Artificial intelligence and machine learning applications for cultured meat

Michael E. Todhunter[1†#], Sheikh Jubair[2†#], Ruchika Verma[2], Rikard Saqe[3], Kevin Shen[3], Breanna Duffy[4*†]

[1]Todhunter Scientifics, Minneapolis, MN, USA

[2]Alberta Machine Intelligence Institute, Edmonton, AB, Canada

[3]University of Waterloo, Waterloo, ON, Canada

[4]New Harvest, Sacramento, CA, USA

*** Correspondence:**
Breanna Duffy
breanna@new-harvest.org

[†] These authors contributed equally to this work

[#]These authors share first authorship

Keywords: Machine Learning, Artificial Intelligence, Culture Meat, Cell Culture, Culture Media Design, Microscopy, Bioprocessing, Food Science

Abstract

Cultured meat has the potential to provide a complementary meat industry with reduced environmental, ethical, and health impacts. However, major technological challenges remain which require time- and resource-intensive research and development efforts. Machine learning has the potential to accelerate cultured meat technology by streamlining experiments, predicting optimal results, and reducing experimentation time and resources. However, the use of machine learning in cultured meat is in its infancy. This review covers the work available to date on the use of machine learning in cultured meat and explores future possibilities. We address four major areas of cultured meat research and development: establishing cell lines, cell culture media design, microscopy and image analysis, and bioprocessing and food processing optimization. This review aims to provide the foundation necessary for both cultured meat and machine learning scientists to identify research opportunities at the intersection between cultured meat and machine learning.

1 Introduction

Food production generates about a quarter of global greenhouse gas emissions and causes other negative impacts on the environment and human health (1,2). Animal products, including meat, seafood, eggs, and dairy contribute more than 56% of food's emissions, despite providing only 37% of protein and 18% of calorie intake (3). Meat production is a large sector, producing 328 Mt in 2020 and expected to expand to 374 Mt by 2023, based on estimates from the Organisation for Economic Co-operation and Development (OECD) and the Food and Agriculture Organization (FAO) of the United Nations (4). In light of projected growth in the global population and income, these estimates projected that meat consumption will increase by 14%. To meet global meat demand and limit global warming to 1.5°C there is a need for major changes in the production of meat (5,6). Cultured meat

(CM), known by many names including “cell-based” or “cultivated” meat, is an emerging technology that uses tissue engineering and biomanufacturing techniques to produce animal meat through cell culture rather than animal husbandry. Proponents of the technology herald its potential to provide an option for producing animal agriculture products with reduced environmental, ethical, and health impacts (7,8). However, major technological challenges remain in bringing CM products to market and achieving their proposed benefits (9). Many challenges stem from the fact that the technologies for mammalian tissue culture come primarily from the medical field, where the scale is much lower and the market has weaker incentives to reduce the costs of production. However, if these technologies are to be applied to food, the challenges of scale and cost must be addressed.

Some specific improvements to CM can be, and have been, made using traditional experimental approaches. However, tackling some of the more complex research questions requires more advanced approaches and experimental conditions. In recent years, an increasing number of groups have been using methods enhanced by artificial intelligence (AI), and in particular its subset machine learning (ML), for such tasks. ML can streamline experiments, predict optimal results, and reduce experimentation time and resources. There are many opportunities for ML to accelerate research development and reduce costs in CM. Some companies have indicated the use of ML in CM product development or CM-associated services (10–17), but very little of this progress in applying ML to CM has been shown or validated in the public domain. Academic and government-supported research in this space is emerging, including at the University of California at Davis, Virginia Tech, Tufts University, and The CentRe of Innovation for Sustainable banking and Production of cultivated Meats (CRISP Meats). However, few publications on the topic exist to date (18–21). An increase in open public research on the use of ML to optimize and scale CM production would greatly accelerate the application of ML to the CM field.

In this review, we aim to provide a foundation necessary for researchers, from the CM or ML fields, to identify research opportunities at the intersection between CM and ML. We first provide a brief overview of both fields. Subsequent sections delve into both existing and potential ML applications for optimizing cell lines, formulating culture media, aiding cell culture microscopy and image analysis, and optimizing bioreactor and food processing parameters. Note that we have focused on CM challenges to which ML can be applied, and this should not be seen as a comprehensive review of all challenges in the CM or ML fields. As applications of ML in CM are limited, we discuss how ML methods that have been applied in other areas of bioinformatics can be adopted to solve tasks in CM.

