Biotechnology and Bioengineering

WILEY

BIOTECHNOLOGY BIOENGINEERING

REVIEW OPEN ACCESS

Transfer Learning Approaches in Bioprocess Engineering: Opportunities and Challenges

Daniel Barón Díaz | Anna-Lena Drommershausen | Alexander Grünberger | Dirk Holtmann

Institute of Process Engineering in Life Sciences, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Correspondence: Dirk Holtmann (dirk.holtmann@kit.edu)

Received: 10 December 2025 | **Revised:** 19 February 2026 | **Accepted:** 27 February 2026

Funding: Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Grant/Award Number: 536355345

Keywords: bioprocess engineering | data scarcity | hybrid modeling | machine learning | transfer learning

ABSTRACT

Transfer learning (TL) has recently emerged as a promising approach to overcoming one of the key limitations of bioprocess engineering: data scarcity. By leveraging knowledge from one bioprocess to another, TL allows existing models and data sets to be reused efficiently, accelerating process development, improving prediction accuracy, and enhancing model robustness in situations in which data are limited. This review critically assesses recent advances in the application of TL in bioprocess engineering. From genomic analysis to bioreactor modeling and analytics, TL can increase the accuracy of models aiming to predict protein functions, growth, and product formation as well as retention times in chromatographic processes. Despite its potential, several challenges remain, including data heterogeneity and model transferability. Future research will most likely focus on integrating TL with hybrid and physics-informed modeling frameworks, developing standardized benchmark data sets, and exploiting TL to extract relevant information from publicly available data sets. Overall, TL provides a way forward for creating more data-efficient, generalizable, and interpretable models for bioprocess engineering.

1 | Introduction

Bioprocess engineering is the foundation of industrial biotechnology, facilitating the production of pharmaceuticals, fine chemicals, specialty chemicals, and bio-based materials via microbial and enzymatic processes. Modern bioprocesses are inherently complex, governed by interactions on multiple scales between cellular metabolism, transport phenomena, and process dynamics (Blöbaum et al. 2023). Despite substantial progress in process automation, miniaturization, and analytical technologies, the development and optimization of bioprocesses are still limited by a lack of data and data heterogeneity (Helleckes et al. 2023; Mondal et al. 2023). Although high-throughput cultivation systems have increased experimental capacity, they often provide only a few online measurements, such as biomass concentration, dissolved oxygen, pH, or fluorescence (Fink et al. 2021; Kunze et al. 2014). Furthermore, data sets obtained at different scales or under varying operating conditions are often inconsistent, hindering data integration and model transferability across systems (Butean et al. 2025).

Mechanistic modeling approaches have long provided valuable insight into reaction kinetics and transport processes, but their applicability is constrained by incomplete biological knowledge and high computational demands. Data-driven machine learning (ML) methods have therefore gained attention as powerful tools for modeling nonlinear and complex relationships in bioprocesses in which mechanistic understanding is limited (Karimi Alavijeh et al. 2022; Mondal et al. 2023; Mowbray et al. 2021). In such scenarios, a data-driven model that maps inputs to outputs, without fully modeling the internal physical phenomena, can reduce development time and effort. However, in experimental biotechnology, the conditions for achieving generalizable performance are rarely met by ML methods, which typically require large, structured data sets (Peng et al. 2026). Consequently, models, usually built on small data

sets, often fail to extrapolate to new strains, process conditions, or scales, necessitating costly retraining for each new system (Peng et al. 2026).

Transfer learning (TL) has emerged as a promising approach to overcoming these limitations by reusing knowledge acquired from related processes with sufficiently large data sets to improve learning efficiency in new, data-scarce environments. TL enables the adaptation of pretrained models through parameter fine-tuning, feature-space alignment, or relational knowledge transfer, thereby reducing the need for large target-domain data sets (Figure 1) (Pan and Yang 2010). Additionally, incorporating process knowledge in the form of mechanistic models into TL frameworks allows physical understanding to be exploited, thereby mitigating data scarcity and reducing the experimental effort required to build reliable statistical models (Kay et al. 2023; Rogers et al. 2025). This integration of mechanistic and data-driven approaches is commonly referred to as hybrid modeling (Hodgson 2005).

TL has already transformed diverse fields such as image recognition, ranging from brain tumor detection to cultural heritage classification (Janković Babić 2024; Kim et al. 2022; Weiss et al. 2016; Anwar et al. 2023), as well as natural language processing (Weiss et al. 2016), and recent studies demonstrate its increasing importance in bioprocess engineering. This review aims to provide a comprehensive overview of TL methodologies and their emerging applications in bioprocess engineering. Particular emphasis is placed on the opportunities and challenges of integrating TL and hybrid and physics-informed modeling frameworks, with the ultimate aim of establishing TL as a cornerstone for data-efficient, generalizable, and interpretable bioprocess models.

2 | General Introduction Into Transfer Learning

Traditional ML involves uncovering patterns in large data sets using statistical methods, often without providing mechanistic insight into the underlying systems. Typical ML tasks include classification, regression, or clustering (Alzubi et al. 2018). Data sets in ML are described by features, which are measurable attributes such as temperature, pH, or biomass concentration in the context of a bioprocess (Figure 2). The collection of all possible feature combinations forms the feature space X, and each feature is associated with a marginal probability distribution $P(X)$, describing how likely different values of a feature are to appear (Pan and Yang 2010). A domain $\mathcal{D}$ consists of a feature space X and its corresponding $P(X)$. A task $\mathcal{T}$, on the other hand, is defined by a label space Y and a predictive function $f(\cdot)$, which can be learned from data and used to predict the label y

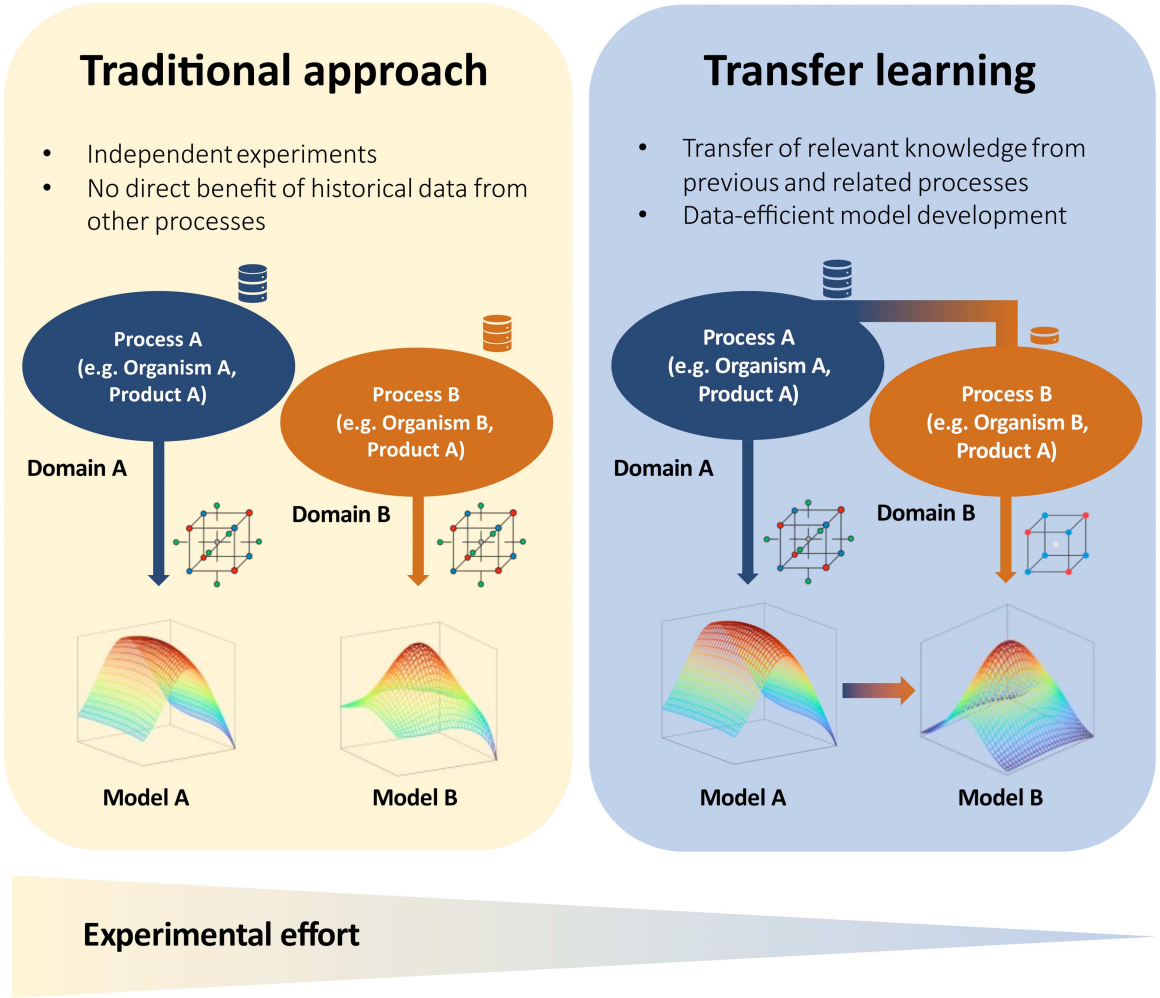

FIGURE 1 | Comparison of the traditional model building workflow with the transfer learning approach.

