Chapter 3
Role of Machine Learning in Bioprocess Engineering: Current Perspectives and Future Directions

Ashutosh Singh and Barkha Singhal

Abbreviations

AAD	Absolute average deviation
ANN	Artificial neural network
CNN	Convolutional neural networks
CT	Classification tress
DLAB	Deep learning for antibodies
GBSA	Generalized Born surface area
HER2	Human epidermal growth factor receptor 2
K_{cat}	Catalytic turnover number
Mabs	Monoclonal antibodies
ML	Machine learning
MM	Molecular mechanic
NIR	Near-infrared spectroscopy
RF	Random forest
RMSE	Root mean squared error
SAX	Symbolic aggregate approximation
SCM	Set covering machine
SG	Spatial graph
SSF	Simultaneous saccharification and fermentation
SVM	Support vector machine

A. Singh · B. Singhal (✉)
School of Biotechnology, Gautam Buddha University, Greater Noida, Uttar Pradesh, India
e-mail: barkha@gbu.ac.in

D. Singh et al. (eds.), *Design and Applications of Nature Inspired Optimization*, Women in Engineering and Science, https://doi.org/10.1007/978-3-031-17929-7_3

1 Introduction

The stupendous stride to achieve economic and environmental sustainability accelerates the rising demand for bioproducts [1]. The harnessing of bioproducts requires the understanding of holistic development of biological processes from raw materials to synthesis and purification of bioproducts and valorization of biowaste to make value-added products at an industrial scale [2, 3]. Therefore, for accomplishing at such height process, engineering principles integrating with natural sciences, such as physics, chemistry, biology with chemical, and system engineering, are playing a pivotal role in translating biological knowledge into making products of commercial importance. Bioprocess engineering is the thrust area that needs to be updated with technological innovations, though the past decade has envisaged significant developments in formulating extensive mechanistic and physiochemical empirical models for simulating the growth pattern of microbial biomass and product formation, calculating the fermented broth rheological parameters and dynamics for bioreactor scale-up, optimization of bioseparation unit designs, synthesis of new enzymes proteins, and analysis of metabolic flux ([4, 5]. However, the complexity of the biological system still posits certain inherent challenges that need to be addressed for industrial purposes [6].

As the fourth industrial revolution industry 4.0 has already gained momentum, there is a dire need for digitalization and inculcation of advanced process analytics and computational biology tools for accelerating the arena of bioprocess engineering. Currently, the paradigmatic shift from physical modeling to data-driven modeling using machine learning approaches contributed to the voluminous data generation in the bioindustry arena [7]. The elucidation of complex biological relationships in data form demonstrated great potential for the bioprocess research community and engineers to "scale up" as well as "scale down" the bioprocess for explicit commercial use. The arena of bioprocess engineering is quite vast constituting various branches from research and development to biomanufacturing consisting of metabolic engineering, bioreaction engineering, protein engineering, synthetic biology, biomaterials, and biocatalysis. The applications of ML are represented in Fig. 3.1. Currently, various ML algorithms have been utilized to circumvent the biological complexities during bioprocess optimization; thus, the accelerating pace of machine learning is invoking a renaissance in this area [8]. Therefore, the present chapter advocates various embodiments of machine learning applications in the bioprocess engineering sector and also predicate current challenges and future prospects. The technology is still evolving; therefore, this chapter doesn't cover the comprehensive aspects, but various dimensions of ML approaches has been described through various case studies of bioprocess engineering.

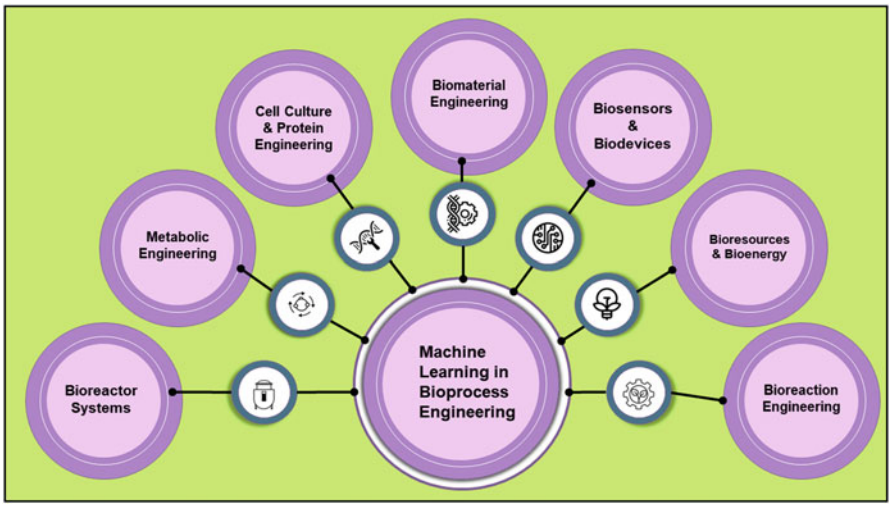

Fig. 3.1 Applications of machine learning is the arena of bioprocess engineering

2 Approaches of Machine Learning in Bioprocess Engineering

Bioprocess engineering is the application of interdisciplinary field to commercialize the bioproduct from lab to industry. The process seems to be quite heterogenous due to the complex requirement of living cells and their prevailed diversity. The commercialization of product various process parameters ingredients and their composition as well as interactions plays pivotal role. There are numerous challenges associated with mathematical modeling and simulations due to multi-parametric nature of biological data. Therefore, applications of ML methods have shown promising potential in tackling complex problems of bio-production at large scale. ML algorithms are categorized into four different learning categories, namely, supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. This classification is based on configuration of various data set based on different problems on which ML algorithm will develop mathematical correlation to build a model followed by the solution of the defined problem. With rising technological advancements computational tools generate voluminous data of biological origin therefore past decade has seen massive growth of various algorithms of ML in the arena of development and manufacturing of bioproducts. The contribution of various ML algorithms in the arena of bioproduct development has been summarized in Fig. 3.2**.**

