Conventional yeast identification based on colony isolation followed by genetic or biochemical characterization	T	Rapid image-based yeast detection and phenotypic classification from microcolony microscopy	High-throughput automated classification and monitoring of spoilage yeasts	+ Classification accuracy (<i>mPrecision</i> of 96.0% and <i>mRecall</i> of 96.3%), + training efficiency, – detection time and labor (6 h)	[35]
<i>K. humilis</i> , <i>S. cerevisiae</i>	YOLO	DL object detection model	Conventional colony counting and yeast identification based on manual observation and morphological evaluation	T	Automated colony detection, counting, and species classification from plate images	Monitoring sourdough yeast mycobiota and fermentation stability during backslopping	+ Counting speed (126 milliseconds to count colonies on one Petri dish), + classification accuracy (>95%), – manual labor	[38]
<i>S. cerevisiae</i>	U-Net, Mask R-CNN	DL for semantic and instance segmentation	Traditional image segmentation approaches	T	Segmentation of yeast cells from microstructures in microfluidic microscopy environments	Automated single-cell monitoring and image analysis in microfluidic synthetic biology experiments	+ Segmentation accuracy (Sørensen-Dice coefficient = 0.96; mean Jaccard index = 0.89), + processing speed, + real-time monitoring capability, – manual image analysis	[40]
<i>S. cerevisiae</i>	YeastNet	U-Net-based CNN segmentation model	Classic non-trainable image segmentation algorithms and existing	T	Automatic segmentation, labeling, and tracking of yeast cells in bright-field microscopy images	High-throughput single-cell analysis, cell tracking, and automated microscopy workflows	+ Segmentation accuracy (10-fold CV of 0.833 ± 0.014), + robustness to focus variation, + tracking performance (tracking accuracy	[41]

yeast-cell segmentation tools					of 0.939 in focus), – dependency on labeled training data			
<i>S. cerevisiae</i>	DYFNet, LipoSeg	Deep CNN for image classification and segmentation	Manual annotation and quantitative analysis of microscopy images	T	Automated quantification of vacuole fusion, lipid droplet segmentation, and lipophagy analysis by fluorescence microscopy and soft X-ray tomography images	High-throughput phenotypic analysis of autophagy, vacuole dynamics, and lipid metabolism	+ Automated 3D analysis, + segmentation accuracy (8-fold CV of 0.8602), + throughput, – manual microscopy analysis	[44]
<i>S. cerevisiae</i>	DL tetrad classification pipeline	CNN-based image classification	Manual tetrad analysis and classification from fluorescence microscopy images	T	Automated detection and classification of fluorescent tetrads for meiotic recombination analysis	High-throughput quantification of crossover frequency, interference, chromosome missegregation, and gene conversion events	+ Throughput meiotic recombination analysis (accuracy = 0.94, precision = 0.95, recall = 0.96, F1 score = 0.9), + automated crossover quantification, – manual tetrad analysis	[45]
<i>S. cerevisiae</i>	CycleNet, DeepLoc	CNN-based image analysis	Manual annotation and microscopy-based classification of cell-cycle stages and protein localization	L	Cell-cycle phase classification and protein subcellular localization prediction from fluorescence microscopy images	Proteome-scale mapping of protein abundance and localization dynamics during the cell cycle	+ Automated classification of >20 cell-cycle/localization states (DeepLoc accuracy = 0.837; evaluated against manually annotated single-cell images), + identification of periodic protein localization and abundance changes	[46]
<i>S. cerevisiae</i>	GAN, future frame prediction model	DL generative model	Direct acquisition of time-lapse fluorescence microscopy data	L	Generation and simulation of time-lapse fluorescence microscopy images of yeast cells	Synthetic dataset generation, in silico modeling of yeast microscopy experiments, and support for training image analysis models	+ Synthetic data generation, + support for model training, – experimental effort, + simulation of microscopy experiments	[47]
<i>S. cerevisiae</i>	YeastSpotter	CNN-based segmentation model	Segmentation software and CellProfiler	T	Automatic segmentation of yeast cells in microscopy images	High-throughput segmentation, automated cell counting, and	Parameter-free operation, + robust segmentation across datasets (matched manual segmentation = 0.975; area	[42]

				without manual parameter tuning	morphological quantification	correlation = 0.969), – manual tuning effort	
Yeast strains	Mask R-CNN, YOLO	DL detection and instance segmentation model	Manual analysis of fluorescence microscopy images	T	Automated segmentation, detection, and classification of yeast cell compartments from fluorescence microscopy images	High-throughput subcellular phenotyping and automated fluorescence microscopy analysis	+ Automation, throughput, speed of image analysis (F1-score of 98% in terms of accuracy and speed), – manual annotation and image processing [43]

DBTL, Design–Build–Test–Learn cycle; D, Design; B, Build; T, Test; L, Learn. Assignment to individual DBTL stages was based on the primary function of the AI application within the workflow. Applications primarily involving data interpretation, pattern recognition, or predictive modeling were assigned to the Learn stage, whereas those supporting experimental evaluation or phenotype measurement were assigned to the Test stage. SVM, Support Vector Machine; kNN, k-nearest neighbor; GED, Mahalanobis distance classifier; CNN, Convolutional Neural Network; GAN, Generative Adversarial Network; DL, Deep learning; CV, Cross-Validation. Plus symbol (+) indicates enhancement or improvement of the evaluated characteristic. Minus symbol (–) indicates a reduction of the evaluated characteristic, such as labor, costs, processing time, or resource consumption.

2.4. AI Has the Potential to Improve Synthetic Biology and Metabolic Engineering by Enhancing the Efficiency and Comprehensiveness of the DBTL Cycle

The engineering of yeast strains for improved biomass formation and product synthesis remains challenging due to the complexity of eukaryotic metabolism, which involves numerous interconnected enzymatic reactions, tightly regulated gene expression programs, and condition-dependent metabolic flux redistribution. These characteristics often limit the predictive power of conventional engineering approaches and necessitate extensive experimental iterations. AI and ML have emerged as powerful tools for addressing these challenges by enabling predictive design, accelerating experimentation, and facilitating the integration of complex biological datasets. As a result, AI is increasingly being incorporated across all stages of the DBTL cycle, thereby improving the efficiency and comprehensiveness of synthetic biology and metabolic engineering for yeast. The AI and DL approaches used for yeast synthetic biology and metabolic engineering are summarized in Table 4.