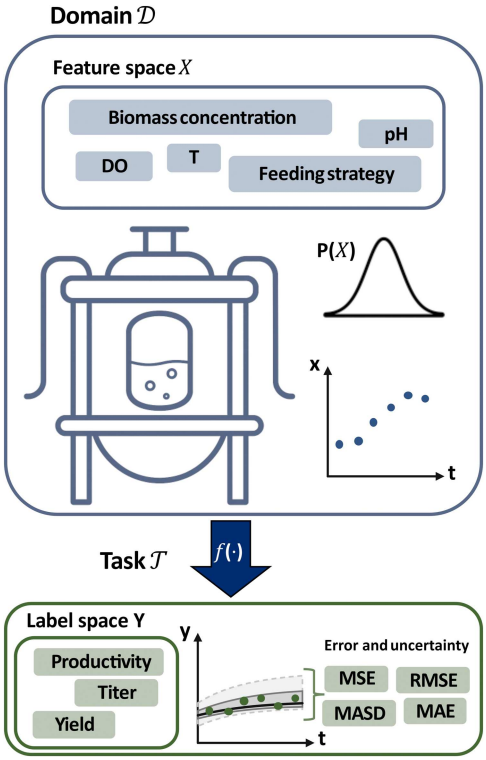

FIGURE 2 | Schematic representation of transfer learning terminology, illustrated within the context of a bioreactor process model. The task $\mathcal{T}$ consists in finding a predictive function $f(\cdot)$, that can map a specific input x to its corresponding label y. In this context, the feature space X is comprised of critical process parameters and attributes such as biomass concentration, pH, dissolved oxygen (DO), temperature (T), and the feeding strategy. The label space Y represents the target attributes assigned to the process, such as product titer, yield, and productivity. There are different metrics available to quantify uncertainty and error of the predictive task, such as the mean squared error (MSE), root mean squared error (RMSE), mean absolute standard deviation (MASD), or mean average error (MAE).

from a new input x (Pan and Yang 2010). Data sets in which each input x is paired with a known outcome y are referred to as labeled data (Pan and Yang 2010). By contrast, unlabeled data consist only of input measurements, for which no corresponding outcome or reference value is available (Pan and Yang 2010). The prediction task aims to minimize the error of $f(\cdot)$ on the training data. For error and uncertainty quantification several options are available, including the root mean squared error (RMSE), mean squared error (MSE), mean average error (MAE), mean average percentage error (MAPE), or mean absolute standard deviation (MASD) (Figure 2).

One critical assumption in traditional ML is that the training and test data come from the same domain $\mathcal{D}$, meaning they share the same X and $P(X)$. For instance, a model trained to predict product formation in a particular microorganism is typically evaluated on data from that same organism under similar bioprocess conditions. As a result, traditional ML approaches require training a new model from scratch for every new domain or task. TL challenges this assumption by leveraging knowledge from a source domain and task to improve performance on a target task, often in a different domain. This is particularly valuable in bioprocessing, for which data in the target domain (e.g., a new bioprocess) is often scarce and expensive to obtain, as data collection is labor-intensive and relies on costly analytical methods such as HPLC, mass spectrometry, or enzymatic assays. Moreover, if a model is trained across multiple domains, it becomes better equipped to identify and generalize key patterns, even when domains differ.

The pioneering work of Pan and Yang categorized TL into three types from a label-setting perspective, based on the relationship between source and target domains and tasks: inductive, transductive, and unsupervised TL (Pan and Yang 2010; Zhuang et al. 2021). In inductive TL, the source and target tasks are different, but labeled data are available in the target domain (Pan and Yang 2010). For example, a model trained to predict the maximum protein yield in a large-scale fermentation can help a second model learn to predict the optimal time for metabolite extraction, since both depend on understanding the same bioprocess conditions, and only a small amount of labeled data is needed for the second task. In transductive TL, the task stays the same, but the domains differ, usually with labeled data only in the source domain and unlabeled data in the target domain (Pan and Yang 2010). For instance, a soft sensor trained on spectral data from small-scale bioreactor experiments might not deliver accurate predictions on large-scale bioreactors due to differences in system dynamics and reactor wall reflection. Instead of training a new model from expensive large-scale data, TL can adapt the existing model by transforming inputs or fine-tuning the original model's parameters. Finally, in unsupervised TL, neither domain has labeled data (Pan and Yang 2010). Instead, structural knowledge is transferred across related unsupervised tasks, such as using a model trained to cluster unlabeled sensor data from one bioreactor by operational mode to cluster kinetic profiles from a different reactor using another bacterial strain. The studies reviewed in this work largely showed that inductive TL dominated across different applications. This reflects the need in bioprocess engineering to adapt models to tasks where target-labeled data are available but scarce.

The surveys by Weiss and Zhuang further systematize TL strategies by introducing a space-setting–based categorization (Weiss et al. 2016; Zhuang et al. 2021). This perspective classifies methods according to the relationship between the feature and label spaces of the source and target domains. In homogeneous TL, the feature and label spaces are identical ($X^S = X^T$ and $Y^S = Y^T$), while the underlying data distributions may differ (Weiss et al. 2016; Zhuang et al. 2021). By contrast, heterogeneous TL addresses cases in which the feature spaces differ and additional feature-space adaptation is required (Weiss et al. 2016; Zhuang et al. 2021). For example, transferring a bioreactor process model between different cell strains changes the marginal data distributions, but the measured variables and targets remain the same, constituting homogeneous TL. In contrast, heterogeneous TL arises when the target process includes additional features absent in the source domain, such as the introduction of a new nutrient in the fermentation process. Across the reviewed literature, mostly homogenous TL is employed, thereby reducing the complexity of the implementation of TL, as feature spaces remain consistent across domains. The key benefit of TL is that it can accelerate model development by reusing prior knowledge. Just as humans apply

experience from one setting to another, TL enables models to generalize information across contexts. A real application example is presented in the work by Riezzo et al. (2025), in which the authors applied TL to enhance the prediction of growth and product formation of a new strain with scarce data. This example has been illustrated in Figure 3.

When implementing TL, it is important to define which part of the information contained in the source data should be transferred. Pan and Yang outline four strategies to perform TL (Figure 4):

- *Instance-transfer* consists in re-weighting source data points using importance sampling, in which instances are weighted in a way that the model focuses on source data most similar to the target data. For example, a model trained to predict product formation for an organism A can be reused by assigning higher weights to process data that resembles that for organism B, for which fewer data points are available.
- *Parameter-transfer* can be achieved, for example, by retaining certain layers of an artificial neural network (ANN) and retraining the remaining layers using target data. In this approach, a model is first pretrained using source data to learn the internal parameters, specifically the weights and biases that define the connections between neurons. These parameters are then used to initialize the model for the target task and they can be adjusted further to better fit the new domain. This approach is known as fine-tuning and the main advantage compared to normal parameter estimation is that the starting parameters for training are not random at the beginning, but already contain information from the source data and it therefore leads to shorter training times and higher accuracy when trained on small data sets compared to training from scratch.
- *Feature-representation-transfer* maps both source and target data into a new feature space in which their distributions are more similar. This can be done using methods like Transfer Component Analysis (TCA) or Maximum Mean Discrepancy Embedding. For example, a model trained on data sets containing thousands of protein sequences to learn general patterns that relate sequence features to physicochemical properties can be reused to improve the solubility prediction of monoclonal antibodies (mAbs), for which data are limited and costly to generate (Feng et al. 2022).
- *Relational-knowledge-transfer* focuses on transferring structural information (e.g., graphs or relationship networks), even when entities differ. For example, a model that learns protein–protein interactions in mice can transfer relational knowledge to predict interactions in human proteins, since the structure of biological networks often follows similar relational patterns across species.

For TL to work effectively, the source and target domains must be meaningfully related. If the domains are too similar, traditional ML might already work well and TL implementation may not impact the model's performance significantly (Bikias et al. 2025). If too different, TL may even be detrimental for the model, a problem known as negative transfer (Ge et al. 2014; Z.

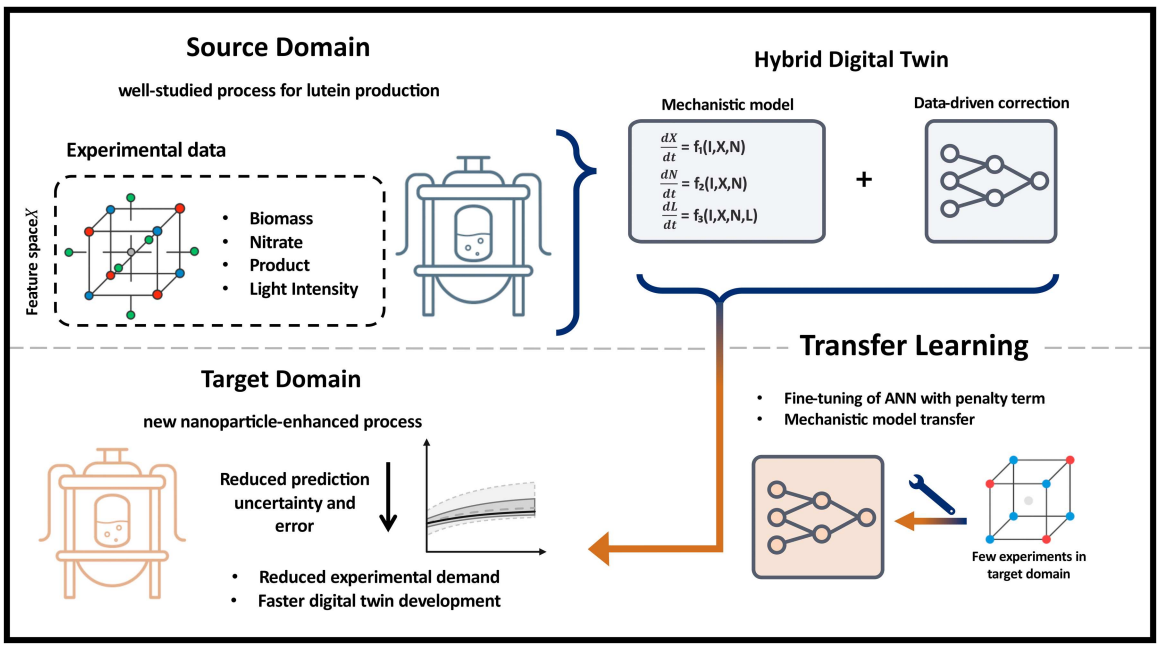

FIGURE 3 | Schematic overview of a concrete application example of transfer learning based on the work by Riezzo et al. (2025). Data from the source domain is used to construct a hybrid digital twin that combines a mechanistic process model with an artificial neural network (ANN). The ANN provides data-driven corrections to selected model parameters, such as the biomass decay and lutein consumption rate, which exhibit high variability and cannot be adequately captured by the mechanistic formulation alone. The mechanistic model structure is then transferred unchanged to a new strain. For the target process involving the nanoparticle-enhanced strain, the ANN is fine-tuned using new data from only two experiments. During retraining, a penalty term is applied to limit changes in the ANN weights and biases, thereby preserving knowledge learned from the source strain while adapting to the new process (Riezzo et al. 2025).