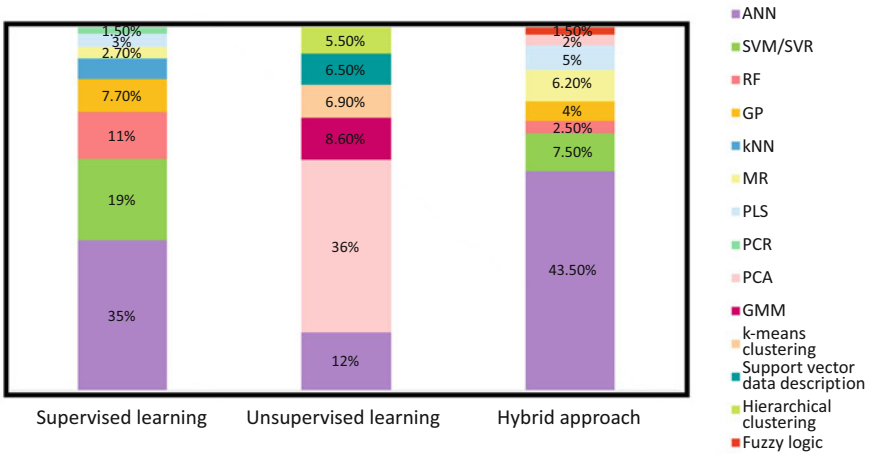

Fig. 3.2 Statistical coverage of various machine learning algorithms in bioprocess engineering

3 Why Machine Learning Strategies Are Needed in Bioprocess Engineering

The essence of bioprocess engineering is the scale-up of cellular factories for the overproduction of commercial metabolites. The scale-up is a multistep method commencing the fermentation and optimization of cells from bench scale (~250 mL–5 L) to pilot scale (~20–200 L) to industrial-scale processes (>1000 L). The fermentation of cells at a larger scale is considered to be a complex and multi-parametric process, in which different process variables, such as pH, temperature, aeration rate, media composition, mixing regime, fermentation time, and feed rate, is affecting the cell growth, product formation, and host cell physiology. Therefore, at industrial scale, the fermentation process is unpredictable; therefore, the central task of scale-up is to fine-tune all these process variables to perform the stable and robust production of desired bioproducts, because the slight change in any process variables confers significant impact on the overall productivity of cells [9, 10]. Thus, scale-up is a time-consuming and costly process; therefore, the industry needs advanced computational scientific methods for accelerating the fermentation process in bioreactors beyond the classical methods. The advent of automation, sensors for controlling, and monitoring the process parameters, comprehensive data collection, and archiving revolutionize the modern fermentation process [11]. Therefore, these huge data can be leveraged for various machine learning algorithms for better prediction, finding the bioprocess failure points, and improving the process outcomes in lieu of better product yield. However, the main bottlenecks of bioprocess data constituting heterogeneity in terms of collection of both online pH, oxygen uptake rate, dissolved oxygen, optical (cell) density, flow rate, off-gas production, etc.) and offline data (various metabolite concentrations, substrate consumption rates). Apart from that, certain data are binary or categorical

(ON/OFF nutrient feed setting), and some data such as the concentration of product has been collected at the final time point. The high variability in data sets with respect to fermentation time and fermentation runs necessitates the preprocessing of data for extracting temporal trends for training into machine learning premises [12, 13].

The research studies reported various preprocessing methods, such as wavelet decomposition methods [14], mean envelope filter methods, vector casting method [15], and Fourier transform and symbolic aggregate approximation (SAX) method, that represents temporal data profile as representative segments for the analysis through machine learning approaches [16]. Initially, decision trees, ANN, and genetic algorithm-based ML were applied for fermentative modeling and identification of optimum input variables by analyzing the data of 69 fed-batch fermentation for predicting the process output, including product concentration, biomass density, and volumetric productivity [17]. Similarly, ANN-based modeling followed by optimization through a genetic algorithm was reported for the production of xylitol. The predictive models for xylose consumption, biomass concentration, and xylitol production were based on analyzing the data of 27 fermentation batches with multiple inputs, and the product titer was enhanced from 59.4 to 65.7 g/L [18]. More recently, the advancement in bioreactor designing enables the generation of continuous online data that is being used for the optimum control and optimization of bioprocess by reinforcement learning. However, this method suffered the limitation of being built on fixed models while requiring continuous updates and improvement with respect to surplus data generation in automated fermentation systems [19]. Therefore, to improvise model-free reinforcement, learning methods have been developed and successfully applied for controlling final ethanol titers during yeast fermentations. Moreover, these methods have been instrumental in controlling coculture species biomass abundances, controlling reactor temperatures [20], optimizing product yields [21], and optimizing a downstream product separation unit [22]. However, the requirement of a large amount of data limits its wider utilization, but there is still a scope of improvement by seeing the marvelous credentials of ML approaches [23]. Thus, despite current challenges, the data of various fermentation systems gives an appealing opportunity to develop various ML algorithms for finding the most appropriate process conditions.