2 Background on the Fields of Cultured Meat and Machine Learning

2.1 Cultured Meat

CM aims to replicate the taste and texture of animal tissue within a manufacturing system using animal cells (Figure 1). First, cells from a species of interest are selected or engineered for desirable growth and differentiation traits in vitro. The selected cells are grown in a suitable medium, providing the nutrients and signaling cues that would normally be provided in the body. At early stages, such as cell selection, the cell culture may start at a small scale in plastic dishes or flasks. Cell culture is eventually scaled up to industrial-scale bioreactors, devices capable of controlling environmental temperature, pH, dissolved oxygen, and nutrient exchange at large volumes (22). Once enough cells are grown, they are differentiated into mature cell types. At this stage, the cell culture may be formed into a tissue (i.e. structured product, such as a steak) or a cell slurry, which is later processed into a meat product (i.e. unstructured product, such as ground beef).

This process leans heavily on medical tissue engineering, an area of research that has been studied for nearly 40 years and has been commercialized at a small scale for simple grafts of cell-laden scaffolds for skin and cartilage. However, complete tissues, such as skin with hair follicles and sebaceous glands or functional muscle, still face technological barriers (23). While tissue function is not critical for CM, using this technology for the production of food comes with unique constraints, especially the need for larger scale, lower costs, and materials that are both edible and palatable.

Commercial and academic interest in CM has grown rapidly in the last decade. The number of companies working on CM grew from 1 to over 170 from 2011 to 2023, with over 3 billion dollars of investment (24). Similarly, academic interest in the field has grown dramatically, with 350+ papers published on CM in the last 2 years, more than all other years prior (with tracking) combined (24). In parallel, cost estimates for CM have come down significantly in the last decade: from the first CM demonstration in 2013 of a 140-gram burger at approximately €250,000 (25), to recent claims of costs as low as $7.70/lb from industry developers (26). However, these industry cost claims have yet to be proven publicly. Furthermore, CM production has yet to be shown on a scale close to that needed to offset even a fraction of current meat consumption (27). Current production capacity is not known, but estimates range from 1-10 kg/y, compared to the 3.2x1011 kg/y produced by conventional meat (28). Highly optimized industrial mammalian cell lines, such as Chinese hamster ovary (CHO), are still produced at a much lower scale and higher cost than needed for food production (28). Given that consumers are unlikely to want to consume hamster ovary cells, significant technological challenges must be overcome for meat and seafood-relevant cells to be produced at a scale and cost needed to replace a meaningful portion of the conventional meat market.

2.2 Machine Learning

An ML workflow involves a series of steps, as illustrated in Figure 2, which starts with preparation of the dataset. A dataset typically consists of datapoints, each one an observation with features that describe the datapoint. The type of data varies widely: numerical, time series, text, images, audio, video, sequential, graph, or any combination of these. Since the data must be in a numerical form for the subsequent steps, a data transformation step converts the data to numerical values. The next step is the preprocessing of raw data, where procedures such as imputing missing values and reducing data dimensionality are undertaken.

Subsequently, an integral component of the ML methodology is dividing data into distinct subsets: training, validation, and testing. Some common approaches, such as K-fold cross-validation, leave-one-out, and holdout validation, can typically be used for most problems (30). Prior to model training, it may be necessary to employ feature selection or extraction techniques on the training set to identify the most informative variables in the data, and these features must be used in both the test and validation sets to maintain the integrity of the evaluation process.

The culmination of this preparatory work is a dataset suitable for training an ML model. Typically, experiments are conducted with different ML algorithms/architectures (such as random forests, k-means, and deep neural networks) to settle on the most performant overall model. However, sometimes the choice of ML model depends on the type of data. For example, convolutional neural networks (CNNs) are preferred for image data since they can extract spatial features from the image (31). On the other hand, recurrent neural networks (RNNs) are typically used for sequential data since they can remember information in a long sequence through their gate mechanism (32).

The training set is instrumental in building the model since the model adjusts its parameters to learn the distribution of the training set. The validation set is required for fine-tuning the model's

hyperparameters and ensuring that the model generalizes well beyond the training data, thereby avoiding overfitting. Finally, the test set provides a measure of the model's predictive accuracy and overall performance in real-world scenarios. Some of the typical performance metrics for classification tasks are accuracy, precision, recall, and F1-Score. Typical performance metrics for regression are R^2 score, mean absolute error, and mean square error (33). However, lots of other performance metrics are used depending on the tasks and types of data. The purpose of performance metrics is to compare the performance between different models and also to understand how a model is performing overall for the specific task.