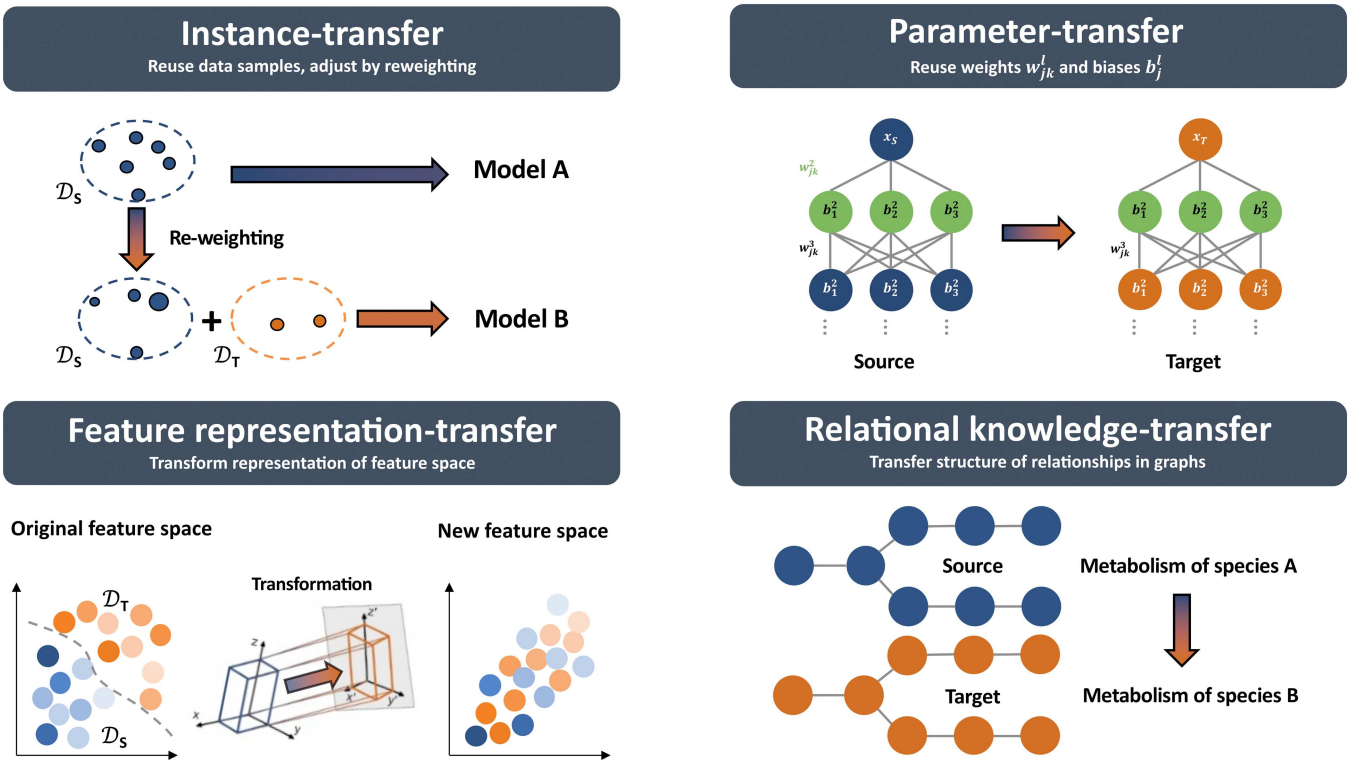

FIGURE 4 | Schematic illustration of transfer learning strategies based on the framework of Pan and Yang. Blue denotes the source domain ($\mathcal{D}_S$) and orange the target domain ($\mathcal{D}_T$). Instance-transfer may be performed by reweighting samples from $\mathcal{D}_S$ according to their similarity to $\mathcal{D}_T$ and incorporating them into target model training. Parameter-transfer, for example, in an ANN, involves transferring parameters from the source model and selectively adapting them using target data, while other parameters can be kept fixed (green neurons). Feature representation-transfer maps source and target data into a shared representation to exploit similarities, and relational knowledge-transfer focuses on reusing structural information learned in the source domain (Pan and Yang 2010).

Wang et al. 2019). This typically arises when the assumptions, data distributions, or learned representations of the source domain are misaligned with those of the target domain (Cai et al. 2020; Kota et al. 2021; Galeazzi et al. 2025). For example, in CRISPR–Cas9 off-target prediction, models trained on high-throughput proxy assays learned features that did not generalize to true in vivo behavior. In this case, increasing model complexity improved performance on proxy data but degraded accuracy on in vivo benchmarks, highlighting the risk of negative transfer when source and target data arise from fundamentally different distributions (Kota et al. 2021).

To explore the relatedness between domains, several strategies have been developed that are based, for example, on label propagation and clustering, modeling relationships between source tasks and the target task using graph theory, and identifying helpful source data through instance-level weighting based on domain similarity (Eaton et al. 2008; Ge et al. 2014; Z. Wang et al. 2019). In the context of bioprocessing, expert knowledge can also play a key role, as domain specialists may be able to assess the relevance of source data based on process understanding before TL is applied. Ultimately, the choice of strategy depends on the task, data type, and the degree of similarity between source and target domains, making it challenging to decide which approach to use. This review includes an overview table that summarizes the main strategies for implementation of TL in bioprocessing along with the example applications mentioned throughout the review, helping to guide method selection based on the task characteristics (Table S1). Additionally, a glossary has been included summarizing ML terminology that is used throughout this work (Table S2).

3 | Transfer Learning for Bioprocess Engineering

The aim of this section is to provide an overview of the current applications of TL in bioprocess engineering, identifying both emerging opportunities and limitations. Based on the conceptual foundations introduced earlier, it examines how TL can overcome the challenges of data scarcity and model transferability across the bioprocess chain. The section is organized by key application domains, including molecular analysis and engineering, biocatalysis, bioreactor modeling and monitoring, as well as chromatography and downstream processing. It demonstrates how knowledge reuse can enable more efficient model training and improved predictive performance across different applications. Through these examples, the section aims to highlight methodological trends and demonstrate how TL can be incorporated into the development of future data-driven bioprocesses.

3.1 | Molecular Analysis and Engineering

Computational biology applies mathematical models, algorithmic approaches, and data analysis techniques to understand and predict the behavior of biological systems. Genomic analysis, in particular, exploits DNA and RNA data sets to uncover genetic sequences, patterns, and functional relationships.

Combined, these strategies enable the prediction of diverse biological mechanisms, thereby supporting the rational design of targeted therapeutic and biotechnological strategies. Typical data structure includes the representation of molecules as graphs with attributes like atom types, charges, or bonds. Other data can include sequences (DNA, RNA, and protein). A majority of experiments deals with regression tasks, such as prediction of binding affinities, toxicity levels, or thermodynamic properties. Classification tasks can include data annotation and image analysis. Here, an overview is provided encompassing applications of TL to CRISPR system modeling, promotor engineering, and biological data annotation.

3.1.1 | *CRISPR System Modeling and Prediction*

CRISPR is a powerful genetic tool that enables precise DNA editing to enhance traits such as productivity, stability, or yield. Its high efficiency and ease of use have revolutionized life sciences and medicine. To ensure precise edits, cutting sites with minimal off-target effects must be accurately predicted, making the modeling and interpretation of CRISPR–Cas9/sgRNA activity a key research focus (Kota et al. 2021).

To improve the prediction of these interactions, TL within deep learning (DL) architectures has been explored. For instance, a DL framework, designed to determine the kinetics of biochemical systems, was developed by Zhang et al. (2023). In the study, the method is used to predict and understand CRISPR–Cas9 off-target reactions. The framework is built by first setting up kinetically interpretable neural networks (KINNs) that are able to predict reaction rates. The KINNs are then used as intermediary layers in a deep convolutional neural network (CNN). In this TL approach, layers of the KINNs, trained on a specialized data set, are transferred to the CNN layers, which are built from in vivo data. Applying this TL strategy provides model interpretability and results in a decreased training effort compared to traditional CNNs, as the optimized KINNs require a factor of 240 fewer parameters. The overall performance of classification was between 150% and 440% higher using TL than the classical optimized DL model across the tested classical benchmark models (Zhang et al. 2023). A different approach was taken by Ham et al. (2023) with the development of a general Cas9/sgDNA prediction model. This model was trained first on large existing data sets (~95,000 sgRNAs) and then fine-tuned using a small amount (~550 sgRNAs) of new, high-quality data. Through this strategy, the bacterial sgRNA prediction accuracy on the new high-quality data set was improved by 21% based on the Spearman rank correlation coefficient, compared to previous models (Guo et al. 2018). However, there are still significant limitations in the ability to predict highly active Cas9/sgRNA interactions (Ham et al. 2023).

Overall, there is a high need for further collection of biological data for CRISPR and related technologies, appropriate treatment of experimental data, and application of TL to leverage information contained in existing data sets. Nevertheless, these examples show how the development of new tools can be accelerated.

3.1.2 | *Promoter Sequence Analysis and Engineering*

Identifying promoters is a fundamental task due to their significant role in the development of various diseases, such as coronary heart disease, diabetic nephropathy, and tumors (X. Liu et al. 2022). Accurate identification can support risk prediction or the detection of tumor suppressor promoters (X. Liu et al. 2022; Saif et al. 2018). While each promoter sequence is species-specific, promoter regions across prokaryotic and eukaryotic organisms share common structural motifs, making prediction particularly challenging for species with limited available data (Xia et al. 2024).

To address the challenges of promoter identification, a hybrid model was developed to recognize the native sequence as well as the morphological outline of promoters (Y. Wang et al. 2022). The model consists of a promoter sequence network and a deep structural profile to model both the sequence and structure at the same time. TL combined with the new developed sampling approach enhanced the identification of promoter subtypes. This strategy achieved a higher correlation coefficient (MCC) of 0.82 on human samples, compared to reported MCCs ranging from 0.64 to 0.73 in previous studies (Y. Wang et al. 2022). However, it is not clearly stated if the higher MCC was achieved due to the availability of higher quality data (better sampling approach) the TL approach or a combination of both.