4 Applications of Machine Learning in Bioprocess Engineering (Case Studies)

Approaches of Machine Learning in Biorefinery: A Case Study

The rising demand of environmental pollution, reduction in fossil fuels, and increasing ecosystem resilience paved the way for finding various avenues for renewable energy sources. Among various sources, lignocellulosic biomass is offering the most

promising feedstock for the development of the bioenergy paradigm. Though bioethanol and biodiesel are the most preferentially utilized product from the lignocellulosic biomass, the compositional variability among various biomass sources offers a diverse array of products that leads to the conception of multiproduct biorefinery [24]. The major operational bottleneck of the biorefinery is the natural heterogeneity and spatial variability of biomass. Recently, machine learning and data analytics has been envisaged as a prospective tool for predicting this biomass variability and easing the way of standardization of biomass properties that leads to the consistency in the biorefining process. Though advanced sophisticated analytical techniques, such as rapid near-infrared (NIR) spectroscopy and hyperspectral imaging, have been used for predicting the chemical composition of the biomass and its conversion performances, these techniques are unable to correlate a large amount of data and higher complexity of biomass [25]. Thus, a machine learning framework based on an ANN has been recently implied for correlating biomass chemical composition and their conversion performances and finding a correlation of physical properties of tissue powders along with handling and grinding performances [26]. It is envisioned that the predictive models will be used to produce conversion ready and highly flowable feedstock and provide decision centric view to researchers and multiple stakeholders. More recently, machine learning approaches, such as random forest, artificial neural networks (ANNs), and classification trees (CTs), have been used for alleviating one of the critical bottlenecks for bioethanol production that is enzymatic hydrolysis. The simultaneous saccharification and fermentation (SSF) process posit a prominent and feasible strategy for reducing the capital cost for the production of bioethanol from lignocellulosic biomass [29].

Thus, ML approaches have been used for visualizing the effects of time, temperature, inoculum size, and biomass on bioethanol fermentation in SSF.

ANN Based Model

An ANN-based model is used for predicting the yield of bioethanol by implementing three layers of data sets and finding optimum conditions using R software and AMORE library (http://cran.r-project.org/web/packages/AMORE/). The coefficient of determination (R) [2], reduction of root mean squared error (RMSE), and absolute average deviation (AAD) have been calculated by Eqs. (3.1), (3.2), and (3.3), respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_i^{\text{calc}} - Y_i^{\text{exp}}\right)^2}{\sum_{i=1}^{n} \left(Y_i^{\text{calc}} - Y_m\right)^2}, \tag{3.1}$$

$$\text{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n} \left(Y_i^{\text{calc}} - Y_i^{\text{exp}}\right)^2\right]^{\frac{1}{2}} = \sqrt{\text{MSE}}, \tag{3.2}$$

$$\mathrm{AAD} = \frac{1}{n}\sum\left|\frac{Y_i^{\mathrm{calc}} - Y_i^{\mathrm{calc}}}{Y_i^{\mathrm{calc}}}\right|, \tag{3.3}$$

where n = number of points, Y_i^{calc} = predicted value, Y_i^{exp} = experimental value, Y_m = average value of all experimental data, and MSE = mean square error.

Based on these equations, the optimal ethanol concentration and the optimal ethanol conversion value were found that lead to the determination of optimal volumetric productivity of ethanol by Eq. (3.4) [27],

$$I_j = \frac{\sum \frac{|W_{jm}^{ih}|}{\sum |W_{km}^{ih}|} \cdot |W_{mn}^{hO}|}{\sum \sum \frac{|W_{km}^{ih}|}{\sum |W_{km}^{ih}|} \cdot |W_{mn}^{hO}|}, \tag{3.4}$$

where I_j = relative importance of jth input variable on the ethanol conc.; N_{i} and N_{h} = number of input and hidden neurons, respectively; W_{s} = connection weights; subscripts i, h, and O refer to input, hidden, and output layers, respectively; and subscripts k, m, and n represent input, hidden, and output neurons, respectively.

Random Forest Model

This model has been used for predicting the effects of variables in SSF using the library of R language [28]. A total of 1000 RFs comprising different numbers of trees and variables in each of the branches has been assessed. The assessment of the optimal RF model was performed using two rando data sets having 2/3 for training and 1/3 for test, and the values of R[2], RMSE, and AAD have been calculated [29].

Classification Tress-Based Model

This model has been used for making decisions based on the entropy of the process. The current study includes the C5.0 script that has been used by utilizing the default library of R (http://cran.r-project.org/web/packages/C50/) for predicting the concentration of ethanol [29].

Thus, by the above discussion, it has been clearly seen that ML methodologies have tremendous potential to evaluate the various process parameters for bioethanol production without prior knowledge of kinetics and inhibition process. An overview of applications of ML approaches in biorefinery sector has been represented in Fig. 3.3. Thus, in the futuristic scenario, more comprehensive studies have been warranted for overcoming the various technical gaps for the commercialization of biorefineries.