In the design phase, AI enables the rational selection and optimization of genetic elements, pathways, and engineering strategies before experimental implementation. Recent advances in CRISPR-based technologies, including CRISPR nucleases, base editors, and prime editors, have significantly expanded genome engineering capabilities; however, variability in editing efficiency and off-target activity remains a major challenge [48]. AI-assisted approaches can improve guide RNA design, predict off-target effects, and optimize editing outcomes by learning from large experimental datasets [49]. An example of this strategy is DeepGuide, a DL framework developed for *Y. lipolytica* that predicts highly active CRISPR sgRNAs using genome-wide experimental data, thereby improving the efficiency of CRISPR-based genetic screens [50]; however, the prediction accuracy for low-activity Cas9 guides was reduced due to bias in the training dataset toward highly active guides, limiting the model's ability to learn features associated with poorly active guides. This highlights the importance of using a balanced and representative training dataset for optimal model training and improved prediction accuracy. Moreover, limited cross-species generalizability is a restriction for DeepGuide. The model requires both organism-specific training data and genome-specific unsupervised pre-training to capture the genomic context of the target organism for achieving optimal predictive performance [50]. In addition to gene editing, AI has also been applied to the design and evaluation of regulatory elements and metabolic engineering strategies. Polygraph, a Python-based framework, was developed to provide a systematic assessment of synthetic regulatory DNA sequences by integrating multiple computational approaches, including sequence analysis, motif identification, predictive neural network models, DNA language modeling, and sequence diversity metrics. This framework enables quantitative comparison and prioritization of engineered regulatory elements based on their predicted functional properties and similarity to native regulatory sequences, facilitating more informed design of synthetic regulatory elements [51]. Similarly, Yang et al. [52] developed Pymaker, an AI framework that leverages the pre-trained DNA language model DNABERT to predict promoter activity and design robust promoter sequences in *S. cerevisiae*. The pre-trained DNA models consistently outperformed non-pre-trained models, particularly when training data were limited, highlighting the advantage of transfer learning in synthetic biology. Importantly, their results showed that the effectiveness of pre-training depends on the relevance of the pre-training data to the target task, as DNA-specific models such as DNABERT and DNABERT2 outperformed BioBERT, which was trained primarily on biomedical text. By integrating promoter activity prediction with a mutation-based sequence generation approach, Pymaker enabled the identification of high-expression and mutation-resistant promoters, which were experimentally validated to enhance protein production.

Beyond regulatory element design, AI-assisted computational approaches have also been applied to optimize metabolic engineering strategies. Domenzain et al. [2] developed an enzyme-constrained metabolic modeling framework (ecFactory) based on ecYeastGEM to predict optimal gene modification strategies for enhancing chemical production in *S. cerevisiae*, facilitating the rational design of industrial platform strains. By incorporating enzyme capacity constraints into metabolic simulations, this approach improved the quantitative prediction of cellular behavior by accounting for enzyme-associated protein and substrate costs, enabling more realistic identification of engineering targets. Compared with conventional stoichiometric models, enzyme-constrained models substantially reduced the number of predicted gene targets, facilitating the selection of experimentally feasible modifications and supporting the development of versatile platform strains. However, the accuracy and broader applicability of these approaches were dependent on the availability of high-quality metabolic reconstructions and enzyme kinetic parameters, particularly for non-model organisms. The integration of DL-based prediction of enzyme parameters and expansion to diverse hosts represent important future directions for improving computational metabolic engineering.

The build phase focuses on implementing engineered designs through genome modification and strain construction. Although advances in CRISPR technologies have greatly simplified genome engineering, successful implementation often depends on selecting highly efficient guide RNAs and minimizing unintended genomic modifications. AI-assisted CRISPR platforms contribute to this process by predicting editing outcomes, improving target selection, and supporting the development of novel genome-editing systems with enhanced precision and efficiency [49]. Consequently, AI is increasingly serving as an enabling technology, improving the reliability and scalability of strain construction workflows.

AI is also transforming the test phase by enabling automated and real-time evaluation of engineered strains and bioprocess performance. Conventional growth monitoring relies heavily on optical density measurements, which are affected by light scattering, turbidity, and the inability to distinguish viable from nonviable cells. To overcome these limitations, AlMarzooqi et al. [53] developed an AI-enhanced electrochemical impedance spectroscopy platform for real-time monitoring of *S. cerevisiae* growth. By integrating impedance biosensing with Gaussian Process Regression, the system accurately predicted growth dynamics while providing uncertainty-aware estimates suitable for continuous process monitoring. The proposed framework also established the foundation for a real-time digital twin, enabling predictive monitoring and supporting future adaptive control strategies for intelligent fermentation processes, thereby facilitating continuous data acquisition and strengthening the transition from the Test to the Learn phase of the DBTL cycle.

The learning phase is one of the most influential applications of AI in metabolic engineering, as it converts large-scale experimental data into actionable biological knowledge. AI models trained on transcriptomic, metabolomic, proteomic, and phenotypic datasets can uncover genotype–phenotype relationships that are difficult to identify using conventional statistical approaches. DL frameworks were used to predict metabolic flux distributions in engineered *S. cerevisiae*, enabling more accurate characterization of cellular metabolic states and supporting iterative strain optimization [54]. Similarly, hybrid mechanistic-ML models integrating genome-scale metabolic and kinetic models with transformer-based architectures were developed to identify pathway bottlenecks and prioritize engineering targets [55]. AI-driven metabolic engineering frameworks that combine genome-scale metabolic modeling with ML were applied to optimize biomass production in *S. cerevisiae*. Nair and D’Souza [56] developed an

integrative framework combining the Yeast9 genome-scale metabolic model with ML and optimization approaches to analyze and optimize single-cell protein production in *S. cerevisiae*. Models including RF, XGBoost, feed-forward neural networks, variational autoencoders, and generative adversarial networks were employed to predict biomass flux, identify key metabolic reactions, explore metabolic patterns, and generate candidate flux configurations. SHAP-based interpretation and Bayesian optimization further enabled identification of important metabolic drivers and favorable cultivation conditions. These studies highlight the potential of AI-integrated metabolic modeling to extract biological insights from complex datasets and provide computational guidance for future strain engineering strategies. However, these AI-driven metabolic frameworks remain limited by the assumptions of genome-scale metabolic models, such as steady-state constraints and incomplete representation of enzyme kinetics and regulatory dynamics. Furthermore, the predicted metabolic strategies require experimental validation, and generative approaches such as GANs still face challenges in producing biologically feasible solutions.

The developments in AI were extended beyond individual DBTL stages toward automating the entire engineering cycle. Brunnsåker et al. [57] developed a hybrid AI framework integrating Large Language Models, relational learning, inductive logic programming, metabolomics analysis, and automated laboratory systems for *S. cerevisiae*. By combining hypothesis generation, experimental prioritization, logical reasoning, and robotic experimentation within a closed-loop workflow, the framework enabled the potential of self-driving laboratories to accelerate hypothesis-driven biological research. Such systems represent an important step toward self-driving laboratories capable of continuously generating hypotheses, designing experiments, interpreting results, and refining biological knowledge with minimal human intervention. However, the current workflow still requires limited human intervention, and further advances are needed to increase hypothesis complexity, prioritize experiments, and fully automate laboratory processes.

The discussed studies demonstrate that AI is becoming an integral component of the DBTL cycle in yeast synthetic biology and metabolic engineering, enabling more rational design, efficient strain evaluation, and data-driven learning. Across different stages, AI approaches have improved the prediction of genetic element function, the optimization of metabolic strategies, the real-time monitoring of engineered strains, and the extraction of biological insights from complex datasets. However, the successful implementation of AI-driven engineering remains constrained by several factors, including limited availability of high-quality biological datasets, challenges in model transferability across organisms and conditions, incomplete representation of biological complexity in computational models, and the continued need for experimental validation. Future advances will likely arise from the integration of AI with mechanistic models, multi-dimensional biological data, and automated experimental platforms, ultimately enabling more predictive, scalable, and iterative DBTL workflows for yeast engineering.