3.1.3 | *Biological Data Annotation*

Beyond sequence-level prediction, TL has also advanced biological data annotation tasks such as cell line identification and subcellular classification. To identify cell lines from cultivation images, Tong et al. (2022) developed a multitask DL framework. By using TL, the model was easily adapted to new cell lines, reducing computational cost and training time from 36.4 to 0.078 h by pretraining the model on more than 45,000 images and then fine-tuning on under 900 images (Tong et al. 2022). In another strategy, an automated analysis of microscopy data based on DL was developed to identify yeast cells and perform classification of protein subcellular localization (Kraus et al. 2017). When a pretrained model on over 20,000 yeast cell images was fine-tuned using only 5 additional samples per class, the TL approach achieved an average classification accuracy of 62.7%, which represents a 63.4% performance improvement over training the CNN from scratch (Kraus et al. 2017).

In microbial genomics, TL has also provided new insight into uncultivable organisms, also known as microbial dark matter. To address this, Hoarfrost et al. (2022) developed a DL model trained on microbial genomes to generate sequence representations (embeddings) of short DNA reads, distinguishing sequences by their molecular and functional roles. A TL-enhanced oxidoreductase classifier identified putative enzymes with 82.3% accuracy, including those with very low sequence similarity to known examples.

In computational biology, TL has emerged as a powerful strategy to overcome data scarcity and high dimensionality. It has improved in vivo CRISPR/Cas9 activity prediction, promoter identification with limited sequence data, and cell image classification. However, challenges persist in achieving model generalizability and bridging the gap between proxy training data and true in vivo biological behavior.

3.2 | Biocatalysis

Biocatalysis, the use of enzymes as natural catalysts for chemical reactions, is a key technology for sustainable chemistry and

industrial biotechnology. The establishment of modern high-throughput screening, cultivation, and characterization methods has greatly increased the amount of enzyme-related data. However, the complexity of enzymatic systems, such as the large number of possible substrates and reaction conditions, still makes the optimization of biocatalytic processes challenging. Typical data sets are structured multimodally and hierarchically. Amino acid sequences and 3D structures are represented as graphs, substrates, and products as molecular graphs. Typical regression tasks include the prediction of enzyme performance parameters like reaction rate constants, turnover numbers, conversion, temperature and pH optima, and binding affinities. Classification tasks refer to the prediction of activity to support decisions. Collection of data in biocatalysis is difficult and time-consuming, because it is labor-intensive, expensive, and often involves multiple experimental steps. Results strongly depend on experimental conditions, so-called metadata, which makes standardization in experimentation and data storage difficult. ML has enabled data-driven modeling of enzymatic reactions, advancing both sequence-to-function and function-to-sequence predictions. Comprehensive overviews of ML applications and challenges in biocatalysis are provided by Sampaio and Fernandes and Mazurenko (Mazurenko et al. 2020; Sampaio and Fernandes 2023). However, the scarcity of data for specific enzyme functions remains a key limitation, highlighting TL as a promising approach for modeling enzymatic processes with limited data. To illustrate its versatility, studies spanning protein engineering, synthesis route prediction, and drug discovery are discussed.

3.2.1 | *Protein Engineering*

A neural network which predicts the function of biocatalysts was developed by X. Wang et al. (2025). The structure-based DL network was designed to predict protein and ligand interactions by combining a co-attention network and a transformer-based module. The source model was trained on a large data set (Kd data set > 50,000 entries) and then fine-tuned for the small function-specific data set (initial CalB 233, negative samples 1200 samples). By transferring knowledge from the large source data set, higher prediction accuracy was achieved even with limited data, such that the TL model achieved an average accuracy of 90% for numerical binary classification tasks. This significantly outperformed classical methods tested on a comparable regression task, achieving an RMSE of 12.18 and R^2 of 0.37 (CalB data set) (X. Wang et al. 2025).

A task-transfer approach for the prediction of substrates for ribosomally synthesized and posttranslationally modified peptide biosynthetic enzymes was presented by Clark et al. (2024). Here, a masked language model was fine-tuned to predict substrate preferences. It could be shown that the functional forms that were learned by the model are transferable between related enzymes in the same biosynthetic pathway, achieving an accuracy of 69.7% when classifying substrates without further training (Clark et al. 2024).

In order to find a connection of a certain model and a suitable TL technique, Bikias et al. (2025) set up a comprehensive analysis, which investigates the pairing of three state-of-the-art protein language models with three different TL methods for six different protein engineering data sets. Fine-tuning provided substantial performance gains, especially with increasing task complexity, improving the performance by up to 117.82% over the baseline methods. This strategy can serve as a platform to streamline new processes in protein engineering (Bikias et al. 2025).

3.2.2 | *Prediction of Synthesis Routes*

Kreutter et al. (2021) presented a TL approach, in which a molecular transformer is trained to predict if a certain molecule can serve as a substrate for a particular enzyme. By further developing computer-assisted synthetic planning, the planning of retrosynthesis routes is simplified. The molecular transformer is a sequence-to-sequence ML model, which is trained with chemical reactions from the US patent office (1.8 million reactions) and a special enzymatic biotransformation training set (25,700 reactions) with the help of multitask TL. By integrating the knowledge of chemical reactions with specific enzymatic conversions the model combines comprehensive knowledge with higher prediction accuracies. The best prediction Top 1 accuracy reached 62.2% when using TL with the complete enzyme information, significantly surpassing the 34.3% Top 1 accuracy achieved when the model was trained with the enzymatic data set alone (Kreutter et al. 2021).

Similar to this approach, Probst et al. (2022) constructed a data-driven model for the planning of retrosynthesis routes. The molecular transformer model was trained on a large data set including chemical reactions and after that fine-tuned on specific enzymatic data set, again with multitask TL. The molecular transformer model was trained on the USPTO dataset, containing one million organic chemical reactions, and on an enzymatic reaction dataset (over 56,000 reactions). These datasets were utilized with a 9:1 weighting favoring the USPTO data. The combined data were split into 90% training, 5% validation, and 5% test sets. Due to the prior training on chemical reaction, the model is generalized and can be used even with a limited enzymatic data set. Apart from that, the fine-tuning enabled flexibility and faster fitting times (Probst et al. 2022).

3.2.3 | *Drug Discovery*

In silico drug discovery research often faces the challenge of limited data, making it well-suited for leveraging existing databases through TL. Cai et al. (2020) provide an overview of the various TL methods used in drug discovery, recent advances in the field, and highlight the progress of deep TL. In this field, TL is often the only viable approach, since training data sets typically contain only a small number of molecules (Awale et al. 2019). Nevertheless, TL applications in drug discovery remain in their early stages, and direct comparisons with other ML methods are hindered by the lack of standardized metrics and benchmark data sets (Cai et al. 2020). An application for TL in drug discovery is demonstrated by X. Wang et al. (2023), who developed a DL model for molecular docking to generate novel molecules with improved binding affinities. By pretraining a generative transformer on amino acid composition, sequence order, and physicochemical properties, and subsequently fine-tuning (1400 active compounds) it via TL, they successfully produced candidate molecules with higher predicted docking scores (X. Wang et al. 2023).

Overall, TL in biocatalysis has emerged as a powerful approach to address data scarcity and improve predictive accuracy across enzyme-related applications. It enables efficient adaptation of pretrained models for protein engineering, synthesis route design, and drug discovery, achieving significant gains in accuracy and data efficiency. However, the reliability of TL is consistently challenged by the imbalance and scarcity observed within target data sets (e.g., transferases heavily dominating EC enzyme classes (Probst et al. 2022)), and the success of the transfer hinges fundamentally on the inherent relatedness between the source and target tasks.

3.3 | Bioreactor Modeling and Monitoring

Bioreactor engineering is central to biotechnology, enabling the controlled cultivation of microorganisms, fungi, mammalian cells, and algae for producing chemicals, pharmaceuticals, and other high-value products. Accurate modeling facilitates process optimization, control, and scale-up. Bioreactor data are predominantly a multivariate time-series organized into discrete batches, combining high-frequency online measurements (e.g., pH and DO) with sparsely sampled offline variables that capture key performance indicators (KPIs), such as viable cell density (VCD) and product titer (Baako et al. 2024; Mbiki 2022; Yu et al. 2025). Despite the abundance of raw sensor data, the number of informative labeled instances mapping inputs to outputs remains very limited (Mbiki 2022; Baako et al. 2024; Riezzo et al. 2025). Data collection is limited by long culture durations and labor-intensive and costly offline analyses which limit the availability of high-quality training data. Most modeling tasks focus on regression problems aimed at quantitatively predicting continuous variables such as growth and productivity, with classification reserved for tasks like species identification, image-based morphology analysis, and fault detection (Zhu et al. 2018; Koksal and Aydin 2025; Y. Liu et al. 2024; X. Wang, Zhou et al. 2024). While still emerging, early applications of TL in bioreactor modeling and monitoring are highly promising to tackle these challenges, spanning cell growth prediction, process control, and real-time analytics.

3.3.1 | *Modeling Cell Growth and Product Formation*

A major application involves predicting KPIs directly from process data. Mbiki (2022) compared traditional long short-term memory (LSTM) ANNs with fine-tuning and feature extraction TL for CHO cell cultures producing mAbs. Using process variables (pH, temperature, gas flow) and metabolite concentrations as inputs, fine-tuning outperformed traditional ML models, especially with limited data, and was most effective for VCD prediction (Mbiki 2022). However, the study did not report quantitative error metrics for these improvements. The performance differences are only qualitatively recognized from figures in the original publication (Mbiki 2022).

Similarly, Rogers et al. (2022) applied TL via parameter fine-tuning to microalgal lutein production, transferring process knowledge from eight batches from a well-characterized *Chlorella sorokiniana* (*C. sorokiniana*) strain to a newly isolated strain. Data augmentation prevented overfitting of small data sets by adding Gaussian noise to the data, while a forward ANN architecture updated network parameters at each epoch. Two parameter transfer strategies were explored for the two different use cases. First, exploring different model structures while adding extra nodes to the original ANN trained on the source data. Either in input, output, or hidden layers and fine-tuning while penalizing large deviations from the original parameters. The approach improved prediction accuracy of lutein production by up to 50% (MAE) and reduced uncertainty by over 80% (MASD), enabling robust simulations with only one to three batches of data in the target domain (Rogers et al. 2022).