Types of Machine Learning:

ML methodologies can be broadly classified into three categories: supervised, unsupervised, and reinforcement learning (RL). In supervised learning, the model is trained on a labeled dataset, where each example is paired with an outcome or label that aligns with the objective of the specific task at hand. Consider the goal of predicting gene expression levels from DNA sequences; here, the data point would be the DNA sequence, while the associated expression level serves as the label. The model hones its predictive capability by learning the relationship between input features and their corresponding labels. When it comes to evaluation, the trained model is tasked with predicting labels for new, unseen data points. Key traditional supervised learning models include logistic/linear Regression, k-nearest neighbours, and support vector machines. The supervised ML approach has made significant contributions across various domains in bioinformatics, enabling advancements in DNA segmentation, gene expression prediction, and protein structure prediction (34).

Unsupervised learning, in contrast, does not rely on labeled data. It is particularly useful when the goal is to unearth underlying patterns or structures within the data, independent of predefined outcomes. This makes unsupervised learning a potent tool for exploratory analysis, especially in scenarios where labeled data is scarce or when the structure of the data is not fully understood. Some of the key unsupervised learning approaches are k-means, hierarchical clustering, and density-based spatial clustering of applications with noise. Unsupervised learning has been successfully applied in grouping functionality-related genes, microarray analysis, and biological image segmentation (34,35).

RL is a dynamic and adaptive approach well-suited for situations where a machine is required to make a series of decisions to achieve a desired goal or perform an optimization task, offering a framework for learning through interaction (36). Within this paradigm, an agent—often an advanced ML model in deep RL —engages in a sequential decision-making process, each time interacting with a complex environment represented by a set of variables that define the current state of the environment. The agent executes actions, transitioning between states, and ultimately may reach a terminal state, signaling the conclusion of the decision sequence. The RL framework incorporates a system of rewards and penalties, with the agent receiving feedback in the form of rewards for beneficial actions or penalties for undesirable outcomes. The objective for the agent is to devise a strategy that maximizes cumulative rewards, thus steering towards the most optimal actions to attain its final goal. RL has recently been applied to bioinformatics (37–39) and gained lots of attention because of its success in areas, such as sequence alignment (40) and protein loop sampling (41).

Neural Networks:

Recent advances in applying ML to biology are based on neural networks, a subset of ML methods that can be employed in all three ML categories: supervised, unsupervised, and RL (42). ML models

with neural networks are often referred to as deep neural networks when there are multiple layers of neural networks in the architecture of the model, a method more broadly known as deep learning. These neural network layers typically attempt to mimic the activity of brain neurons, where each neuron employs a mathematical function that alters the data it receives from the previous layer (43,44). At first, the data is fed into an input layer, which then connects to hidden layer(s) (used for computing), and finally an output layer, designed to deliver the final prediction. The model learns by optimizing the function parameters of each node. Some popular neural network architectures are feed forward networks, CNNs, RNNs, and transformers (43–45). Deep learning models typically capture complex biological processes and incorporate heterogeneous data in the model through its different layers, which may be a necessity for many optimization and prediction tasks in CM (46). It can be applied to both supervised and unsupervised scenarios. In RL, deep learning models can be employed as agents as well (47).

Generative AI:

Generative AI is another subfield of unsupervised learning which typically aims to generate new data or samples based on the patterns learned from the training data. Variational Autoencoders (VAEs) (48–50), which employ deep learning, are the first generation of generative AI models. VAEs typically employ two deep learning models: i) an encoder that encodes the input into a latent space and ii) a decoder that predicts the input by matching the distribution of the latent space by applying variational inference techniques. Together, the encoder and decoder minimize the reconstruction loss of the input. Figure 3 shows the general architecture of VAE.

The second generation of generative models is deep adversarial networks (51,52). These architectures also have two networks: a generator and a discriminator. Initially, generators generate new data randomly and discriminators classify whether the generated sample is original or generated. Iteratively, both the generator and discriminator learn how to generate a better sample and distinguish between original or generated sample. Eventually, the generator learns to generate realistic data samples that are difficult for the discriminator to distinguish from real data.

Recent advances in generative AI are mostly based on transformers (45), which have revolutionized various fields including natural language processing, computer vision, and bioinformatics (53,54). Transformers have demonstrated exceptional capabilities in capturing long-range dependencies and modeling complex sequential data. In bioinformatics, methodologies heavily draw inspiration from NLP techniques due to the inherent similarities between biological sequences and natural language texts. By employing transformers, researchers are able to effectively model biological sequences such as DNA, RNA, and protein sequences, leading to significant advancements in tasks such as sequence generation, structure prediction, and drug discovery (38,53–56).