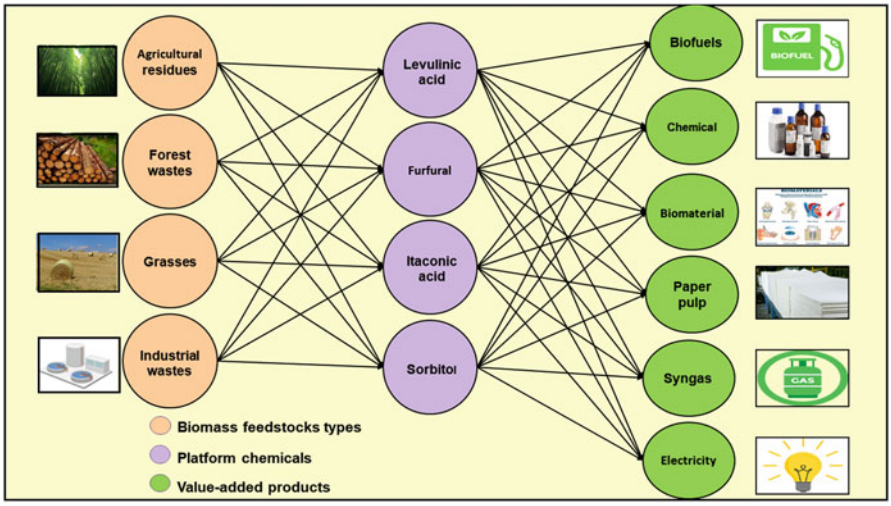

Fig. 3.3 An overview of applications of machine learning approaches in biorefinery

Approaches of Machine Learning in Monoclonal Antibody Production: A Case Study

The past decade has seen a stupendous ride in the production of biotherapeutics, more preferably monoclonal antibodies, for the treatment of a variety of chronic disorders like cancers and autoimmune and inflammatory diseases [30]. This continuous surge has been attributed to higher efficacy, specificity, reduced toxicity, and less side effects conferred by monoclonal antibodies. Apart from that, the production of monoclonal antibodies is considered to be a costly, time-consuming, and fastidious endeavor due to the requirement of the high standards and stability during production, storage, and transportation [31–34]. Due to the proteinaceous nature, these antibodies always remain susceptible to various physical and chemical degradation pathways with varied conditions encountered during the whole life cycle [35, 36]. Thus, there is a pressing need to overcome these challenges for the sustainable production of this important class of biotherapeutics. There are various avenues from the design and prediction of antigen specificity of monoclonal antibodies to the prediction of various liquid formulations for effective delivery of these compounds inside the body in which various domains of machine learning have been used. Recently, the structure-based deep learning for antibodies (DLAB) database has been developed for virtual screening and prediction of putative binding of antibodies against antigen as a target [37]. Based on this database, Reddy et al. reported the prediction of antigen specificity of therapeutic antibody trastuzumab against human epidermal growth factor receptor 2 (HER2) as an antigen. The studies involved the screening of thousands of lead molecules by analyzing 1 × 108

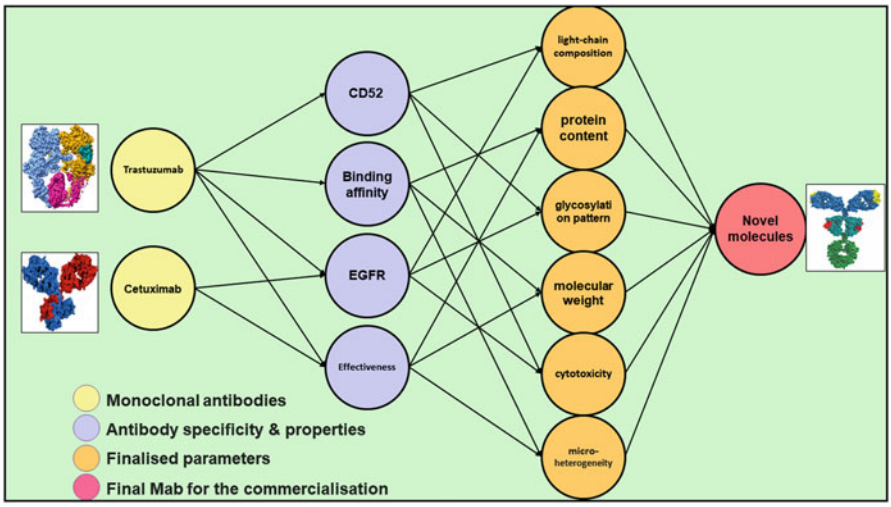

Fig. 3.4 Use of machine learning methods in commercial production of monoclonal antibodies

trastuzumab variants against 1 × 106 variants of HER2 based on viscosity, solubility, clearance, and immunogenicity [38]. Similarly, more comprehensive studies related to the determination of molecular descriptors affecting the viscosity of monoclonal antibodies have been reported by Trout et al. A decision tree-based machine learning framework has been used for predicting the net charge and high viscosity index of monoclonal antibodies [39].

That studies significantly contribute for the assessment of rheological behavior that affects the delivery of these therapeutic molecules. More recently, a Bayesian optimization algorithm has been developed for the screening of formulations of mAbs. The formulation comprises various excipients such as thermal stabilizers, amino acids buffering agents, surfactants and tonicity modifiers that imparts a significant effect on the stability of proteins and their storage [40]. Thus, this approach of ML leads to the acceleration in the design of formulations with optimum excipients and parallelization of operations in mAbs development. Figure 3.4 is representing the applications of ML algorithms in monoclonal antibody manufacturing.

Thus, based on the above discussion, it is conceivable to comprehend that ML approaches provide a novel, innovative, and accelerated platform for the discovery, development, and manufacturing of monoclonal antibodies and can be used for other biotherapeutics.