Using the same process as a case study, lutein production in *C. sorokiniana*, Kay et al. (2023) integrated hybrid modeling and TL to transfer knowledge from a detailed kinetic model built on eight experiments using *Desmodesmus* sp. Using only two experiments for training, the TL model accurately reproduced process trends and provided reliable state estimates and uncertainty bounds. Compared to the benchmark model, the hybrid TL approach effectively compensated for missing information (Kay et al. 2023). Although the original work does not report explicit quantitative metrics comparing the TL model to the baseline approach, the improved performance is apparent from a comparison of the predicted time series (Kay et al. 2023). Recently, a similar strategy for hybrid digital twin development for lutein production using *C. sorokiniana* was proposed, fine-tuning ANN components while retaining shared kinetic structures (Figure 3) (Riezzo et al. 2025). The hybrid model was adapted via TL to a nanoparticle-enhanced strain of the same species, using two batches of experimental data in the target domain. The TL model achieved a 27% MAPE reduction and halved prediction uncertainty (from 18.9% to 9.8%) in lutein prediction compared to the benchmark (Riezzo et al. 2025).

Similarly, Jiang et al. (2024) introduced a kinetic-assisted ML framework for predicting *Chlamydomonas reinhardtii* growth curves, integrating a kinetic model with instance-based TL. Among several ML algorithms tested, such as decision trees, polynomial regression, ANN, and RF, RF performed best. By aligning 121 synthetic kinetic data sets with sparse experimental data from 25 experiments, the TL model improved prediction fit from $R^2 = 0.60$ to $R^2 = 0.91$, demonstrating strong performance under data-scarce conditions (Jiang et al. 2024).

In wastewater treatment, Koksal and Aydin (2025) developed a physics-informed TL method for dissolved oxygen prediction via parameter transfer. A source model trained on simulated or plant data transferred selected neural network parameters to the target model. Embedding mass balance equations improved robustness and physical consistency under noisy, small-data conditions. The TL model reduced test and validation error (MSE) by 27% and up to 59%, respectively, compared to a standard LSTM model, while using the same datapoints available (Koksal and Aydin 2025).

Beyond parameter transfer, Rogers et al. (2025) proposed an interpretable model-structural TL framework that modifies mechanistic model kinetics using symbolic regression and ANN feature attribution. Applied to yeast fermentation for astaxanthin production, the method converted black-box ML corrections into interpretable differential equations, revealing metabolic differences between strains using only three experiments. The framework identified structural shifts such as Contois instead of Monod kinetics and introduced biomass decay and product reversal terms. This method accelerates

model identification, enhances interpretability, and supports model-based experimental design, facilitating high-fidelity digital twin development for novel processes. Furthermore, it offers regulatory advantages by enhancing transparency and interpretability, addressing the increasing demand for explainable ML in process engineering (Rogers et al. 2025).

3.3.2 | *Cross-Recipe and Cross-Condition Process Monitoring*

Another key application involves adapting monitoring models across similar but not identical batch processes. Zhu et al. (2018) developed a systematic TL framework for penicillin fermentation monitoring, combining Principal Component Analysis-based and k-means clustering for similarity assessment. Depending on quantitative or qualitative similarity, the framework applies phase-based Procrustes analysis or multiphase Bayesian networks for knowledge transfer. Using this strategy, model adaptation from a source domain composed of 80 batches to a target domain of 40 batches was over 200 times faster than full retraining and maintained a false-alarm rate below 2% (Zhu et al. 2018).

Similar challenges in *Pichia pastoris* fermentation were addressed by using TCA to align data distributions from different operating conditions and create a soft sensor. This TL-based soft sensor improved robustness by aligning marginal data distributions from source and target fermentation conditions before modeling. By leveraging data from multiple known fermentation conditions, the model accurately predicts cell and product concentrations under new, unseen conditions, resulting in a 26%–29% reduction in prediction error compared to the optimized soft sensor trained without domain adaptation (B. Wang, Yu, et al. 2024). Nonetheless, the authors do not clearly state the total number of individual batches used to populate the source and target sets for the final analysis.

Expanding TL toward interpretable models, Yu et al. (2025) introduced an Ensemble Kalman Filter (EnKF) that dynamically adapts a mechanistic CHO source model across different cell lines, culture scales, and operating conditions. By assimilating sparse daily measurements (e.g., VCD, mAb titer, and key metabolites) and incorporating uncertainty in 8 state variables and 24 kinetic parameters based on one single data set, the EnKF updates the model in real time and produces reconstructed trajectories and parameter estimates. Applied to six different data sets, it delivered increasingly accurate long-term forecasts, capturing scale effects, temperature-shift metabolism, and cell line–specific lactate consumption. Although this approach is implemented through online model adaptation, with parameters being recursively updated as new data become available, the explicit reuse of prior knowledge from a source domain (100 mL shake flasks) to a distinct target domain (900 mL bioreactors) justifies its inclusion as a TL technique, even though it does not follow the classical static domain adaptation (Yu et al. 2025).

3.3.3 | *Process Optimization and Control*

Beyond prediction and monitoring, TL can also enhance optimization and control strategies. Petsagkourakis et al. (2020) applied policy-gradient reinforcement learning (RL) to optimize fed-batch processes, fine-tuning pretrained policies from simpler mechanistic models. Three case studies of increasing complexity, including *Arthrospira platensis* phycocyanin production with nonsmooth, phase-dependent dynamics, demonstrated that the TL-enhanced RL approach required fewer batch runs, outperformed nonlinear model predictive control (NMPC) from the onset of online implementation, and computed control actions in 0.002 s versus NMPC's 2–4 min. Training used four epochs with 25 batches each (100 batches total) and could have been stopped earlier without loss of performance. These results demonstrate that combining RL with TL can significantly accelerate the development of adaptive and efficient control strategies in bioprocess engineering, reducing experimental costs and enabling real-time decision-making for complex, nonlinear systems (Petsagkourakis et al. 2020).

3.3.4 | *Emerging Bioprocess Analytics*

Beyond conventional variables, TL supports image-based and spectroscopic process monitoring. Wang and colleagues used a CNN pretrained on a general data set containing over 328,000 images (Lin et al. 2015) to segment CHO cells in situ, fine-tuning only selected layers. Morphological descriptors such as eccentricity, aspect ratio, circularity, and equivalent diameter were extracted from the segmented images. Using data augmentation of the target domain from 46 to 184 labeled images by flipping and rotation and TL, the precision and sensitivity score was enhanced by over 23%, compared to the benchmark model. This strategy led to significant improvements in the segmentation of clustered cells and a reduction in the need for manual annotation (X. Wang, Zhou, et al. 2024).

Additionally, a soft sensor for real-time monitoring of recombinant protein production in *Escherichia coli* was established by Y. Liu et al. (2024) using Raman spectroscopy. Combining unsupervised spectral labeling, feature extraction, regression, and TL, the approach leveraged an existing optical density source model trained on 1302 samples to predict protein concentration using only 20 offline samples from 2 experiments. It achieved an RMSE of 0.028 g/L and a maximum deviation of 0.0017 g/L, providing a noninvasive, accurate, and robust solution for fermentation monitoring with minimal data and preprocessing (Y. Liu et al. 2024).

For bioreactor engineering, TL shows a trend toward with physics-informed and hybrid modeling, combining mechanistic insight with data-driven flexibility (Kay et al. 2023; Koksal and Aydin 2025; Riezzo et al. 2025). TL has proven effective for KPI prediction, model adaptation across scales and organisms, and accelerating digital twin and control strategy development. Integrating TL with RL enhances process optimization (Petsagkourakis et al. 2020), while coupling with mechanistic models improves interpretability and uncertainty quantification. However, noisy or mismatched source data, insufficient feature overlap, and suboptimal architectures can reduce transfer efficiency (Koksal and Aydin 2025; Rogers et al. 2022, 2025). While the physics-informed TL approach increases model interpretability, it may also complicate convergence, depending on the model architecture (Koksal and Aydin 2025).

3.4 | Chromatography and Downstream Processing

Chromatography is one of the core unit operations in bioprocessing, essential for purifying and analyzing complex

biomolecules such as mAbs, peptides, and metabolites. However, modeling chromatographic processes for performance optimization remains challenging due to complex nonlinear interactions among stationary-phase chemistry, mobile-phase composition, pH, temperature, and molecular structure. Across the specific applications, the typical data structure varies. Small molecules are commonly represented as text-based SMILES strings, molecular graphs with atoms and bonds as nodes and edges, three-dimensional conformations, or numerical molecular descriptors and fingerprints (Fedorova et al. 2022; Osipenko et al. 2021; Ju et al. 2021; Kwon et al. 2023; Yang, Ji, Fan, et al. 2021; Wu et al. 2025). Protein and peptide data are typically encoded as amino acid sequences, which are often transformed into vector representations to capture biochemical properties (Ma et al. 2018; Feng et al. 2022). For bioprocess design tasks, data structures are based on sets of critical process parameters, including variables such as loading time, feed flow rate, and inlet concentration (Galeazzi et al. 2025). TL is used for both regression tasks, such as predicting retention times and classification tasks including design space feasibility (Fedorova et al. 2022; Kwon et al. 2023; Long et al. 2022; Galeazzi et al. 2025). Source models are often pretrained on thousands to millions of samples from public data sets (Fedorova et al. 2022; Osipenko et al. 2021; Long et al. 2022; Feng et al. 2022), while target data sets typically comprise only a few dozen to a few hundred labeled samples, reflecting the high cost and effort of experimental data generation (Kwon et al. 2023; Feng et al. 2022; Galeazzi et al. 2025). Data acquisition is often constrained by material scarcity and resource-intensive experimental procedures, raising interest for TL approaches (Osipenko et al. 2021; Wu et al. 2025; Galeazzi et al. 2025). While TL applications in preparative chromatography are still limited, studies in analytical liquid chromatography for metabolomics highlight its potential to exploit historical data.

3.4.1 | *Retention Time (RT) Prediction for Small Molecules and Peptides*

DL combined with TL has shown strong promise for predicting RTs of small molecules and peptides. For small molecules, TL has been implemented mainly through parameter transfer and fine-tuning. Several strategies use the METLIN Small Molecule Retention Time (SMRT) data set, which contains over 80,000 small molecules analyzed via reversed-phase liquid chromatography, as their primary source domain (Ju et al. 2021; Fedorova et al. 2022; Yang, Ji, Lu, et al. 2021; Kwon et al. 2023). Ju et al. (2021) introduced a TL model, which used weighted autoencoders pretrained on the METLIN SMRT data set for RT prediction. It outperformed classical ML methods such as RF by over 40% in the median absolute error and surpassed RF, Gradient Boosting, and standard deep neural networks on 17 diverse, small data sets containing maximal 665 compounds (Ju et al. 2021). Using the same source data set, Federova et al. (2022) used a one-dimensional CNN pretrained freezing early feature-extraction layers and retraining later ones on target data, which was composed of ~100–500 molecules. This approach retained general molecular features while adapting to specific chromatographic systems, improving generalization across stationary phases and eluents and resulted in an error reduction of 38% on prediction (Fedorova et al. 2022).