Table 4. AI-assisted synthetic biology and metabolic engineering of yeast.

Organism	Tool/Model	Type	Conventional Approach	DBTL Role	Function/Application	Yeast Eng. Application	Performance Gain	Reference
<i>Y. lipolytica</i>	DeepGuide	DL (CNN, unsupervised representation learning)	Existing sgRNA activity prediction tools	L	Prediction of high-activity CRISPR-Cas9 and Cas12a guide RNAs using genomic context, sequence features	Improved CRISPR guide selection for yeast genome engineering	+ Accuracy in guide efficiency prediction (Pearson coefficients of 0.5 and 0.66 for Cas9 and Cas12a cutting score values), + genome editing success rate, + scalable guide design	[50]
<i>S. cerevisiae</i>	RF, XGBoost, VAE, GAN	Hybrid ML/DL framework	Genome-scale metabolic modeling using Yeast9 and FBA	L	Prediction and optimization of biomass/single-cell protein production using genome-scale metabolic simulation	Metabolic engineering for biofuel and single-cell protein production, strain optimization, flux reprogramming, and nutrient optimization	+ Predictive accuracy for biomass flux (RF $R^2 = 0.999$; XGBoost $R^2 = 0.999$), + discovery of key metabolic drivers, + generation of feasible novel flux states via GAN	[56]
<i>S. cerevisiae</i>	CellFactory-ecYeastGEM pipeline	Genome-scale metabolic modeling (enzyme-constrained GEM)	–	D	Prediction of metabolic engineering targets and gene-modification strategies for chemical production	Metabolic engineering of yeast cell factories for chemical production, platform strain design, and multi-product strain optimization	Systematic identification of gene targets across 103 chemicals, + rational design of platform strains, + prediction of metabolic engineering interventions, + scalable chemical production pathway discovery	[2]
<i>S. cerevisiae</i>	GPR	Supervised ML for biosensor signal analysis	Optical density measurement (OD600) and conventional optical monitoring of yeast growth	T	Quantification of yeast cell concentration from impedance biosensor signals using ML-based signal decoding and calibration models	Real-time monitoring of yeast growth in fermentation and bioprocessing systems, biosensor-driven process optimization, and non-invasive biomass estimation	+ Accuracy in biomass estimation from raw impedance signals (RMSE = 0.79 min; $R^2 = 0.999$; $r = 0.999$), + real-time monitoring capability (classification accuracy of 100% for 15 min growth intervals), – manual sampling in fermentation monitoring	[53]
<i>S. cerevisiae</i>	Agentic AI framework	LLM, symbolic relational learning	Manual hypothesis generation based on	L	Automated hypothesis generation, experimental design, execution, and	Autonomous discovery of metabolic interactions	+ Hypothesis generation and validation cycles, + experimental throughput via automation, +	[57]

			literature review and expert knowledge		iterative learning from experimental outcomes	and systems biology knowledge	iterative refinement of biological hypotheses, + integration of AI reasoning with wet-lab feedback	
S. <i>cerevisiae</i>	Polygraph	Regulatory sequence evaluation framework	Manual evaluation of designed regulatory sequences based primarily on known regulatory elements	D	Systematic assessment and optimization of synthetic regulatory DNA elements using predictive neural network models, motif analysis, and guided sequence evolution	Synthetic promoter engineering, regulatory sequence optimization	+ Evaluation of synthetic promoters, + generation of more biologically realistic high-activity promoters, + sequence diversity analysis	[51]
S. <i>cerevisiae</i>	Pymaker (built on DNABERT)	Pre-trained transformer for DNA sequences	Conventional promoter screening and experimental promoter optimization	D	Prediction of promoter expression levels and generation of novel promoter sequences	Synthetic promoter design for gene expression optimization and metabolic control	+ Accuracy than non-pre-trained models, + promoter design efficiency, + protein expression compared to conventional promoters	[52]

DBTL, Design–Build–Test–Learn cycle; D, Design; B, Build; T, Test; L, Learn. Assignment to individual DBTL stages was based on the primary function of the AI application within the workflow. Applications primarily involving data interpretation, pattern recognition, or predictive modeling were assigned to the Learn stage, whereas those supporting experimental evaluation or phenotype measurement were assigned to the Test stage. RF, Random Forest; VAE, Variational Autoencoder; GAN, Generative Adversarial Network; GEM, Genome-scale metabolic model; GPR, Gaussian Process Regression; DL, Deep learning; CNN, Convolutional Neural Network; ML, Machine Learning; LLM, Large Language Model; FBA, Flux Balance Analysis; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats. Plus symbol (+) indicates enhancement or improvement of the evaluated characteristic. Minus symbol (–) indicates a reduction of the evaluated characteristic, such as labor, costs, processing time, or resource consumption.

2.5. AI-Driven Fermentation Optimization Can Lead to More Efficient Metabolite Production, Reduced Wastage, and Lower Environmental Impacts

Fermentation is a highly dynamic and multifactorial process in which environmental and operational variables, including temperature, pH, oxygen availability, nutrient composition, and feeding strategy, collectively determine yeast metabolism and ultimately influence product yield, quality, and process scalability. Although conventional kinetic models remain valuable for describing substrate consumption and microbial growth, they are generally insufficient to capture the complex nonlinear interactions governing industrial fermentation systems. Consequently, AI and ML have emerged as powerful tools for modelling these complex relationships, enabling accurate prediction of fermentation behaviour, optimization of operating conditions, and implementation of intelligent control strategies that improve productivity while reducing waste and environmental impact [58,59]. Among industrial applications, the brewing sector is one of the most advanced examples of AI-assisted fermentation, in which ML methods are routinely used to improve product consistency, process efficiency, and operational reliability [60]. The AI and DL approaches used in yeast fermentation monitoring and optimization, together with their corresponding improvements, are summarized in Table 5.

A major advancement enabled by AI is the transition from conventional offline measurements toward continuous, non-invasive, and real-time fermentation monitoring. Rather than relying on individual analytical measurements, modern AI frameworks integrate information from multiple sensing platforms to construct soft sensors capable of estimating key process variables with high accuracy while simultaneously providing insight into yeast physiology and metabolic activity [58]. Spectroscopic approaches combined with ML algorithms have proven particularly effective for predicting glucose, ethanol, glycerol, biomass, and other fermentation-related variables directly from spectral information, achieving high predictive accuracy while eliminating the need for destructive sampling [3]. Likewise, ultrasonic sensing integrated with ANN and DL architectures, particularly LSTM networks, has demonstrated the ability to continuously estimate alcohol concentration throughout fermentation, highlighting the advantage of temporal neural networks in capturing the dynamic behavior of biological processes. However, prediction errors were still observed at certain fermentation stages, and larger training datasets and validation under industrial-scale conditions were suggested as requirements to improve model robustness [61]. Collectively, these approaches demonstrate that AI-driven soft sensing not only accelerates data acquisition but also provides reliable process information suitable for real-time industrial control.

Computer vision has emerged as another important AI-assisted monitoring strategy by enabling direct observation of fermentation phenomena without physical sampling. Image analysis combined with CNNs allows automated detection of CO₂ bubble formation and accurate quantification of fermentation progress, facilitating precise identification of fermentation endpoints while simultaneously capturing the influence of operational variables such as yeast inoculum and glucose concentration on fermentation kinetics [62]. Compared with conventional sensor-based monitoring, vision-based systems provide an inexpensive and highly automated alternative capable of supporting continuous process supervision and reducing operator intervention.