Approaches of Machine Learning for Antibiotic Production: A Case Study

The serendipitous discovery of penicillin as a life-saving drug during world war has been proved to be a cornerstone discovery in modern medicine. Then, the golden era of antibiotics has been visualized, but their overwhelming use leads to a deadly menace of antibiotic resistance, and it is estimated that by 2050, 10 million death per year will occur due to drug resistance diseases [41]. The discovery of novel antibiotics is from a natural source, which is plagued by dereplication problems [42]. Thus, the approaches of machine learning are proving to be eye-opening methods that have the capacity to search large amounts of data with accelerating speed. Recently, genotype-based machine learning models, such as support vector machine (SVM) and set covering machine (SCM), have been used as a promising diagnostic tool to predict the resistance of commonly used antibiotics, including tetracycline, ampicillin, sulfisoxazole, trimethoprim, and enrofloxacin, against the whole genome of 96 isolates of *Actinobacillus pleuropneumoniae* [43]. Moreover, halicin molecule was identified through the screening of 6000 chemical compounds that not only have the potency to treat diabetes but also found to exhibit strong activity against *Mycobacterium tuberculosis* and other hard-to-treat microbes [44]. ML is not only to accelerate the discovery of novel antibiotics, but different algorithms can be helpful for predicting the susceptibility towards antibiotics. Figure 3.5 represents the role of ML in antibiotic discovery for finding novel antimicrobials. Recently, the single centric study was performed to assess the eight algorithms of ML for predicting the resistance toward antimicrobials by taking

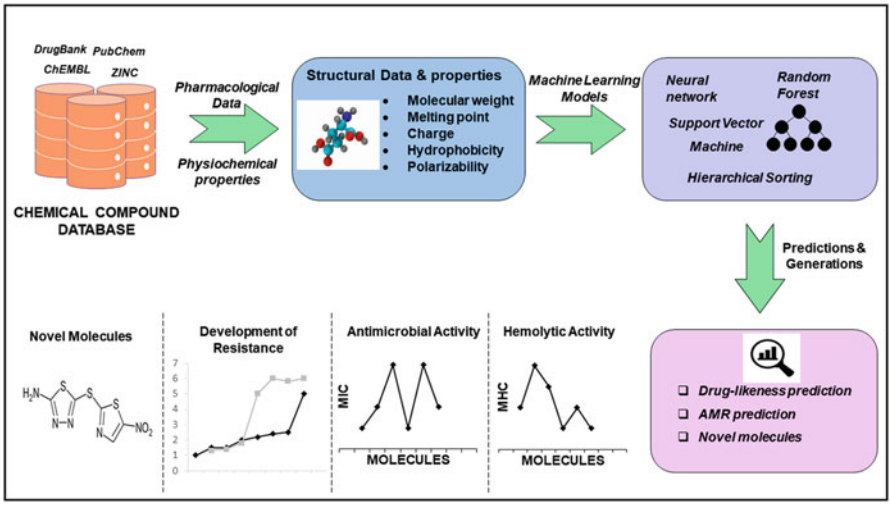

Fig. 3.5 Applications of machine learning methods in antibiotic production

demographic data from patients, gram staining, and site of infection [45]. These studies will be helpful for clinicians in the selection of appropriate antibiotic therapy. Thus, in the future ML holds tremendous potential to alleviate the global threat of antibiotic resistance and is helpful in maintaining the stewardship of antibiotics.

Machine Learning in Protein Engineering: A Case Study

The continuous surge in the production of bioproducts needs sustainable bioprocessing portfolios. The development of industrial strains requires a thorough understanding of genome organization, cellular metabolism, and enzymes. The overproduction of various products requires engineering of their biosynthetic pathway and enzymes that are still unknown to the scientific community. Thus, the development of novel biosynthetic pathways and the engineering of enzymes can spur the overproduction of industrially important metabolites. Recently, catalytic turnover (K_{cat}) of enzymes has been evaluated in *E. coli* through machine learning approaches. The diverse properties of enzymes, such as structural properties, network properties, assay conditions, and biochemical mechanism information, have been considered for generating ML models. This in vivo and in vitro prediction of K*cat* will be helpful for implementing the information of genome-scale metabolic models for correlating the expression of the proteome in *E. coli*. [46] Furthermore, the scope of substrate specificity of enzymes has also been predicted with ML models. Four machine learning models, along with molecular modeling and docking tools, namely, support vector machines, random forest, logistic regression, and gradient-boosted decision trees, have been developed for evaluating the substrate specificity of bacterial nitrilases that hydrolyzed the nitrile compounds to the corresponding carboxylic acids and ammonia. The accuracy of substrate prediction leads to a better catalytic activity of enzymes that facilitates the overproduction of metabolites [47]. Recently, the affinities of protein-ligand binding have been performed with deep learning ML models, including three-dimensional (3D)-convolutional neural networks (3D-CNNs), spatial graph neural networks (SG-CNNs), and their fusion models. These models predicted the binding free energies based on docking pose coordinates, docking scores, and molecular mechanic/generalized Born surface area (MM/GBSA) calculations. [48]. An overview of the utilization of ML methods in the realm of protein engineering has been summarized in Fig. 3.6. These studies will be playing a pivotal role in the drug discovery paradigm. Thus, based on the above discussion, enzyme engineering is the backbone in improving the bioprocess design and development and fosters the path for sustainable biomanufacturing.