Graph neural networks (GNNs) have also proven effective for small-molecule RT prediction. Yang and colleagues developed the GNN-RT model, pretrained on the METLIN SMRT data set and fine-tuned to new chromatographic methods via parameter transfer. By freezing general layers and adapting system-specific ones using ~200–500 target RT datapoints, the model improved prediction accuracy by at least 30% over traditional methods such as RF (Yang, Ji, Lu, et al. 2021). Another GNN pretrained on synthetic data for hydrophilic interaction chromatography data on 306,000 molecules and fine-tuned on limited experimental data from the target domain (880 molecules) improved prediction accuracy and reduced false positives (Yang, Ji, Fan, et al. 2021). This strategy reduced the MAE by ~45% compared to training from scratch on the test set (Yang, Ji, Fan, et al. 2021). Furthermore, it outperformed existing traditional models by 18%–32% in prediction error (MAE), and the predicted RTs filtered out nearly 60% false positive candidates on average (Yang, Ji, Fan, et al. 2021). Kwon et al. (2023) systematically evaluated multiple TL strategies for graph isomorphism networks using METLIN SMRT as a source data set, finding that full parameter fine-tuning yielded the most stable and accurate predictions even with small data sets ranging from 38 to 532 datapoints in the target domain (Kwon et al. 2023).

An alternative method applies natural language processing principles using SMILES strings to represent molecular structures. A transformer model pretrained on one million molecules and fine-tuned on smaller data sets (~250–500 molecules) outperformed classical ML models such as RF and Gradient Boosting. The approach directly learned molecular embeddings without predefined descriptors like pKa or logD and reduced the median average error by 34%–43% compared to projection-based methods, depending on the data set (Osipenko et al. 2021).

Additionally, TL has been applied for the scale-up of preparative chromatography of small molecules. Using a GNN, models were built on more than 4500 datapoints from experiments in small columns (4 g) and were then fine-tuned using ~500 datapoints on larger scales (8–40 g), improving prediction accuracy from near-zero to an R^2 of approximately 0.75 and generalizing across eluent systems and bonded column types (Wu et al. 2025). This highlights TL's potential to reduce experimental runs during process development by exploiting physical similarities across scales.

For peptide RT prediction, Ma et al. (2018) developed a model using convolutional and capsule networks pretrained on a peptide data set containing over 145,000 peptides and fine-tuned on smaller data sets ranging from ~3000 to ~14,000 peptides. The model achieved highly accurate RT predictions with only a few hundred peptides, capturing subtle effects such as phosphorylation or oxidation. By embedding amino acids analogously to words in large language models, it required only around 100 peptides for fine-tuning, compared to the several thousand typically needed. Leveraging TL improved accuracy, reducing the time deviation window ($\Delta t_{95\%}$) by approximately 40%, while maintaining $R^2 = 0.975$–0.987 even under limited data conditions (Ma et al. 2018).

3.4.2 | *Monoclonal Antibody Purification and Solubility Prediction*

Although to a smaller extent, TL has also been applied to mAb purification, particularly protein A chromatography,

which is a major cost driver in mAb production.Galeazzi et al. (2025) developed a ML-enhanced design-space identification framework integrating experimental and synthetic data. Two pathways were compared: a fully data-driven route for large data sets, and a TL route for limited data using mechanistic model simulations as pretraining. An ANN pretrained on ~4000 simulated data points (using the mechanistic model from (Grom et al. 2018)) and fine-tuned with a small set of experimental protein A breakthrough data accurately identified feasible operating regions based on dynamic binding capacity. The TL-based model matched the accuracy of data-rich models that were built on six or more runs but outperformed them under low-data conditions, with three or fewer experimental runs, by up to 27.3% in accuracy. By transferring knowledge from mechanistic models instead of re-parameterizing them, this approach reduced both experimental and computational demands, aligning with Quality by Digital Design principles (Galeazzi et al. 2025).

In addition to purity, solubility is a key quality attribute in mAb process development. Feng et al. (2022) applied TL using embeddings from a protein language model, pretrained on 250 million unlabeled protein sequences, to predict mAb solubility directly from amino acid sequences. Applied to 260 labeled antibodies across different subclasses, the model showed strong correlations with experimental solubility data obtained using polyethylene glycol as a destabilizing agent. The pretrained embeddings substantially improved performance compared to traditional one-hot protein encoding, increasing R^2 by up to 60% and reducing the RMSE by 21%. Even when using more complex neural network architectures, TL further enhanced performance, increasing R^2 from 0.66 to 0.69 and lowering RMSE from 4.60 to 4.40. These results demonstrate that pretrained molecular representations effectively capture key physicochemical properties relevant to mAb purification and formulation (Feng et al. 2022).

In chromatography and downstream processing, TL is mainly implemented by pretraining on large data sets to learn molecular embeddings, followed by fine-tuning on small, system-specific data sets. Its integration with convolutional, transformer-based, and GNNs has enhanced predictions for small molecules and peptides, while hybrid TL models have improved generalization and enabled data-efficient Quality by Digital Design for mAbs (Galeazzi et al. 2025). In small-molecule applications, fine-tuning requires on the order of 100–200 labeled target samples (Fedorova et al. 2022; Yang, Ji, Fan, et al. 2021). In some cases, fine-tuning with fewer than 50 samples markedly reduced prediction errors compared to alternative approaches (Kwon et al. 2023). In contrast, training deep models from scratch on such limited data sets is generally ineffective, leading to overfitting, poor generalization, and unstable predictions (Osipenko et al. 2021; Wu et al. 2025; Yang, Ji, Fan et al. 2021). However, in high-data settings, the added complexity of TL offers little gain (Galeazzi et al. 2025) and fine-tuning complex models on small data sets risks overfitting (Wu et al. 2025). Generalized models also often face challenges in extrapolating to under-represented molecules and lack mechanistic interpretability (Feng et al. 2022).

4 | Challenges and Future Expectations for the Application of Transfer Learning in Bioprocess Engineering

The successful implementation of TL in bioprocessing remains limited by model adaptability and insufficient standardization. Accessible and standardized databanks are essential to advance TL applications (Mazurenko et al. 2020; Wu et al. 2025). However, existing bioprocess data sets are often small or inconsistent, lacking comprehensive benchmarks (Cai et al. 2020), which complicates model training, validation, and comparison. To determine which TL techniques are most suitable for a given application, standardized and openly accessible data sets are required, along with clearly defined evaluation metrics to assess whether TL provides a tangible benefit. In particular, TL models should be systematically compared against non-TL approaches in terms of reductions in training time, improvements in prediction accuracy (e.g., RMSE), or decreases in predictive uncertainty. Studies should also transparently report the number of additional data points used during TL to clarify its effectiveness. At present, the lack of such standardized benchmarks makes it difficult to assess which approaches adhere to best practices for each application field.

From a computational perspective, training advanced models from scratch is often highly time- and resource-intensive. For example, pretraining molecular structure prediction models can require up to 90 h (Osipenko et al. 2021). Although deeper neural networks can improve accuracy, increasing model complexity may raise computational costs by more than 30% (Baako et al. 2024). Moreover, completely retraining large protein language models is frequently impractical due to GPU memory constraints, restricting TL adoption to environments with access to high-end hardware, although selective fine-tuning strategies, such as only retraining the final layer of an ANN, can substantially reduce these requirements (Bikias et al. 2025). Additionally, fine-tuning of complex architectures can be prone to overfitting under data-scarce conditions (Cai et al. 2020). Moreover, combining these complex architectures with physics-informed strategies could hinder convergence of the numerical solution and reduce model reliability (Koksal and Aydin 2025). Moreover, the absence of standardized data flow architectures complicates the transformation of raw bioprocess data into analysis-ready formats, making fine-tuning workflows technically challenging (Mbiki 2022; Bikias et al. 2025).

Practical implementation also introduces methodological challenges, including how to optimally design experiments in the target domain to maximize information gain given a specific source domain, and how and when to update TL models as new data from additional domains become available. Finally, TL performance depends heavily on the similarity between source and target systems, yet still no standardized metric for measuring this similarity has been established for bioprocess models (Cai et al. 2020). Most TL approaches focus on fine-tuning or updating parameters rather than modifying model structures, even when structural changes could be necessary to capture critical differences (Riezzo et al. 2025). Nonetheless, some studies report differences between TL approaches for the same task, particularly between fine-tuning and feature extraction. In protein engineering, performance strongly depends on task complexity. For complex and diverse data sets, where training

data span diverse protein families, fine-tuning offered clear advantages, whereas for simpler tasks with closely related sequence distributions, feature extraction applied to only part of the model achieved comparable results while being more computationally efficient (Bikias et al. 2025). In RT prediction for liquid chromatography, fine-tuning consistently outperformed both feature extraction and training from scratch, especially when the target data was scarce (Kwon et al. 2023). Similarly, in CHO cell line bioprocessing, fine-tuning enhanced predictions of VCD and titer with fewer epochs and limited data compared to feature extraction (Mbiki 2022).

Negative transfer remains poorly understood and is typically identified only after the modeling effort has been completed (Rogers et al. 2025). In bioprocessing, only a limited number of studies explicitly report negative effects of TL. Negative transfer should be more systematically addressed in future work, as understanding its causes can provide valuable guidance on when and how TL should be applied. While studies from other disciplines have explored methods to assess domain relatedness prior to transfer (Eaton et al. 2008; Ge et al. 2014; Z. Wang et al. 2019), such assessments are largely absent from the bioprocessing literature reviewed in this article. The underlying causes of negative transfer are commonly linked to domain misalignment (Cai et al. 2020), including shifts in data distributions, for example, discrepancies between in vitro proxy measurements and in vivo biological behavior (Kota et al. 2021). To mitigate the risk of negative transfer, several strategies have been proposed, including quantifying task relatedness prior to transfer using chemical or molecular similarity measures (Cai et al. 2020; Clark et al. 2024), selectively freezing model layers to retain generalizable knowledge (Bikias et al. 2025; Cai et al. 2020; Fedorova et al. 2022), and applying parameter regularization to prevent excessive deviation from a reliable source model (Kay et al. 2023; Riezzo et al. 2025; Rogers et al. 2022). Additional approaches include combining multiple source domains to reduce bias (Cai et al. 2020; B. Wang, Yu et al. 2024) and embedding TL within hybrid or physics-informed frameworks to enforce biological consistency (Kay et al. 2023; Koksal and Aydin 2025; Riezzo et al. 2025). Applying interpretable TL strategies can identify the potential of negative transfer, as substantial and unintuitive structural modifications required to fit new data might indicate that the source model's physical assumptions may not hold for the target system (Rogers et al. 2025). Collectively, these strategies help reduce the likelihood of negative transfer while enabling more robust and reliable knowledge reuse across bioprocessing applications.