Beyond monitoring process variables, AI has increasingly been employed to characterize cellular physiology, metabolic activity, and product formation through the integration of multimodal analytical platforms. Recent study into oleaginous yeast *Rhodotorula mucilaginosa* combining Field Emission Scanning Electron Microscopy (FESEM), Transmission Electron Microscopy (TEM), Raman spectroscopy, Liquid Chromatography–Mass Spectrometry (LC-MS), and Thin-Layer Chromatography (TLC)

with ML algorithms has demonstrated that CNNs, RF, SVM, and ANN models can accurately classify cell morphology, characterize lipid composition, and quantify intracellular lipid accumulation with high predictive performance [63]. The integration of complementary analytical techniques substantially improves model robustness by extracting biologically meaningful features that cannot be obtained from individual analytical methods alone, illustrating the growing importance of multimodal AI frameworks in industrial biotechnology. Nevertheless, the generalizability of these approaches remains constrained by relatively small datasets, limited species diversity, and insufficient external validation, emphasizing the need for larger multi-omics datasets and standardized validation protocols [63].

In parallel with process monitoring, ML has become an effective tool for predicting final product quality directly from fermentation data. Various nonlinear learning algorithms, including back-propagation neural networks (BP-NN), radial basis function neural networks (RBF-NN), and least squares support vector machine (LS-SVM), consistently outperform conventional linear regression methods when modelling complex relationships between fermentation conditions and quality attributes such as acetic acid concentration [64]. Although neural networks generally provide excellent predictive capability, their performance often depends on large training datasets and may suffer from limited reproducibility because of random weight initialization. In contrast, SVM-based approaches frequently demonstrate superior robustness and generalization when applied to relatively small, nonlinear, and high-dimensional industrial fermentation datasets [64]. These findings indicate that algorithm selection should be guided not only by prediction accuracy but also by dataset characteristics, computational requirements, and industrial implementation constraints.

The reviewed studies demonstrate a clear evolution in the application of AI for yeast fermentation, shifting from the prediction of individual fermentation parameters toward integrated platforms capable of real-time monitoring, process optimization, and intelligent process control. Conventional ML algorithms have shown robust performance for modelling nonlinear relationships and predicting fermentation variables and product quality, whereas DL models have expanded AI capabilities by enabling automated analysis of image, spectral, and time-series data. Furthermore, the integration of advanced sensing technologies with multimodal AI frameworks has improved predictive accuracy and provided deeper insights into fermentation dynamics and yeast physiology. Despite these advances, broader industrial implementation still requires larger and more diverse datasets, standardized validation, and improved model transferability. Overall, AI is becoming an essential tool for developing more autonomous, efficient, and sustainable fermentation processes in industrial yeast biotechnology.

Table 5. The studies used AI and ML tools in yeast fermentation process optimization and the benefits gained.

Organism	Tool/Model	Type	Conventional Approach	DBTL Role	Function/Application	Yeast Eng. Application	Performance Gain	Reference
<i>S. cerevisiae</i>	ML-assisted Raman spectroscopy	ML-based Raman spectroscopy modeling	Conventional Raman spectroscopy using PLS regression and offline fermentation parameter measurements	T	Real-time prediction of fermentation parameters from Raman spectra	Real-time monitoring and control of yeast fermentation processes	+ Prediction accuracy of fermentation metabolites and biomass ($R^2 = 0.89$ – 0.97)	[3]
<i>S. cerevisiae</i>	CFMM, RFQS	DL-based computer vision	Conventional fermentation monitoring (MQ-3 alcohol sensor, bubble-counting sensor)	T	Automated image-based detection of fermentation progress and endpoint determination	Real-time monitoring and quantitative assessment of yeast fermentation	+ Real-time quantitative fermentation monitoring, + real-time fermentation endpoint detection accuracy	[62]
<i>S. cerevisiae</i>	Neural network	A supervised ML technique	Conventional fermentation monitoring methods (Raman spectroscopy)	T	Prediction of metabolite production from yeast morphological features	AI-assisted fermentation monitoring and forecasting	+ Accuracy of forecasting metabolite yield ($R^2 = 0.9$), + speed of forecasting (coefficient determination > 0.9)	[65]
<i>Rhodotula mucilaginosa</i>	CNNs, RF, SVM	ML models	Conventional lipid characterization methods (FESEM, TEM, Raman spectroscopy, LC-MS, and TLC analyzed independently)	T	Automated lipid classification and characterization	AI-assisted lipid profiling and characterization of oleaginous yeast	+ Accuracy of morphological classification (95%), spectral analysis (91.4%), and lipid quantification (97%)	[63]

<i>S. cerevisiae</i>	LSTM, ANN	DL models	Periodic sampling and offline density/alcohol measurements	T	Real-time prediction of alcohol concentration during fermentation	AI-assisted fermentation monitoring and process control	+ Accuracy of alcohol prediction ($R^2 = 0.952$), + non-invasive monitoring ($R^2 = 0.948$ without temperature sensor), – manual sampling	[61]
<i>S. cerevisiae</i>	BP-NN, RBF-NN, LS-SVM	ML models	PLS	T	Prediction of acetic acid concentration in the final beer fermentation process	AI-assisted monitoring, quality prediction, and optimization of beer fermentation	+ improvement in reliability of prediction for both the calibration set and the validation in RBF-NN and LS-SVM.	[64]

DBTL, Design–Build–Test–Learn cycle; D, Design; B, Build; T, Test; L, Learn. Assignment to individual DBTL stages was based on the primary function of the AI application within the workflow. Applications primarily involving data interpretation, pattern recognition, or predictive modeling were assigned to the Learn stage, whereas those supporting experimental evaluation or phenotype measurement were assigned to the Test stage. ML, Machine learning; CFMM, Convolutional Fermentation Monitoring Model; RFQS, Real-time Fermentation Quantification System; CNN, Convolutional Neural Network; RF, Random Forest; SVM, Support Vector Machine; LSTM, Long Short-Term Memory; ANN, Artificial Neural Network; BP-NN, Back-propagation neural network; RBF-NN, Radial basis function neural network; LS-SVM, least squares Support Vector Machine; DL, Deep learning; PLS, Partial Least Squares Regression; FESEM, Field Emission Scanning Electron Microscopy; TEM, Transmission Electron Microscopy; LC-MS, Liquid Chromatography–Mass Spectrometry; TLC, Thin-Layer Chromatography; GBDT, Gradient Boosting Decision Tree. Plus symbol (+) indicates enhancement or improvement of the evaluated characteristic. Minus symbol (–) indicates a reduction of the evaluated characteristic, such as labor, costs, processing time, or resource consumption.