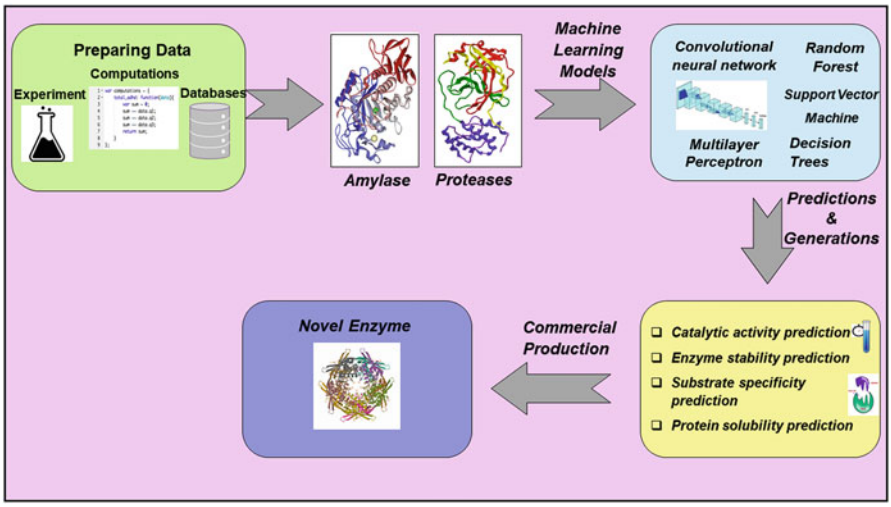

Fig. 3.6 An overview of applications of machine learning methods in protein engineering

5 Current Challenges and Future Prospects

The world is moving toward digitalization and bioproduct development, and manufacturing is no longer the exception for adoption of advanced technologies. Though machine learning methodologies have proven their mettle in other sectors in an efficient manner, biomanufacturing sector is still reluctant to adopt ML as the standardized tool for the development of bioprocess. The skepticism related to catastrophic consequences of defective products inhibited their wider adoption. There are technological challenges such as lack of representative datasets for development, eating, and validation of model limits their operability at commercial scale. The uncertainty of prediction due to the multi-parametric nature of biological data confers additional roadblock for the acceptance. The complexity of models further limits their correlation analysis with biological process. Inspite of these challenges, rising technological innovations in ML and computing will definitely overcome these challenges, and it is clearly envisioned that the ML approaches hold a bright future in upscaling the development of bioproducts through process engineering approaches.

6 Conclusion

Machine learning not only transformed the scientific paradigm but also leads to a gigantic leap in the productivity of industrial manufacturing. The digitalization accompanied with machine learning approaches creates novel history in the

biomanufacturing too. The various approaches such as unsupervised and supervised models both are quite useful in various facets from process development to purification of bioproducts from living cells. The present chapter entails the current application of ML in various bioproducts and their commercial manufacturing through process engineering principles. The chapter represents various case studies of diversified bioproducts portfolios from high value to low value. The immersive applications of ML proved its utility in reducing the cost and time of industry that is considered to be a major economic consideration. The road of utilizing ML is not smooth currently, but in future, the vision of using these concepts for bringing the transformations in bioprocess engineering sector is certainly on horizon.

Acknowledgment The authors greatly endorsed the support of Gautam Buddha University for writing this chapter.

Conflicts of Interest/Competing Interests Authors declare that there is no conflict of interest in this work.

Glossary

Antibiotics	It is a class of antimicrobial substances that are used to kill infectious bacteria.
Antibiotic resistance	It is the type of resistance that is developed by microorganism against the effect of an antibiotic.
Biomanufacturing	It is a type of industrial production that utilizes biological systems to create commercially-important bioproducts.
Bioprocess engineering	A bioprocess is any method that uses living cells or their elements (e.g., enzymes, chloroplasts) to produce a product, whereas engineering is the science of coming up with complex machines or processes.
Bioreactor	It is an apparatus used to grow microorganisms in a controlled environment.
Docking	It is a tool for predicting the interaction, conformation, and orientation of a ligand in binding site of protein.
Entropy	It is a measurable property that is associated with the degree of randomness of a system.
Fermentation	Fermentation is the process by which molecules such as glucose are broken down into a simpler substance.
Hyperspectral imaging	It is a spectroscopic technique that captures and processes an image at very large number of wavelengths.
Monoclonal antibodies	These are laboratory-made proteins that can be used as substitutes for antibodies to enhance or modify the immune system.

Near-infrared spectroscopy	It is a spectroscopic technique that deals with the electromagnetic spectrum within the near-infrared area (780–2500 nm).
Protein engineering	It is the method of developing novel proteins with desired properties.
Scale-up	It is the process of increasing the scale of fermentation.
Scale-down	It is the process of decreasing the scale of fermentation.
Simultaneous saccharification and fermentation	It is a procedure that mixes enzymatic hydrolysis with fermentation to gain value-added products in an individual step.
Surfactants	These are the chemical compounds that are used to lower down the surface tension between two phases.
Synthetic biology	It is a multidisciplinary research area that involves engineering of organisms for producing novel useful substances.
Tonicity	It is the potential of a solution to change the water content surrounding the cell.
Viscosity	It is a measure of resistance of a fluid toward deformation by shear stress.