Integrating data-based approaches into industrial biomanufacturing workflows faces cultural and regulatory barriers. ML models are frequently perceived as opaque "black boxes," undermining trust among users and conflicting with regulatory requirements for explainability and traceability (Di Bonito et al. 2024; Baako et al. 2024; Rogers et al. 2025). This challenge is reinforced by a disciplinary knowledge gap between process engineers and data scientists (Mbiki 2022). Additionally, bioprocess development remains largely product-focused rather than knowledge-driven, making the adoption of ML methods unlikely unless clear incentives, such as substantial reductions in experimental effort, are demonstrated (Mbiki 2022; Helleckes et al. 2024).

Future developments in TL for bioprocessing are expected to focus on hybrid and physics-informed modeling to improve data efficiency, interpretability, and generalization (Galeazzi et al. 2025; Kay et al. 2023; Riezzo et al. 2025; Rogers et al. 2025). Integrating mechanistic understanding with data-driven modeling will enhance model transparency and uncertainty quantification while reducing experimental demands. TL is also expected to play a key role in digital twin development (Riezzo et al. 2025) and automated process control (Petsagkourakis et al. 2020), while protein stability and substrate prediction models could benefit from the use of pre-trained embedding techniques that capture physicochemical and structural molecular properties (Bikias et al. 2025; Clark et al. 2024; Feng et al. 2022). Data augmentation techniques can further mitigate data scarcity (Riezzo et al. 2025; Rogers et al. 2022), and the establishment of standardized benchmarks, open-access databases, and explainability frameworks will be crucial for ensuring reproducibility, facilitating regulatory approval, and promoting the broader adoption of TL-based tools in bioprocess engineering.

5 | Conclusion

As illustrated by numerous applications, TL is emerging as a promising approach for flexible and data-efficient model development in bioprocessing, for which labeled data are scarce and costly to obtain. By leveraging existing data sets, TL can extract meaningful information from historical data and enhance model robustness through increased variability during training. Looking ahead, TL is expected to play a key role in advancing ML-based bioprocess optimization by extending the applicability of ML methods to small-data scenarios. Furthermore, the development of explainable TL models through integration with mechanistic frameworks is likely to strengthen regulatory confidence and accelerate the industrial adoption of data-driven approaches.

Author Contributions

Daniel Barón Díaz: conceptualization, writing – original draft (Sections 1, 2, 3.3, 3.4, and 4), figure preparation, writing – review and editing. **Anna-Lena Drommershausen:** writing – original draft (Sections 3.1 and 3.2), figure preparation, writing – review and editing. **Alexander Grünberger:** conceptualization, review and editing. **Dirk Holtmann:** conceptualization, writing – review and editing. All authors have read and approved the final manuscript.

Acknowledgments

Large language models were used to assist with language editing, structuring, and summarizing general content in this manuscript. These tools were not used to create the original draft or develop original scientific interpretations. All content was critically reviewed, verified, and approved by the authors, who take full responsibility for the integrity and accuracy of the work. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant No. 536355345. Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

This article is a review and does not report original experimental data. Therefore, no data sets were generated or analyzed during the current study.

References

Alzubi, J., A. Nayyar, and A. Kumar. 2018. "Machine Learning From Theory to Algorithms: An Overview." *Journal of Physics: Conference Series* 1142: 012012.

Anwar, R. W., M. Abrar, and F. Ullah. 2023. "Transfer Learning in Brain Tumor Classification: Challenges, Opportunities, and Future Prospects." In *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*, 24–29. IEEE. https://ieeexplore.ieee.org/document/10392830/.

Awale, M., F. Sirockin, N. Stiefl, and J.-L. Reymond. 2019. "Drug Analogs From Fragment-Based Long Short-Term Memory Generative Neural Networks." *Journal of Chemical Information and Modeling* 59: 1347–1356.

Baako, T.-M. D., S. K. Kulkarni, J. L. McClendon, S. W. Harcum, and J. Gilmore. 2024. "Machine Learning and Deep Learning Strategies for Chinese Hamster Ovary Cell Bioprocess Optimization." *Fermentation* 10: 234.

Blöbaum, L., C. Haringa, and A. Grünberger. 2023. "Microbial Lifelines in Bioprocesses: From Concept to Application." *Biotechnology Advances* 62: 108071.

Bikias, T., E. Stamkopoulos, and S. T. Reddy. 2025. "PLMFit: Benchmarking Transfer Learning With Protein Language Models for Protein Engineering." *Briefings in Bioinformatics* 26, no. 4). https://doi.org/10.1093/bib/bbaf381.

Butean, A., I. Cutean, R. Barbero, J. Enriquez, and A. Matei. 2025. "A Review of Artificial Intelligence Applications for Biorefineries and Bioprocessing: From Data-Driven Processes to Optimization Strategies and Real-Time Control." *Processes* 13: 2544.

Cai, C., S. Wang, Y. Xu, et al. 2020. "Transfer Learning for Drug Discovery." *Journal of Medicinal Chemistry* 63: 8683–8694.

Clark, J. D., X. Mi, D. A. Mitchell, and D. Shukla. 2024. "Substrate Prediction for RiPP Biosynthetic Enzymes via Masked Language Modeling and Transfer Learning." Preprint, arXiv. http://arxiv.org/abs/2402.15181.

Di Bonito, L. P., L. Campanile, F. Di Natale, M. Mastroianni, and M. Iacono. 2024. "eXplainable Artificial Intelligence in Process Engineering: Promises, Facts, and Current Limitations." *Applied System Innovation* 7: 121.

Eaton, E., M. desJardins, and T. Lane. 2008. "Modeling Transfer Relationships Between Learning Tasks for Improved Inductive Transfer." In *Machine Learning and Knowledge Discovery in Databases*, Lecture Notes in Computer Science, edited by W. Daelemans, B. Goethals, K. Morik, Vol. 5211, 317–332. Springer. http://link.springer.com/10.1007/978-3-540-87479-9_39.

Fedorova, E. S., D. D. Matyushin, I. V. Plyushchenko, A. N. Stavrianidi, and A. K. Buryak. 2022. "Deep Learning for Retention Time Prediction in Reversed-Phase Liquid Chromatography." *Journal of Chromatography A* 1664: 462792.

Feng, J., M. Jiang, J. Shih, and Q. Chai. 2022. "Antibody Apparent Solubility Prediction From Sequence by Transfer Learning." *iScience* 25: 105173.

Fink, M., M. Cserjan-Puschmann, D. Reinisch, and G. Striedner. 2021. "High-Throughput Microbioreactor Provides a Capable Tool for Early Stage Bioprocess Development." *Scientific Reports* 11: 2056.

Galeazzi, A., S. Sachio, E. Edwards, D. Hilton, and M. M. Papathanasiou. 2025. "Machine Learning Enhanced Process Design in Protein A Chromatography." *Journal of Chromatography A* 1759: 466193.

Ge, L., J. Gao, H. Ngo, K. Li, and A. Zhang. 2014. "On Handling Negative Transfer and Imbalanced Distributions in Multiple Source Transfer Learning." *Statistical Analysis and Data Mining* 7: 254–271.

Grom, M., M. Kozorog, S. Caserman, A. Pohar, and B. Likozar. 2018. "Protein A Affinity Chromatography of Chinese Hamster Ovary (CHO) Cell Culture Broths Containing Biopharmaceutical Monoclonal Antibody (mAb): Experiments and Mechanistic Transport, Binding and Equilibrium Modeling." *Journal of Chromatography B* 1083: 44–56.

Guo, J., T. Wang, C. Guan, et al. 2018. "Improved sgRNA Design in Bacteria via Genome-Wide Activity Profiling." *Nucleic Acids Research* 46: 7052–7069.

Ham, D. T., T. S. Browne, P. N. Banglorewala, et al. 2023. "A Generalizable Cas9/sgRNA Prediction Model Using Machine Transfer Learning With Small High-Quality Datasets." *Nature Communications* 14: 5514.

Helleckes, L. M., J. Hemmerich, W. Wiechert, E. Von Lieres, and A. Grünberger. 2023. "Machine Learning in Bioprocess Development: From Promise to Practice." *Trends in Biotechnology* 41: 817–835.

Helleckes, L. M., C. Wirnsperger, J. Polak, G. Guillén-Gosálbez, A. Butté, and M. Von Stosch. 2024. "Novel Calibration Design Improves Knowledge Transfer Across Products for the Characterization of Pharmaceutical Bioprocesses." *Biotechnology Journal* 19. https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/biot.202400080.

Hoarfrost, A., A. Aptekmann, G. Farfañuk, and Y. Bromberg. 2022. "Deep Learning of a Bacterial and Archaeal Universal Language of Life Enables Transfer Learning and Illuminates Microbial Dark Matter." *Nature Communications* 13: 2606.

Hodgson, B. J. 2005. "Hybrid Modelling of Bioprocesses." PhD diss., University College London. https://api.semanticscholar.org/CorpusID:106591871.

Janković Babić, R. 2024. "A Comparison of Methods for Image Classification of Cultural Heritage Using Transfer Learning for Feature Extraction." *Neural Computing and Applications* 36: 11699–11709.

Jiang, M., X. Cao, Z. Wang, et al. 2024. "A Kinetic-Assisted Growth Curve Prediction Method for *Chlamydomonas reinhardtii* Incorporating Transfer Learning." *Bioresource Technology* 394: 130246.

Ju, R., X. Liu, F. Zheng, X. Lu, G. Xu, and X. Lin. 2021. "Deep Neural Network Pretrained by Weighted Autoencoders and Transfer Learning for Retention Time Prediction of Small Molecules." *Analytical Chemistry* 93: 15651–15658.