2.6. Challenges and Limitations

Despite the substantial contributions of AI in improving the efficiency and predictive capability of synthetic biology workflows, several technical and sociological challenges continue to limit its widespread implementation [66]. Among technical challenges, the availability of high-quality, standardized, and sufficiently annotated biological datasets remains one of the most significant limitations. Biological data are often generated under different experimental conditions, stored in heterogeneous formats, and distributed across multiple platforms, complicating data integration and reducing the reliability of AI model training. Therefore, effective data engineering, including data curation, normalization, and standardization, is essential for developing robust and reproducible AI-driven biological models [67]. Furthermore, many biological datasets, particularly omics and imaging datasets, are characterized by high dimensionality but limited numbers of labelled samples, increasing the risk of overfitting and reducing model generalizability [68,69]. Approaches such as data augmentation can partially address limited dataset size in specific applications; however, they cannot fully replace the need for biologically representative training data [70,71]. Training large DL and foundation models often requires substantial computational resources, including high-performance GPUs, large memory capacity, and considerable training time. These requirements may limit access to advanced AI approaches for many academic laboratories and increase the environmental and financial costs associated with model development.

Beyond data limitations, AI models face challenges related to interpretability, uncertainty estimation, and biological complexity. Many ML models provide accurate predictions but offer limited insight into the underlying biological mechanisms, which is particularly problematic in complex systems characterized by stochastic behaviour and context-dependent regulation. Moreover, conventional performance metrics, such as prediction accuracy or regression scores, may not fully reflect whether a model captures biologically meaningful relationships. The limited ability of AI models to extrapolate beyond their training domain represents another major challenge. For example, DeepGuide demonstrated that optimal performance required organism-specific training data and genomic context, indicating that models developed for one species may require additional training before application to other organisms [50]. Biological datasets are also frequently affected by class imbalance, in which rare phenotypes, mutations, or cellular states are substantially underrepresented compared with common classes. Such an imbalance can bias model training toward dominant classes, reducing predictive performance for biologically important but infrequent events. Similarly, neural network-based models may perform poorly under biological conditions substantially different from those represented in the training dataset, highlighting the importance of evaluating model transferability across strains, environments, and experimental conditions [50,67]. Models developed using laboratory-scale datasets may not be well applicable to industrial fermentation processes, where cultivation conditions, strain backgrounds, media composition, and process variability differ considerably.

In addition to predictive limitations, AI-driven biological design raises important biosafety and reliability considerations. AI-generated genetic sequences or engineering strategies may produce unintended biological effects that are difficult to predict, while biases within training datasets can compromise model reliability in unexplored biological contexts. Consequently, rigorous experimental validation, transparent model development, and appropriate biosafety assessments remain essential for responsible application of AI in synthetic biology [52]. The limited transparency of many DL models also presents another challenge in regulatory and industrial settings, where decisions related to strain development, food production, or biotechnology applications often require interpretable and traceable evidence.

Successful adoption of AI in synthetic biology requires addressing sociological and interdisciplinary challenges. Computational scientists and experimental biologists often differ in their training backgrounds, research priorities, and approaches to problem-solving, which may create communication barriers and limit effective collaboration. While computational approaches emphasize automation, scalability, and predictive modelling, experimental researchers frequently prioritize biological feasibility, practical implementation, and mechanistic understanding. Addressing these challenges will be critical for the reliable and scalable implementation of AI-driven synthetic biology.

2.7. Future Perspectives

The integration of AI and synthetic biology is expected to progress from predictive modelling toward increasingly autonomous and data-driven biological engineering. Rather than serving solely as analytical tools, future AI systems are likely to support the entire DBTL cycle by enabling inverse biological design, automated experimental planning, real-time data analysis, and adaptive decision-making. As biological datasets continue to expand in both size and complexity, AI will play an increasingly important role in accelerating strain engineering, optimizing bioprocesses, and improving the efficiency and reproducibility of yeast biotechnology [67].

A key priority for future research is the development of more robust and generalizable AI models. This will require greater availability of high-quality, standardized, and well-annotated biological datasets encompassing diverse strains, cultivation conditions, and analytical platforms. Emerging approaches such as self-supervised learning, continual learning, transfer learning, and foundation models offer promising opportunities to reduce dependence on large manually labelled datasets while improving model adaptability across different biological systems and experimental conditions.

Another important direction is the integration of multimodal biological data with mechanistic knowledge. Combining genomic, transcriptomic, metabolomic, proteomic, imaging, and fermentation data within unified AI frameworks, together with genome-scale metabolic and kinetic models, is expected to improve the prediction of genotype–phenotype relationships while increasing model interpretability and biological relevance. Such hybrid approaches could facilitate more reliable biological design and support rational optimization of engineered yeast strains.

Future advances are also expected to accelerate the implementation of AI in industrial biotechnology. The integration of AI with advanced sensing technologies, automated bioreactors, robotics, and digital twin platforms will enable real-time monitoring, closed-loop process control, and autonomous optimization of fermentation processes. In parallel, improvements in computational efficiency and high-throughput microscopy will facilitate online single-cell analysis, adaptive cultivation strategies, and rapid strain characterization. Collectively, these developments are expected to transform AI from a predictive tool into a core component of intelligent, autonomous, and sustainable yeast biotechnology.

3. Conclusions

Yeast synthetic biology is increasingly being transformed by AI, which has expanded from a data analysis tool to a key enabling technology across the DBTL cycle, strain engineering, microscopy, metabolic engineering, and fermentation optimization. By integrating ML with experimental biology, AI has improved predictive accuracy, accelerated biological discovery, enhanced process automation, and increased the efficiency of industrial yeast biotechnology. Despite these advances, challenges related to data quality, model interpretability, generalizability, and interdisciplinary collaboration

continue to limit broader implementation. Addressing these challenges through improved data resources, explainable AI, and closer integration of computational and experimental approaches is essential for translating AI into reliable and scalable biological applications. Overall, the convergence of AI and synthetic biology is expected to drive the next generation of intelligent, autonomous, and sustainable yeast biotechnology.