Graph Neural Networks:

Graph neural networks (GNNs) are another area of ML that works with graph structured data and biological networks, such as via protein interaction networks, gene coexpression networks, and metabolic networks (57,58). A biological network or biological graph typically is arranged in terms of nodes and edges, where the nodes are biological entities (i.e. genes, proteins, metabolites, etc) and the edges indicate how these entities relate to one another. GNNs are based on the principle of message passing, where each node in the graph aggregates all embeddings (messages) of its neighbour nodes and updates the weights of pooled messages through a neural network. There are three main tasks in graph structured data. The first one is link prediction between two biological

entities. Examples of link prediction are predicting the interactions between proteins or predicting interactions between genes/regulatory elements (59,60). The second task is node functionality prediction. For example, predicting an unknown function of a protein based on the physical interactions between proteins in the protein interaction network (61). The third application of ML in network analysis is to classify sub-network functionality. Figure 4 shows different ML tasks in a graph or network structured data. Muzio et al. classified functions of subnetwork based on a molecule's toxicity or solubility (61). In addition, ML is also applied to obtain subnetwork embedding where the subnetwork is represented as a vector preserving the important information within the subnetwork in a numeric form. This embedding can facilitate further analysis of the subnetwork and is used for various downstream tasks (62).

3 Cell Lines

Meat consists of various cell types, predominantly muscle cells (approximately 90%), with fat and connective tissue cells accounting for the remaining 10% (63). Additionally, there are some vascular, neural, and tissue-resident immune cells present in small amounts (63–65). To provide the taste, texture, and aroma expected in meat, most CM development has focused on muscle, fat, and connective tissue production. Cells will also likely contribute to the nutritional properties of CM products, and cell optimization may be used to tailor the nutritional profile of CM (66,67).

The cell types used to produce CM range from lineage-committed progenitor cells (e.g. muscle satellite cells, myoblasts, or preadipocytes) to stem cells able to differentiate into a broader set of cells (e.g. mesenchymal stem cells, embryonic stem cells, or induced pluripotent stem cells) (64). Developers may use primary cells, meaning cells that are isolated directly from an animal. However, cell lines, which are established cultures of cells that have been selected and optimized, are ideal because they are more consistent, characterized, and reduce the use of animals in the supply chain. Furthermore, immortalized cell lines or pluripotent stem cells are of particular interest due to their ability to escape the typical limits on population doublings seen in most primary cells (i.e. Hayflick limit) (68,69). A detailed review of cells used in the production of CM can be found elsewhere (65).

Few CM-relevant cell lines are currently well characterized and commercially available, with most coming from model organisms used in biomedical research, such as mouse, rat, or zebrafish, and cell lines for many agricultural species still need to be developed (70). Cell lines for marine species, especially invertebrates such as mollusks or crustaceans, are especially underdeveloped, with only a few reported cell lines and not all are food-relevant species (71–80). For many species, there is a lack of basic knowledge of their physiology and the biochemistry required for in vitro culture or immortalization (70,81). The lack of knowledge of molecular and genetic markers as well as few species-specific antibodies available makes identifying and curating cell lines difficult for under-studied species. ML can help biologists analyze complex cellular data to assist with identifying ideal cell line populations and optimizing existing cell lines through gene perturbations.

3.1 Network analysis is a tool to model biological interactions

Optimization of cell lines is often challenging because it requires understanding the “state” of a cell or cell population (i.e. what the gene and protein networks are doing) and selecting or engineering for desired cell states. Measuring and predicting these states involves interpreting complex interactions between genes and proteins, identifying those that are important for specific qualities, and predicting how perturbations will affect the whole network. ML can be used to model this complex network through an area of study known as network analysis. Network analysis is used throughout sections 3 and 4, so a general overview is first presented here.

Most of the earlier works in network analysis, such as DeepWalk (82–84), LINE (85), Node2Vec (86,87) are inspired by NLP or community detection methods of social networks. These methods primarily use multiple random walks from a specific node to obtain a subnetwork. The objective is to obtain a vector embedding of the specific node by following a skip-gram strategy employed in NLP (88). The obtained embedding maximizes the probability of predicting the neighboring nodes obtained by random walks. This embedding can be used for most of the network analysis tasks described previously. However, these approaches often do not preserve the graph structure information which may lead to information loss.