References

1. Gao, S., Song, W., Guo, M.: The integral role of bioproducts in the growing bioeconomy. Ind. Biotechnol. **16**(1), 13–25 (2020)
2. Petrides, D.: Bioprocess design and economics. Bioseparat. Sci. Eng., 1–83 (2000)
3. Chavan, S., Yadav, B., Atmakuri, A., Tyagi, R.D., Wong, J.W., Drogui, P.: Bioconversion of organic wastes into value-added products: a review. Bioresour. Technol. **344**, 126398 (2022)
4. Mears, L., Stocks, S.M., Albaek, M.O., Sin, G., Gernaey, K.V.: Mechanistic fermentation models for process design, monitoring, and control. Trends Biotechnol. **35**(10), 914–924 (2017)
5. Sakthiselvan, P., Meenambiga, S.S., Madhumathi, R.: Kinetic studies on cell growth. Cell Growth. **13** (2019)
6. Brooks, S.M., Alper, H.S.: Applications, challenges, and needs for employing synthetic biology beyond the lab. Nat. Commun. **12**(1), 1–16 (2021)
7. Montáns, F.J., Chinesta, F., Gómez-Bombarelli, R., Kutz, J.N.: Data-driven modeling and learning in science and engineering. Comptes Rendus Mécanique. **347**(11), 845–855 (2019)
8. Mowbray, M., Savage, T., Wu, C., Song, Z., Cho, B.A., Del Rio-Chanona, E.A., Zhang, D.: Machine learning for biochemical engineering: a review. Biochem. Eng. J. **172**, 108054 (2021)
9. Crater, J.S., Lievense, J.C.: Scale-up of industrial microbial processes. FEMS Microbiol. Lett. **365**(13), fny138 (2018)
10. Humphrey, A.: Shake flask to fermentor: what have we learned? Biotechnol. Prog. **14**(1), 3–7 (1998)
11. Carbonell, P., Radivojevic, T., Garcia Martin, H.: Opportunities at the intersection of synthetic biology, machine learning, and automation. ACS Synth. Biol. **8**(7), 1474–1477 (2019)
12. Cheung, J.Y., Stephanopoulos, G.: Representation of process trends—Part I. A formal representation framework. Comput. Chem. Eng. **14**(4–5, 495), –510 (1990a)
13. Cheung, J.Y., Stephanopoulos, G.: Representation of process trends—Part II. The problem of scale and qualitative scaling. Comput. Chem. Eng. **14**(4–5), 511–539 (1990b)

14. Bakshi, B.R., Stephanopoulos, G.: Representation of process trends—III. Multiscale extraction of trends from process data. Comput. Chem. Eng. **18**(4), 267–302 (1994)
15. Gebrekidan, M.T., Knipfer, C., Braeuer, A.S.: Vector casting for noise reduction. J. Raman Spectrosc. **51**(4), 731–743 (2020)
16. Charaniya, S., Hu, W.S., Karypis, G.: Mining bioprocess data: opportunities and challenges. Trends Biotechnol. **26**(12), 690–699 (2008)
17. Coleman, M.C., Buck, K.K., Block, D.E.: An integrated approach to optimization of Escherichia coli fermentations using historical data. Biotechnol. Bioeng. **84**(3), 274–285 (2003)
18. Pappu, S.M.J., Gummadi, S.N.: Artificial neural network and regression coupled genetic algorithm to optimize parameters for enhanced xylitol production by Debaryomyces nepalensis in bioreactor. Biochem. Eng. J. **120**, 136–145 (2017)
19. Qin, S.J., Badgwell, T.A.: MPC. 4th generation. MPC. Fig. 1 Approximate genealogy of linear MPC algorithms. Control. Eng. Pract. **11**, 733–764 (2003)
20. Xie, H., Xu, X., Li, Y., Hong, W., Shi, J.: Model predictive control guided reinforcement learning control scheme. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2020)
21. Treloar, N.J., Fedorec, A.J., Ingalls, B., Barnes, C.P.: Deep reinforcement learning for the control of microbial co-cultures in bioreactors. PLoS Comput. Biol. **16**(4), e1007783 (2020)
22. Hwangbo, S., Sin, G.: Design of control framework based on deep reinforcement learning and monte-carlo sampling in downstream separation. Comput. Chem. Eng. **140**, 106910 (2020)
23. Shin, J., Badgwell, T.A., Liu, K.H., Lee, J.H.: Reinforcement learning–overview of recent progress and implications for process control. Comput. Chem. Eng. **127**, 282–294 (2019)
24. Ulonska, K., König, A., Klatt, M., Mitsos, A., Viell, J.: Optimization of multiproduct biorefinery processes under consideration of biomass supply chain management and market developments. Ind. Eng. Chem. Res. **57**(20), 6980–6991 (2018)
25. Schimleck, L., Dahlen, J., Yoon, S.C., Lawrence, K.C., Jones, P.D.: Prediction of Douglas-fir lumber properties: Comparison between a benchtop near-infrared spectrometer and hyperspectral imaging system. Appl. Sci. **8**(12), 2602 (2018)
26. Ighalo, J.O., Adeniyi, A.G., Marques, G.: Application of artificial neural networks in predicting biomass higher heating value: an early appraisal. Energy Sources. **10**(5), 933–944 (2020)
27. Garson, D.G.: Interpreting neural network connection weights. AI Expert. **6**(4), 46–51 (1991)
28. Breiman, L.: Random forests. Machine learning. J. Biomed. Sci. Eng. **45**(1), 5–32 (2001)
29. Fischer, J., Lopes, V.S., Cardoso, S.L., Coutinho, U., Cardoso, V.L.: Machine learning techniques applied to lignocellulosic ethanol in simultaneous hydrolysis and fermentation. Braz. J. Chem. Eng. **34**, 53–63 (2017)
30. Walsh, G.: Biopharmaceutical benchmarks 2018. Nat. Biotechnol. **36**(12), 1136–1145 (2018)
31. Elgundi, Z., Reslan, M., Cruz, E., Sifniotis, V., Kayser, V.: The state-of-play and future of antibody therapeutics. Adv. Drug Deliv. Rev. **122**, 2–19 (2017)
32. Chi, E.Y., Krishnan, S., Randolph, T.W., Carpenter, J.F.: Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharm. Res. **20**(9), 1325–1336 (2003)
33. Randolph, T.W., Carpenter, J.F.: Engineering challenges of protein formulations. AICHE J. **53**(8), 1902–1907 (2007)
34. Gentiluomo, L., Svilenov, H.L., Augustijn, D., El Bialy, I., Greco, M.L., Kulakova, A., Indrakumar, S., Mahapatra, S., Morales, M.M., Pohl, C., Roche, A.: Advancing therapeutic protein discovery and development through comprehensive computational and biophysical characterization. Mol. Pharm. **17**(2), 426–440 (2019)
35. Krause, M.E., Sahin, E.: Chemical and physical instabilities in manufacturing and storage of therapeutic proteins. Curr. Opin. Biotechnol. **60**, 159–167 (2019)
36. Jiskoot, W., Randolph, T.W., Volkin, D.B., Middaugh, C.R., Schöneich, C., Winter, G., Friess, W., Crommelin, D.J., Carpenter, J.F.: Protein instability and immunogenicity: roadblocks to clinical application of injectable protein delivery systems for sustained release. J. Pharm. Sci. **101**(3), 946–954 (2012)