Author Contributions: Conceptualization, H.Z.B., M.K., and M.P.; validation, A.N. and L.K.; writing—original draft preparation, H.Z.B., M.P., and M.K.; writing—review and editing, A.N. and L.K.; visualization, H.Z.B.; supervision, M.P.; project administration, H.Z.B. and M.P.; funding acquisition, H.Z.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00379.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gaikani, H.K.; Stolar, M.; Kriti, D.; Nislow, C.; Giaever, G. From Beer to Breadboards: Yeast as a Force for Biological Innovation. *Genome Biol.* **2024**, *25*, 10. <https://doi.org/10.1186/s13059-023-03156-9>.
2. Domenzain, I.; Lu, Y.; Wang, H.; Shi, J.; Lu, H.; Nielsen, J. Computational Biology Predicts Metabolic Engineering Targets for Increased Production of 103 Valuable Chemicals in Yeast. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2417322122. <https://doi.org/10.1073/pnas.2417322122>.
3. Wu, D.; Xu, Y.; Xu, F.; Shao, M.; Huang, M. Machine Learning Algorithms for In-Line Monitoring during Yeast Fermentations Based on Raman Spectroscopy. *Vib. Spectrosc.* **2024**, *132*, 103672. <https://doi.org/10.1016/j.vibspec.2024.103672>.
4. Yan, X.; He, Q.; Geng, B.; Yang, S. Microbial Cell Factories in the Bioeconomy Era: From Discovery to Creation. *Biodes. Res.* **2024**, *6*, 0052. <https://doi.org/10.34133/bdr.0052>.
5. Nielsen, J. Yeast Cell Factories on the Horizon. *Science* **2015**, *349*, 1050–1051. <https://doi.org/10.1126/science.aad2081>.
6. Nielsen, J. Systems Biology of Metabolism. *Annu. Rev. Biochem.* **2017**, *86*, 245–275. <https://doi.org/10.1146/annurev-biochem-061516-044757>.
7. Matzko, R.; Konur, S. Technologies for Design-Build-Test-Learn Automation and Computational Modelling across the Synthetic Biology Workflow: A Review. *Netw. Model. Anal. Health Inform. Bioinforma.* **2024**, *13*, 22. <https://doi.org/10.1007/s13721-024-00455-4>.
8. Kitano, S.; Lin, C.; Foo, J.L.; Chang, M.W. Synthetic Biology: Learning the Way toward High-Precision Biological Design. *PLoS Biol.* **2023**, *21*, e3002116. <https://doi.org/10.1371/journal.pbio.3002116>.
9. Ramos, M.C.; Collison, C.J.; White, A.D. A Review of Large Language Models and Autonomous Agents in Chemistry. *Chem. Sci.* **2025**, *16*, 2514–2572. <https://doi.org/10.1039/D4SC03921A>.
10. Ataii, N.; Bakshi, S.; Chen, Y.; Fernandez, M.; Shao, Z.; Scheftel, Z.; Tou, C.; Vega, M.; Wang, Y.; Zhang, H.; et al. Enabling AI in Synthetic Biology through Construction File Specification. *PLoS ONE* **2023**, *18*, e0294469. <https://doi.org/10.1371/journal.pone.0294469>.
11. Camacho, D.M.; Collins, K.M.; Powers, R.K.; Costello, J.C.; Collins, J.J. Next-Generation Machine Learning for Biological Networks. *Cell* **2018**, *173*, 1581–1592. <https://doi.org/10.1016/j.cell.2018.05.015>.
12. Martinez, J.P.O.; Speight, R.E. High-Throughput Yeast Engineering in Biofoundries: Towards Autonomous and Scalable Synthetic Biology. *FEMS Yeast Res.* **2026**, *26*, foag003. <https://doi.org/10.1093/femsyr/foag003>.
13. Renzo, T.D.; Reale, A. Process Optimization and Quality Improvement of Fermented Foods and Beverages. *Foods* **2025**, *14*, 1238. <https://doi.org/10.3390/foods14071238>.
14. Mears, L.; Stocks, S.M.; Sin, G.; Gernaey, K.V. A Review of Control Strategies for Manipulating the Feed Rate in Fed-Batch Fermentation Processes. *J. Biotechnol.* **2017**, *245*, 34–46. <https://doi.org/10.1016/j.jbiotec.2017.01.008>.

15. Pennington, O.; Zhang, D. Comparing Different Modelling Approaches for Metabolic Network Dynamic Simulation under Uncertainty. In *33 European Symposium on Computer Aided Process Engineering*; Kokossis, A.C., Georgiadis, M.C., Pistikopoulos, E., Eds.; Computer Aided Chemical Engineering: Amsterdam, The Netherlands; Elsevier: Amsterdam, The Netherlands, 2023; Volume 52, pp. 2589–2594.
16. Almquist, J.; Cvijovic, M.; Hatzimanikatis, V.; Nielsen, J.; Jirstrand, M. Kinetic Models in Industrial Biotechnology—Improving Cell Factory Performance. *Metab. Eng.* **2014**, *24*, 38–60. <https://doi.org/10.1016/j.ymben.2014.03.007>.
17. Huang, J.-C.; Guo, Q.; Li, X.-H.; Shi, T.-Q. A Comprehensive Review on the Application of Neural Network Model in Microbial Fermentation. *Bioresour. Technol.* **2025**, *416*, 131801. <https://doi.org/10.1016/j.biortech.2024.131801>.
18. Libbrecht, M.W.; Noble, W.S. Machine Learning Applications in Genetics and Genomics. *Nat. Rev. Genet.* **2015**, *16*, 321–332. <https://doi.org/10.1038/nrg3920>.
19. Kang, M.; Ko, E.; Mersha, T.B. A Roadmap for Multi-Omics Data Integration Using Deep Learning. *Brief. Bioinform.* **2022**, *23*, bbab454. <https://doi.org/10.1093/bib/bbab454>.
20. Fan, X.; Xiang, G.; Liao, W.; Xiao, L.; He, S.; Luo, N.; Lu, H.; Yan, X. Decoding Yeast Transcriptional Regulation via a Data-and Mechanism-Driven Distributed Large-Scale Network Model. *Synth. Syst. Biotechnol.* **2025**, *10*, 1140–1149. <https://doi.org/10.1016/j.synbio.2025.06.005>.
21. Chen, L.; Cai, C.; Chen, V.; Lu, X. Learning a Hierarchical Representation of the Yeast Transcriptomic Machinery Using an Autoencoder Model. *BMC Bioinform.* **2016**, *17*, S9. <https://doi.org/10.1186/s12859-015-0852-1>.
22. Sakharova, H.; Lareau, L.F. Protein Language Models Reveal Evolutionary Constraints on Synonymous Codon Choice. *bioRxiv* **2025**, 2025.08.05.668603. <https://doi.org/10.1101/2025.08.05.668603>.
23. Dash, T.; Bornelöv, S. Predicting Gene Expression Using Millions of Yeast Promoters Reveals Cis-Regulatory Logic. *Bioinform. Adv.* **2025**, *5*, vbaf130. <https://doi.org/10.1093/bioadv/vbaf130>.
24. Jagtap, S.; Pirayre, A.; Bidard, F.; Duval, L.; Malliaros, F.D. BRANenet: Embedding Multilayer Networks for Omics Data Integration. *BMC Bioinform.* **2022**, *23*, 429. <https://doi.org/10.1186/s12859-022-04955-w>.
25. Han, C.; Fu, S.; Tang, D.; Chen, Y.; Liu, D.; Feng, Z.; Gou, Y.; Zhang, C.; Zhang, W.; Xiao, L.; et al. Omic AI Reveals New Autophagy Regulators from the Atg1 Interactome in *Saccharomyces Cerevisiae*. *Front. Cell Dev. Biol.* **2025**, *13*, 1554958. <https://doi.org/10.3389/fcell.2025.1554958>.
26. Mendes, I.; Sanchez, I.; Franco-Duarte, R.; Camarasa, C.; Schuller, D.; Dequin, S.; Sousa, M.J. Integrating Transcriptomics and Metabolomics for the Analysis of the Aroma Profiles of *Saccharomyces Cerevisiae* Strains from Diverse Origins. *BMC Genom.* **2017**, *18*, 455. <https://doi.org/10.1186/s12864-017-3816-1>.
27. Tian, D.; Wang, P.; Tang, B.; Teng, X.; Li, C.; Liu, X.; Zou, D.; Song, S.; Zhang, Z. GWAS Atlas: A Curated Resource of Genome-Wide Variant-Trait Associations in Plants and Animals. *Nucleic. Acids. Res.* **2020**, *48*, D927–D932. <https://doi.org/10.1093/nar/gkz828>.
28. Tam, V.; Patel, N.; Turcotte, M.; Bossé, Y.; Paré, G.; Meyre, D. Benefits and Limitations of Genome-Wide Association Studies. *Nat. Rev. Genet.* **2019**, *20*, 467–484. <https://doi.org/10.1038/s41576-019-0127-1>.
29. Boulesteix, A.-L.; Janitzka, S.; Kruppa, J.; König, I.R. Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics. *WIREs Data Min. Knowl. Discov.* **2012**, *2*, 493–507. <https://doi.org/10.1002/widm.1072>.
30. Harrison, M.-C.; Ubbelohde, E.J.; LaBella, A.L.; Oplente, D.A.; Wolters, J.F.; Zhou, X.; Shen, X.-X.; Groenewald, M.; Hittinger, C.T.; Rokas, A. Machine Learning Enables Identification of an Alternative Yeast Galactose Utilization Pathway. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2315314121. <https://doi.org/10.1073/pnas.2315314121>.
31. Khaiwal, S.; De Chiara, M.; Barré, B.P.; Barrio-Hernandez, I.; Stenberg, S.; Beltrao, P.; Warringer, J.; Liti, G. Predicting Natural Variation in the Yeast Phenotypic Landscape with Machine Learning. *Mol. Syst. Biol.* **2025**, *21*, 1466–1489. <https://doi.org/10.1038/s44320-025-00136-y>.
32. Zelezniak, A.; Vowinkel, J.; Capuano, F.; Messner, C.B.; Demichev, V.; Polowsky, N.; Mülleder, M.; Kamrad, S.; Klaus, B.; Keller, M.A.; et al. Machine Learning Predicts the Yeast Metabolome from the Quantitative Proteome of Kinase Knockouts. *Cell Syst.* **2018**, *7*, 269–283.e6. <https://doi.org/10.1016/j.cels.2018.08.001>.
33. He, Y.; Degraeve, P.; Oulahal, N. Bioprotective Yeasts: Potential to Limit Postharvest Spoilage and to Extend Shelf Life or Improve Microbial Safety of Processed Foods. *Heliyon* **2024**, *10*, e24929. <https://doi.org/10.1016/j.heliyon.2024.e24929>.
34. Kovács, M.; Pomázi, A.; Taczman-Brückner, A.; Kiskó, G.; Dobó, V.; Kocsis, T.; Mohácsi-Farkas, C.; Belák, Á. Detection and Identification of Food-Borne Yeasts: An Overview of the Relevant Methods and Their Evolution. *Microorganisms* **2025**, *13*, 981. <https://doi.org/10.3390/microorganisms13050981>.