In recent years, GNNs have been employed in graph-structured data and biological networks. While the objective remains the same for network analysis with GNN, the primary difference between the NLP approach and the GNN is that since GNNs are built for working with graph structured data, each node in the graph aggregates information from its neighbours while preserving the connectivity information in the GNN. GNNs have been implemented to predict protein interactions (89,90), molecular interactions (91,92), metabolite-disease associations (93) and obtain subnetwork embeddings (94) that can be used to identify the functions of a biological subnetwork. Generative deep learning strategies, such as VAEs and deep adversarial network-based models, are also trained on biological networks by employing GNNs (95,96). The most famous example to date is AlphaFold2, a deep learning GNN model that was employed on amino acid sequences to predict more than 200 million protein structures (38). This is a significant advance in the field of structural biology which can be used in protein design, drug target prediction, cell type identification, and antibody development. AlphaFold2 can play a big role in designing new proteins and understanding the physical relationships between proteins (97).

Network analysis can employ multiple types of biological data within a multi-omics setting. Since each omics technique typically captures a specific biological process, integrating multi-omics data can provide a holistic overview of the biological process. In a multi-omics ML approach, different ML models are employed on different types of data (such as genomics, transcriptomics, proteomics, and metabolomics) to obtain numerical representations. These numerical representations are then combined together to obtain a more informative representation used to predict final outputs. For example, MOGONET employs a GNN on mRNA expression, DNA methylation, and miRNA expression data to predict disease information (98). Similar datasets can also be used for CM to predict cell line features. More applications of multi-omics ML methods are discussed in section 3.2.

Furthermore, network analysis can help determine aroma and flavor (99). Since meat aroma and flavor are largely controlled by metabolic pathways (100), network analysis can be applied to biosynthetic pathways to enhance or add flavors to CM products. Network optimization has been used to design yeast that overexpress licorice glycoside (101), and although licorice is a flavor that few would want in a steak, one could imagine more savory corollaries. In other yeast experiments, transcriptomic network analysis has been used to improve acid resistance, and acid resistance is just as important to mammalian culture as to microbial culture (102).

3.2 Machine Learning Can Help to Analyze Complex Omics Data to Identify and Characterize New Cell Lines

"Omics" approaches, such as genomics, transcriptomics, proteomics, and metabolomics, are powerful tools for identifying and characterizing cell lines. In particular, RNA sequencing (RNA-seq) technologies are commonly used to quantify cellular gene expression to validate and optimize cell lines. However, analyzing RNA-seq or other -omics data across many candidate cells is a complex

and daunting analytical task. ML can help these analyses in multiple ways, including by grouping functionality-related cells using an unsupervised approach (94), profiling gene expression using a supervised approach (103), and identifying different tissue types using unsupervised (104,105) and semi-supervised approaches (106).

Using gene expression data to cluster cells based on cell type or behavior can help explain heterogeneity among cell populations and discover subpopulations with beneficial characteristics. When establishing cells for CM production, scientists may want to isolate only certain cell types with optimal attributes or remove undesirable cell types. For example, using single cell RNA-seq (scRNA-seq) Messemer et al. found that an isolation from cattle muscle contained 11 distinct cell types (107). This work led to a better understanding of the cells derived from a primary isolation, as well as cell surface markers suitable for the identification and separation of populations by flow cytometry. Additionally, in a recent preprint, Melzener et al. used RNA-seq to study subpopulations during muscle differentiation, understanding cell fates with an aim to improve the efficiency of cell differentiation and maturation (108). ML is also increasingly being explored for the modeling of cell trajectories via scRNA-seq (109).

Unsupervised ML can help to map cellular heterogeneity by grouping functionality-related cells, identifying cell sub-populations, and performing dimensionality reduction (104,105,110). Typically, the input to the unsupervised ML model is gene expression data obtained from RNA-seq. In conventional ML frameworks, which are not based on neural networks, the outcome is generally an assigned cluster number for each cell or gene. For an in-depth exploration of how traditional ML techniques are applied in this context, we direct readers to the comprehensive review by Petegrosso et al. (111). In contrast, unsupervised deep learning methods predominantly leverage autoencoders, which compress high-dimensional cellular data into a more manageable lower-dimensional space while retaining essential information (104,105,110). This lower dimensional representation of the encoder can be used to obtain clusters of cells (104,105,112–114) and can be further fine tuned for gene expression profiling (42,115).

Recently, another family of autoencoders that uses GNNs has been employed to obtain the lower dimensional representation of transcriptomics data. These GNN-based autoencoders use the knowledge of biological networks, such as gene-gene relationships, cell interaction networks, protein interaction networks and biological pathways, along with gene expression data to obtain a more robust and informative representation of cells and genes (94,110,116–120). The GNN autoencoders are unique compared to traditional autoencoders as they encode information on biological interaction between entities along with structural information and biological properties.