Karimi Alavijeh, M., I. Baker, Y. Y. Lee, and S. L. Gras. 2022. "Digitally Enabled Approaches for the Scale Up of Mammalian Cell Bioreactors." *Digital Chemical Engineering* 4: 100040.

Kay, S., H. Kay, A. W. Rogers, and D. Zhang. 2023. "Integrating Hybrid Modelling and Transfer Learning for New Bioprocess Predictive Modelling." In *Computer Aided Chemical Engineering*, edited by A. C. Kokossis, M. C. Georgiadis, and E. Pistikopoulos, 52, 2595–2600. Elsevier. https://www.sciencedirect.com/science/article/pii/B9780443152740504121.

Kim, H. E., A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt. 2022. "Transfer Learning for Medical Image Classification: A Literature Review." *BMC Medical Imaging* 22: 69.

Koksal, E. S., and E. Aydin. 2025. "A Hybrid Approach of Transfer Learning and Physics-Informed Modelling: Improving Dissolved Oxygen Concentration Prediction in an Industrial Wastewater Treatment Plant." *Chemical Engineering Science* 304: 121088.

Kota, P. K., Y. Pan, H.-A. Vu, M. Cao, R. G. Baraniuk, and G. Bao. 2021. "The Need for Transfer Learning in CRISPR-Cas Off-Target Scoring." Preprint. http://biorxiv.org/lookup/doi/10.1101/2021.08.28.457846.

Kraus, O. Z., B. T. Grys, J. Ba, et al. 2017. "Automated Analysis of High-Content Microscopy Data With Deep Learning." *Molecular Systems Biology* 13: 924.

Kreutter, D., P. Schwaller, and J.-L. Reymond. 2021. "Predicting Enzymatic Reactions With a Molecular Transformer." *Chemical Science* 12: 8648–8659.

Kunze, M., S. Roth, E. Gartz, and J. Büchs. 2014. "Pitfalls in Optical On-Line Monitoring for High-Throughput Screening of Microbial Systems." *Microbial Cell Factories* 13: 53.

Kwon, Y., H. Kwon, J. Han, et al. 2023. "Retention Time Prediction Through Learning From a Small Training Data Set With a Pretrained Graph Neural Network." *Analytical Chemistry* 95: 17273–17283.

Lin, T.-Y., M. Maire, S. Belongie, et al. 2015. "Microsoft COCO: Common Objects in Context." Preprint, arXiv. http://arxiv.org/abs/1405.0312.

Liu, X., Y. Xu, Y. Luo, and L. Teng. 2022. "Prokaryotic and Eukaryotic Promoters Identification Based on Residual Network Transfer Learning." *Bioprocess and Biosystems Engineering* 45: 955–967.

Liu, Y., X. Zhou, T. Wang, et al. 2024. "Genetic Algorithm-Based Semisupervised Convolutional Neural Network for Real-Time Monitoring of *Escherichia coli* Fermentation of Recombinant Protein Production Using a Raman Sensor." *Biotechnology and Bioengineering* 121: 1583–1595.

Long, X., C. Ma, H. Sheng, et al. 2022. "Transfer Learning Analysis for Subvisible Particle Flow Imaging of Pharmaceutical Formulations." *Applied Sciences* 12: 5843.

Ma, C., Y. Ren, J. Yang, Z. Ren, H. Yang, and S. Liu. 2018. "Improved Peptide Retention Time Prediction in Liquid Chromatography Through Deep Learning." *Analytical Chemistry* 90: 10881–10888.

Mazurenko, S., Z. Prokop, and J. Damborsky. 2020. "Machine Learning in Enzyme Engineering." *ACS Catalysis* 10: 1210–1223.

Mbiki, S. 2022. "Scalable Data-Driven Predictive Modelling and Analytics for CHO Process Development Optimisation." PhD diss., Clemson University. https://onlinelibrary.wiley.com/doi/10.1002/biot.201200065.

Mondal, P. P., A. Galodha, V. K. Verma, et al. 2023. "Review on Machine Learning-Based Bioprocess Optimization, Monitoring, and Control Systems." *Bioresource Technology* 370: 128523.

Mowbray, M., T. Savage, C. Wu, et al. 2021. "Machine Learning for Biochemical Engineering: A Review." *Biochemical Engineering Journal* 172: 108054.

Osipenko, S., K. Botashev, E. Nikolaev, and Y. Kostyukevich. 2021. "Transfer Learning for Small Molecule Retention Predictions." *Journal of Chromatography A* 1644: 462119.

Pan, S. J., and Q. Yang. 2010. "A Survey on Transfer Learning." In *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22, 1345–1359. IEEE. https://doi.org/10.1109/tkde.2009.191.

Peng, J., T. T. Khuat, K. Musial, and B. Gabrys. 2026. "Machine Learning Methods for Small Data and Upstream Bioprocessing Applications: A Comprehensive Review." *Biotechnology Advances* 87: 108749.

Petsagkourakis, P., I. O. Sandoval, E. Bradford, D. Zhang, and E. A. Del Rio-Chanona. 2020. "Reinforcement Learning for Batch Bioprocess Optimization." *Computers & Chemical Engineering* 133: 106649.

Probst, D., M. Manica, Y. G. Nana Teukam, A. Castrogiovanni, F. Paratore, and T. Laino. 2022. "Biocatalysed Synthesis Planning Using Data-Driven Learning." *Nature Communications* 13, no. 1). https://doi.org/10.1038/s41467-022-28536-w.

Riezzo, L., H. Kay, Y. Feng, K. Jing, and D. Zhang. 2025. "Accelerating Bioprocess Digital Twin Development by Integrating Hybrid Modelling With Transfer Learning." *Chemical Engineering Journal* 511: 162018.

Rogers, A. W., F. Vega-Ramon, A. Lane, P. Martin, and D. Zhang. 2025. "Interpretable-AI-Based Model Structural Transfer Learning to Accelerate Bioprocess Model Construction." *Biotechnology and Bioengineering* 122: 2819–2831. https://doi.org/10.1002/bit.70026.

Rogers, A. W., F. Vega-Ramon, J. Yan, E. A. Del Río-Chanona, K. Jing, and D. Zhang. 2022. "A Transfer Learning Approach for Predictive Modeling of Bioprocesses Using Small Data." *Biotechnology and Bioengineering* 119: 411–422.

Saif, I., Y. Kasmi, K. Allali, and M. M. Ennaji. 2018. "Prediction of DNA Methylation in the Promoter of Gene Suppressor Tumor." *Gene* 651: 166–173.

Sampaio, P. S., and P. Fernandes. 2023. "Machine Learning: A Suitable Method for Biocatalysis." *Catalysts* 13: 961.

Tong, L., A. Corrigan, N. R. Kumar, et al. 2022. "An Automated Cell Line Authentication Method for AstraZeneca Global Cell Bank Using Deep Neural Networks on Brightfield Images." *Scientific Reports* 12: 7894.

Wang, Z., Z. Dai, B. Poczos, and J. Carbonell. 2019. "Characterizing and Avoiding Negative Transfer." In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11285–11294. IEEE. https://ieeexplore.ieee.org/document/8953565/.

Wang, X., C. Gao, P. Han, et al. 2023. "PETrans: De Novo Drug Design With Protein-Specific Encoding Based on Transfer Learning." *International Journal of Molecular Sciences* 24: 1146.

Wang, Y., Q. Peng, X. Mou, et al. 2022. "A Successful Hybrid Deep Learning Model Aiming at Promoter Identification." *BMC Bioinformatics* 23: 206.

Wang, B., A. Yu, H. Wang, and J. Liu. 2024. "Modeling and Optimization of an Enhanced Soft Sensor for the Fermentation Process of *Pichia pastoris*." *Sensors* 24: 3017.

Wang, X., G. Zhou, L. Liang, et al. 2024. "Deep Learning-Based Image Analysis for In Situ Microscopic Imaging of Cell Culture Process." *Engineering Applications of Artificial Intelligence* 129: 107621.

Wang, X., J. Zhou, J. Mueller, et al. 2025. "BioStructNet: Structure-Based Network With Transfer Learning for Predicting Biocatalyst Functions." *Journal of Chemical Theory and Computation* 21: 474–490.

Weiss, K., T. M. Khoshgoftaar, and D. Wang. 2016. "A Survey of Transfer Learning." *Journal of Big Data* 3: 9.

Wu, W., H. Xu, Y. Xu, et al. 2025. "Intelligent Column Chromatography Prediction Model Based on Automation and Machine Learning." *Chem* 11: 102598.

Xia, Y., X. Du, B. Liu, S. Guo, and Y.-X. Huo. 2024. "Species-Specific Design of Artificial Promoters by Transfer-Learning Based Generative Deep-Learning Model." *Nucleic Acids Research* 52: 6145–6157.

Yang, Q., H. Ji, X. Fan, Z. Zhang, and H. Lu. 2021. "Retention Time Prediction in Hydrophilic Interaction Liquid Chromatography With Graph Neural Network and Transfer Learning." *Journal of Chromatography A* 1656: 462536.

Yang, Q., H. Ji, H. Lu, and Z. Zhang. 2021. "Prediction of Liquid Chromatographic Retention Time With Graph Neural Networks to Assist in Small Molecule Identification." *Analytical Chemistry* 93: 2200–2206.

Yu, L., A. Rio Chanona, and C. Kontoravdi. 2025. "Model-Enabled Knowledge Transfer Across Cell Lines, Culture Scales and Conditions." *Systems Biology*. Preprint. http://biorxiv.org/lookup/doi/10.64898/2025.11.30.691385.

Zhang, Z., A. R. Lamson, M. Shelley, and O. Troyanskaya. 2023. "Interpretable Neural Architecture Search and Transfer Learning for Understanding CRISPR–Cas9 Off-Target Enzymatic Reactions." *Nature Computational Science* 3: 1056–1066.

Zhu, J., Y. Yao, and F. Gao. 2018. "Transfer of Qualitative and Quantitative Knowledge for Similar Batch Process Monitoring." *IEEE Access* 6: 73856–73870.

Zhuang, F., Z. Qi, K. Duan, et al. 2021. "A Comprehensive Survey on Transfer Learning." *Proceedings of the IEEE* 109: 43–76.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.
Supplementary Material.