35. Park, H.W.; Mason Earles, J.; Nitin, N. Deep Learning Enabled Rapid Classification of Yeast Species in Food by Imaging of Yeast Microcolonies. *Food Res. Int.* **2025**, *201*, 115604. <https://doi.org/10.1016/j.foodres.2024.115604>.
36. Moen, E.; Bannon, D.; Kudo, T.; Graf, W.; Covert, M.; Van Valen, D. Deep Learning for Cellular Image Analysis. *Nat. Methods* **2019**, *16*, 1233–1246. <https://doi.org/10.1038/s41592-019-0403-1>.
37. Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; Aila, T. Training Generative Adversarial Networks with Limited Data. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12104–12114.
38. Aiki, K.; Tsuchiya, R.; Kushida, A.; Tominaga, T. Rapid Counting of *Kazachstania humilis* and *Saccharomyces cerevisiae* in Sourdough by Deep Learning-Based Classifier. *J. Microbiol. Methods* **2025**, *236*, 107183. <https://doi.org/10.1016/j.mimet.2025.107183>.
39. Yang Yu, B.; Elbuken, C.; Ren, C.L.; Huissoon, J.P. Image Processing and Classification Algorithm for Yeast Cell Morphology in a Microfluidic Chip. *J. Biomed. Opt.* **2011**, *16*, 066008. <https://doi.org/10.1117/1.3589100>.
40. Prangemeier, T.; Wildner, C.; Frančani, A.O.; Reich, C.; Koepl, H. Yeast Cell Segmentation in Microstructured Environments with Deep Learning. *Biosystems* **2022**, *211*, 104557. <https://doi.org/10.1016/j.biosystems.2021.104557>.
41. Salem, D.; Li, Y.; Xi, P.; Phenix, H.; Cuperlovic-Culf, M.; Kærn, M. YeastNet: Deep-Learning-Enabled Accurate Segmentation of Budding Yeast Cells in Bright-Field Microscopy. *Appl. Sci.* **2021**, *11*, 2692. <https://doi.org/10.3390/app11062692>.
42. Lu, A.X.; Zarin, T.; Hsu, I.S.; Moses, A.M. YeastSpotter: Accurate and Parameter-Free Web Segmentation for Microscopy Images of Yeast Cells. *Bioinformatics* **2019**, *35*, 4525–4527. <https://doi.org/10.1093/bioinformatics/btz402>.
43. Haja, A.; Schomaker, L.R.B. A Fully Automated End-to-End Process for Fluorescence Microscopy Images of Yeast Cells: From Segmentation to Detection and Classification. In *Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021)*; IEEE: New York, NY, USA, 2022; pp. 37–46. https://doi.org/10.1007/978-981-16-3880-0_5.
44. Egebjerg, J.M.; Szomek, M.; Thaysen, K.; Juhl, A.D.; Kozakijevic, S.; Werner, S.; Pratsch, C.; Schneider, G.; Kapishnikov, S.; Ekman, A.; et al. Automated Quantification of Vacuole Fusion and Lipophagy in *Saccharomyces Cerevisiae* from Fluorescence and Cryo-Soft X-Ray Microscopy Data Using Deep Learning. *Autophagy* **2024**, *20*, 902–922. <https://doi.org/10.1080/15548627.2023.2270378>.
45. Szűcs, B.; Selvan, R.; Lisby, M. High-Throughput Classification of *S. Cerevisiae* Tetrads Using Deep Learning. *Yeast* **2024**, *41*, 423–436. <https://doi.org/10.1002/yea.3965>.
46. Litsios, A.; Grys, B.T.; Kraus, O.Z.; Friesen, H.; Ross, C.; Masinas, M.P.D.; Forster, D.T.; Couvillion, M.T.; Timmermann, S.; Billmann, M.; et al. Proteome-Scale Movements and Compartment Connectivity during the Eukaryotic Cell Cycle. *Cell* **2024**, *187*, 1490–1507.e21. <https://doi.org/10.1016/j.cell.2024.02.014>.
47. Reich, C. Generation and Simulation of Yeast Microscopy Imagery with Deep Learning. *arXiv* **2021**, arXiv:2103.11834.
48. Wu, X.; Wan, X.; Yu, H.; Liu, H. Recent Advances in CRISPR-Cas System for *Saccharomyces Cerevisiae* Engineering. *Biotechnol. Adv.* **2025**, *81*, 108557. <https://doi.org/10.1016/j.biotechadv.2025.108557>.
49. Kim, M.; Go, M.; Kang, S.-H.; Jeong, S.; Lim, K. Revolutionizing CRISPR Technology with Artificial Intelligence. *Exp. Mol. Med.* **2025**, *57*, 1419–1431. <https://doi.org/10.1038/s12276-025-01462-9>.
50. Baisya, D.; Ramesh, A.; Schwartz, C.; Lonardi, S.; Wheeldon, I. Genome-Wide Functional Screens Enable the Prediction of High Activity CRISPR-Cas9 and -Cas12a Guides in *Yarrowia Lipolytica*. *Nat. Commun.* **2022**, *13*, 922. <https://doi.org/10.1038/s41467-022-28540-0>.
51. Lal, A.; Gunsalus, L.; Gupta, A.; Biancalani, T.; Eraslan, G. Polygraph: A Software Framework for the Systematic Assessment of Synthetic Regulatory DNA Elements. *Genome. Biol.* **2025**, *26*, 114. <https://doi.org/10.1186/s13059-025-03584-9>.
52. Yang, G.; Chen, Y.; Guo, Q.; Li, X.; Zhou, Z. Leveraging Pre-Trained AI Models for Robust Promoter Sequence Design in Synthetic Biology. *swvlxb* **2026**, *12*, 126–135. <https://doi.org/10.52601/bpr.2025.240033>.
53. AlMarzooqi, A.A.; Ahmad, M.A.; Chalissery, J.; Hassan, A.H. AI-Assisted Impedance Biosensing of Yeast Cell Concentration. *Biosensors* **2025**, *16*, 18. <https://doi.org/10.3390/bios16010018>.
54. Kim, G.B.; Kim, W.J.; Kim, H.U.; Lee, S.Y. Machine Learning Applications in Systems Metabolic Engineering. *Curr. Opin. Biotechnol.* **2020**, *64*, 1–9. <https://doi.org/10.1016/j.copbio.2019.08.010>.
55. Song, Y.; Xu, H.; Li, C.; He, Q.; Tian, Z.; Liu, X. HMT: Hybrid Mechanistic Transformer for Bio-Fabrication Prediction under Complex Environmental Conditions. *Expert Syst. Appl.* **2024**, *255*, 124389. <https://doi.org/10.1016/j.eswa.2024.124389>.
56. Nair, N.K.; D'Souza, A. An Integrative Genome-Scale Metabolic Modeling and Machine Learning Framework for Predicting and Optimizing Single-Cell Protein Production in *Saccharomyces Cerevisiae*. *arXiv* **2026**, arXiv:2603.25561.