Semi-supervised approaches have been employed in many tasks, such as learning responses due to gene perturbation, different functional score prediction due to gene knockouts, identifying different tissue types, and transcriptome analysis

(106,121–126). These models are adept at leveraging both labeled and unlabeled data, typically employing the unlabeled data to train an autoencoder and then using the labeled data to fine-tune the autoencoder toward the target outcomes (106,123,127). Fine-tuning is the process of adjusting pretrained model parameters for a specific task by utilizing a small labeled dataset. The semi-supervised approach is notably different from its unsupervised counterpart as it provides a degree of guided learning, which is crucial when the available labeled data is sparse but critical for the identification of specific labels. For instance, while the categorization of cells based on functional similarities may fall under unsupervised learning—grouping cells by inherent characteristics—the

task of pinpointing cell doublets could benefit from a semi-supervised model that utilizes a limited set of known doublet samples to enhance its predictive accuracy.

In addition to transcriptomic data, further single-cell multimodal sequencing technologies have been developed that provide cell-specific information, such as chromatin accessibility (scATAC-seq) (128) and surface proteins (CITE-seq) (129). These data types provide complimentary insight into scRNA-seq, such as improving accuracy in modeling gene regulatory networks in the case of scATAC-seq (130,131). This understanding can aid in CM-related tasks, such as identifying transcriptomic/epigenetic markers predictive of high proliferation and differentiation potential given previously observed variability in primary cell culture performance (132–134).

In an ML context, autoencoders and GNN-based deep learning models are mostly applied to this multimodal data (135). Some of these autoencoders employ individual encoders and decoders for each modality. We refer the readers to a review of multimodal single-cell models, and a review on general best practices for single cell analysis, to learn more about the topic (135,136).

3.3 Antibody Design for Characterization and Isolation of Novel Species can be Aided by Machine Learning

In cell line development it is crucial to both characterize cells, commonly via omics data (as discussed above) or visual identification (discussed in section 5), and isolate the cells of interest, typically using flow cytometry cell sorting. Visual identification and flow cytometry both require the use of antibodies with specific binding affinity to known cellular markers. Markers of muscle cell differentiation are well understood for most mammalian species (137). However, in under-explored species, such as fish or aquatic invertebrates, these proteins are not always shared with mammalian cells or, to the degree they are, have low sequence conservation (138). For example, recent work with an Atlantic mackerel (Scomber scombrus) skeletal muscle cell line found that antibodies for the muscle satellite cell marker paired-box protein 7 (PAX7) was successful, while early myogenic marker myoblast determination protein 1 (MYOD) was not (80). Surface markers, which are most commonly used during flow cytometry cell sorting, are particularly understudied in CM-relevant cell types, with most existing research focused on immunology (138,139).

Attempts at validating commonly used mammalian antibodies in fish have led to issues with cross-reactivity and specificity, suggesting that establishing cell lines for CM research will require the repurposing of existing antibodies, if not the development of fully novel antibodies (140). The use of ML on omics data has been shown to aid antibody development, such as via improving species cross-reactivity, antibody co-optimization, and binding affinity (141–143). The autoencoder model totalVI, trained on joint scRNA-seq and CITE-seq data, is able to provide insight into many key variables that would aid antibody development: the identification of novel differentially expressed features to target in a given cell subpopulation, the improved prediction of false positive/negative surface proteins, the reduction of technical bias common in antibody-based measurements such as "background", and the improvement of experimental design by helping determine optimal antibody titrations/sequencing depths for balancing cost and signal-to-noise ratio (144). For a full review of computational methods relevant to antibody development, readers are pointed to (145).

3.4 Machine Learning Can Map and Enhance Genetic Traits to Optimize Cell Lines

Gene editing can be used to enhance or alter cellular traits to generate cell lines optimized for CM production, such as accelerating growth, extending growth (i.e. immortalization), reducing input costs, or tailoring the flavor and nutrition. For example, manipulating cellular metabolism has a large

potential for increasing the efficiency (and therefore reducing costs) of CM products (9,146). As another example, Stout et al. engineered cells to overexpress FGF2, eliminating the need for FGF2 supplementation in media through autocrine signaling (147). Because many of the species used in CM are under-studied, their genome regulatory networks are not yet well understood, slowing efforts for gene editing. ML offers an opportunity to accelerate gene editing technology.