37. Schneider, C., Buchanan, A., Taddese, B., Deane, C.M.: DLAB: deep learning methods for structure-based virtual screening of antibodies. Bioinformatics. **38**(2), 377–383 (2022)
38. Mason, D.M., Friedensohn, S., Weber, C.R., Jordi, C., Wagner, B., Meng, S.M., Ehling, R.A., Bonati, L., Dahinden, J., Gainza, P., Correia, B.E.: Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. Nature Biomed. Eng. **5**(6), 600–612 (2021)
39. Lai, P.K., Fernando, A., Cloutier, T.K., Gokarn, Y., Zhang, J., Schwenger, W., Chari, R., Calero-Rubio, C., Trout, B.L.: Machine learning applied to determine the molecular descriptors responsible for the viscosity behavior of concentrated therapeutic antibodies. Mol. Pharm. **18**(3), 1167–1175 (2021)
40. Narayanan, H., Dingfelder, F., Condado Morales, I., Patel, B., Heding, K.E., Bjelke, J.R., Egebjerg, T., Butté, A., Sokolov, M., Lorenzen, N., Arosio, P.: Design of biopharmaceutical formulations accelerated by machine learning. Mol. Pharm. **18**(10), 3843–3853 (2021)
41. O'Neill, J.: Antimicrobial resistance: tackling a crisis for the health and wealth of nations. Rev. Antimicrob. Resist. **2014**(4), 1–20 (2014)
42. Cox, G., Sieron, A., King, A.M., De Pascale, G., Pawlowski, A.C., Koteva, K., Wright, G.D.: A common platform for antibiotic dereplication and adjuvant discovery. Cell Chem. Biol. **24**(1), 98–109 (2017)
43. Liu, Z., Deng, D., Lu, H., Sun, J., Lv, L., Li, S., Peng, G., Ma, X., Li, J., Li, Z., Rong, T.: Evaluation of machine learning models for predicting antimicrobial resistance of Actinobacillus pleuropneumoniae from whole genome sequences. Front. Microbiol. **11**, 48 (2020)
44. Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M., MacNair, C. R., French, S., Carfrae, L.A., Bloom-Ackermann, Z., Tran, V.M.: A deep learning approach to antibiotic discovery. Cell. **180**(4), 688–702 (2020)
45. Feretzakis, G., Loupelis, E., Sakagianni, A., Kalles, D., Martsoukou, M., Lada, M., Skarmoutsou, N., Christopoulos, C., Valakis, K., Velentza, A., Petropoulou, S.: Using machine learning techniques to aid empirical antibiotic therapy decisions in the intensive care unit of a general hospital in Greece. Antibiotics. **9**(2), 50 (2020)
46. Heckmann, D., Lloyd, C.J., Mih, N., Ha, Y., Zielinski, D.C., Haiman, Z.B., Desouki, A.A., Lercher, M.J., Palsson, B.O.: Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. Nat. Commun. **9**(1), 1–10 (2018)
47. Mou, Z., Eakes, J., Cooper, C.J., Foster, C.M., Standaert, R.F., Podar, M., Doktycz, M.J., Parks, J.M.: Machine learning-based prediction of enzyme substrate scope: application to bacterial nitrilases. Proteins Struct. Funct. Bioinf. **89**(3), 336–347 (2021)
48. Jones, D., Kim, H., Zhang, X., Zemla, A., Stevenson, G., Bennett, W.D., Kirshner, D., Wong, S. E., Lightstone, F.C., Allen, J.E.: Improved protein–ligand binding affinity prediction with structure-based deep fusion inference. J. Chem. Inf. Model. **61**(4), 1583–1592 (2021)