57. Brunnsäker, D.; Gower, A.; Naval, P.; Bjurström, E.; Kronström, F.; Tiukova, I.; King, R. Self-Driven Biological Discovery through Automated Hypothesis Generation and Experimental Validation. *bioRxiv* **2025**. <https://doi.org/10.1101/2025.06.24.661378>.
58. Wang, Z.-Z.; Zeng, D.-W.; Zhu, Y.-F.; Zhou, M.-H.; Kondo, A.; Hasunuma, T.; Zhao, X.-Q. Fermentation Design and Process Optimization Strategy Based on Machine Learning. *BioDesign Res.* **2025**, *7*, 100002. <https://doi.org/10.1016/j.bidere.2025.100002>.
59. Yee, C.S.; Zahia-Azizan, N.A.; Rahim, M.H.A.; Zaini, N.A.M.; Raja-Razali, R.B.; Ushidee-Radzi, M.A.; Ilham, Z.; Wan-Mohtar, W.A.A.Q.I. Smart Fermentation Technologies: Microbial Process Control in Traditional Fermented Foods. *Fermentation* **2025**, *11*, 323. <https://doi.org/10.3390/fermentation11060323>.
60. Nettesheim, P.; Burggräf, P.; Steinberg, F. Applications of Machine Learning in the Brewing Process: A Systematic Review. *Discov. Artif. Intell.* **2024**, *4*, 80. <https://doi.org/10.1007/s44163-024-00177-6>.
61. Bowler, A.; Escrig, J.; Pound, M.; Watson, N. Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning. *Fermentation* **2021**, *7*, 34. <https://doi.org/10.3390/fermentation7010034>.
62. Jeong, J.; Kim, S. Application of Machine Learning for Quantitative Analysis of Industrial Fermentation Using Image Processing. *Food Sci. Biotechnol.* **2025**, *34*, 373–381. <https://doi.org/10.1007/s10068-024-01744-4>.
63. Gedela, R.; Vantaku, V.R.; Dasu, V.V.; Pakshirajan, K. Artificial Intelligence and Machine Learning-Assisted Metabolite Profiling of Oleaginous Yeast *Rhodotorula Mucilaginosa*. *Bioresour. Technol. Rep.* **2025**, *32*, 102431. <https://doi.org/10.1016/j.biteb.2025.102431>.
64. Zhang, Y.; Jia, S.; Zhang, W. Predicting Acetic Acid Content in the Final Beer Using Neural Networks and Support Vector Machine. *J. Inst. Brew.* **2012**, *118*, 361–367. <https://doi.org/10.1002/jib.50>.
65. Itto-Nakama, K.; Watanabe, S.; Kondo, N.; Ohnuki, S.; Kikuchi, R.; Nakamura, T.; Ogasawara, W.; Kasahara, K.; Ohya, Y. AI-Based Forecasting of Ethanol Fermentation Using Yeast Morphological Data. *Biosci. Biotechnol. Biochem.* **2022**, *86*, 125–134. <https://doi.org/10.1093/bbb/zbab188>.
66. Groff-Vindman, C.S.; Trump, B.D.; Cummings, C.L.; Smith, M.; Titus, A.J.; Oye, K.; Prado, V.; Turmus, E.; Linkov, I. The Convergence of AI and Synthetic Biology: The Looming Deluge. *NPJ Biomed. Innov.* **2025**, *2*, 20. <https://doi.org/10.1038/s44385-025-00021-1>.
67. Eslami, M.; Adler, A.; Caceres, R.S.; Dunn, J.G.; Kelley-Loughnane, N.; Varaljay, V.A.; Martin, H.G. Artificial Intelligence for Synthetic Biology. *Commun. ACM* **2022**, *65*, 88–97. <https://doi.org/10.1145/3500922>.
68. Liu, L.; Xia, Y.; Tang, L. An Overview of Biological Data Generation Using Generative Adversarial Networks. In *2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*; IEEE: New York, NY, USA, 2020; pp. 141–144. <https://doi.org/10.1109/TOCS50858.2020.9339748>.
69. Zinchuk, V.; Grossenbacher-Zinchuk, O. Machine Learning for Analysis of Microscopy Images: A Practical Guide and Latest Trends. *Curr. Protoc.* **2023**, *3*, e819. <https://doi.org/10.1002/cpz1.819>.
70. Ma, L.; Yi, J.; Wisuthiphaet, N.; Earles, M.; Nitin, N. Accelerating the Detection of Bacteria in Food Using Artificial Intelligence and Optical Imaging. *Appl. Environ. Microbiol.* **2022**, *89*, e01828-22. <https://doi.org/10.1128/aem.01828-22>.
71. Yi, J.; Wisuthiphaet, N.; Raja, P.; Nitin, N.; Earles, J.M. AI-Enabled Biosensing for Rapid Pathogen Detection: From Liquid Food to Agricultural Water. *Water Res.* **2023**, *242*, 120258. <https://doi.org/10.1016/j.watres.2023.120258>.

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