Genes typically have many regulatory regions, such as upstream and downstream regions, untranslated regions, promoters, and enhancers. These regulatory regions determine where, when, and how much a gene is expressed. Thus, to apply gene editing technology, identifying these regulatory regions is an essential task where ML models have been applied successfully (53,148,149). The most relevant modern ML techniques demonstrated to be useful for these tasks are generative adversarial networks (150,151) and convolutional neural networks (152). ATAC-sequencing is a very useful data modality for informing on this set of tasks (153).

More recently, researchers have adopted transformers and architectures similar to large language models, such as BERT, to segment regulatory regions and predict expression levels (53). These models are typically trained on a large number of DNA sequences or similar kinds of data using a supervised fashion by employing a masking strategy to generate a labeled dataset when labels are not available (53,148). This entails subdividing a DNA sequence into numerous smaller fragments, typically employing a k-mers based approach, where a k-mer is a subsequence of length k within the DNA. Selected k-mers are then masked within the sequence. The objective of the transformer-based model is to predict those masked parts of the sequence while imitating the interactions among different fragments (53,154). By predicting the masked fragment, the model learns the underlying structure of the DNA sequences. The resulting output is a set of numerical vectors, often referred to as embeddings, representing each k-mer, encapsulating the sequence information. These vector representations can be further used to fine-tune the model for various downstream tasks, such as segment identification, sequence alignment, and gene expression prediction where limited labeled data is available (53,148,149).

When labeled data is available, convolutional neural networks and transformers can be combined to predict outcomes. This approach employs convolutional neural networks on the small DNA fragments to obtain an initial representation. Following this, the processed fragments are inputted into transformer layers. The transformer architecture finds how each fragment interacts with other fragments and summarizes these interactions in a vector that represents the fragment. The vector representations of all fragments are finally used to predict different genomic tracks (155). Since the vector representations of the DNA fragments derived from the final transformer layer encapsulate information that is both comprehensive and adaptable across various genomic tracks, these vectors can be effectively utilized for predicting a range of genomic tasks beyond those initially targeted.

Since these approaches can be used for gene expression prediction, they are particularly useful for predicting the expression of an edited DNA sequence. Gene editing involves adding, removing, or substituting a segment of DNA to alter the expression of a specific gene. Identification of replacement DNA segments is a complex task that can be addressed by employing an ML algorithm that can estimate the gene expression. Alternatively, using RL, the discussed ML approaches can be trained to generate gene-edited DNA sequences that maximize the expression by following the same training architecture of Large Language Models, such as ChatGPT (156). Initially, one of these models can be fine-tuned to generate a gene-edited DNA sequence from an input DNA sequence (supervised fine-tuned model). Then, the supervised fine-tuned model can be optimized for predicting gene expression, which will work as the reward function of the RL agent (reward model).

Finally, the supervised fine-tuned model can be further optimized by employing RL optimization techniques, such as Proximal Policy Optimization (157), where outputs of the reward model are used as rewards for generated sequences.

4 Media Design

Cell culture requires media to keep cells alive, promote growth, and direct differentiation. Culture media contains the nutrients and signals that cells receive in the body and is typically composed of carbohydrates, amino acids, vitamins, minerals, buffers, proteins, peptides, fatty acids, lipids, and growth factors. In addition, media components may also contribute to the flavor of the final product (158,159). A detailed discussion of how cells utilize these components has been covered previously (158).

Most existing cell culture media formulations are expensive, not designed for the species of interest for CM production, and include animal-derived ingredients (158). Media design for CM requires optimizations to address these limitations, which are often complex and resource-intensive research efforts. ML is well suited to accelerate research in this space.

4.1 Culture Media Design Can be Treated as a Hyperparameter Optimization Problem

The objective of media design is to identify conditions that are optimized for various parameters, such as price, growth rate, stability, flavor, environmental impacts, and lot-to-lot reproducibility, sometimes optimizing for several of these objectives simultaneously (158,160). In addition, media may be optimized for specific cell responses, such as differentiation or spontaneous immortalization (161,162). In certain cases, media may need to be developed for uncharted parameter spaces; for instance, cells derived from lamb, duck, and deer that were not previously grown for biomedical purposes.

Most culture media contain dozens of ingredients - the classic media Ham's F-12 contains 47 ingredients (163). Certain ingredients, such as fetal bovine serum or B27 (164), are themselves complex mixtures of additional ingredients. In the absence of ML, single specific improvements of large effect can still be sometimes made to culture media, such as a group that swapped animal albumin for plant albumin (165). Big gains in the stability of stem cell media have been made with recombinant peptides (166). Other groups have optimized media, such as bovine myoblast media, using straightforward factorial approaches, and design of experiments for these approaches are quite mature (167,168). However, a medium with 49 ingredients, each with five relevant concentrations, has a design space of over 1034 potential recipes, exceeding the capacity of traditional factorial methods (i.e., Taguchi methods (169)) or high-throughput screening.

Some attempts have been made to apply ML to the culture media design problem as a means of exploring the experimental search space more effectively. A recent preprint (170) used gradient-boosted trees to identify culture media that improve the growth characteristics of suspension-phase HeLa cells. Another recent paper trained a neural network to model the results of fractional factorial culture media experiments (21). Response surface methodology, a classic design-of-experiments method, has been used in conjunction with a genetic algorithm for optimizing media properties such as cost, growth rate, and global warming potential of both zebrafish fibroblast (21) and mouse myocyte (20) culture media. However, these methods, on their own, aren't optimal for dealing with the complexity of full culture media recipe design.

Designing culture media is fundamentally a high-dimensional search problem, and Bayesian optimization can effectively navigate problems of this nature, which has the advantages of requiring only a small number of experiments and dealing with uncertainty in a robust manner. Bayesian optimization has been used to optimize myocyte culture media (18), spirulina culture media (171), and keratinocyte differentiation media (172). Google Vizier, which is Google's hyperparameter tuning service, is a key enabler of Bayesian optimization, as it provides biologists access to this algorithm via API, bypassing the need to code from scratch (173). Additionally, the BoTorch library for Python offers a powerful free toolkit to implement Bayesian optimization (174).

Even given an optimal algorithm for optimization, obtaining and analyzing relevant data from cultured cells remains a challenge in media design. One potential solution is through the use of high-content imaging, which can be enhanced through the implementation of ML methods, as discussed below in Section 5. Another approach involves the application of gene network analysis, which may be the most effective method for gaining detailed insights into the cellular response to a given culture medium.

Bayesian optimization is the leading hyperparameter exploration technique, but it stands among others. Genetic algorithms, which cluster parameters in a way that mimics biological chromosomes to enable recombination (175), have been used to optimize cyanobacteria media (176) and myocyte media (19). Simulated annealing, which is based on a metallurgical concept (177), has been used in conventional animal agriculture to optimize poultry feed (178) and to produce a population balance model for CHO cells (178). However, Bayesian optimization appears to be the most widely used among these techniques, probably because it can work on non-differentiable objective functions without the difficulty of partitioning parameter space into somewhat arbitrary and finicky "chromosomes" (179).

4.2 Machine Learning Can Interpret How Cells Respond to Media Conditions

The heart of media design is attempting to alter the properties of cells by altering their environment. Assessing the properties of cells can be challenging - although RNA-seq can provide insight on the inner state of cells by quantifying expressed genes (161). However, that information is often not readily interpretable because it is high-dimensional and the effect of any single gene is often esoteric or context-dependent. ML-guided network analysis is a solution - it can assess changes to the inner state of cells when performing media design or other optimizations, such as culture vessel geometry or atmosphere composition. Both gene-based network analysis and metabolite-based network analysis can be used for this application.

Gene network analysis has not yet been directly used in CM media research, but it has been used in parallel applications. Network analysis has been used to improve mammalian culture, CHO cells in particular, to identify feed supplements (180). Network analysis has also been used to find gene networks important for feed efficiency or milk production in cattle, with the networks formed from phenotype-genotype analysis (181,182).

To get more details about a gene network, gene network inference (GNI) methods can deduce the regulatory interactions between genes. The potential value of GNI to CM media design is that GNI can aid and simplify the interpretation of an RNA-seq data matrix; instead of looking at the differential expression of tens of thousands of genome features, the analyst can look at the differential activity of dozens of network modules. Among recent GNI methods are dynGENIE3 (random forests) (183), Scribe (184), and TENET (185), all of which are benchmarks for the most recent papers (186). GNI has been used to optimize output in traditional agriculture, both soybean

cultivation (187) and maize cultivation (188), and GNI could be used for analogous output maximization problems in CM.

In biotechnology, metabolic network analysis has, historically, most commonly been done with linear programming. Flux balance analysis, a type of linear programming that is especially prominent, attempts to predict the inputs and outputs of metabolites in cells (189). Linear programming models can incorporate the relationships between different biological networks, assuming the relationships between networks are both known and linear (190,191). Flux balance analysis has been used to model CHO metabolic states (192), to predict the effect of sparging on cultures (193), to evaluate hypotheses regarding metabolic changes (194), and to develop genome-scale metabolic models (195), all of which are relevant aspects of CM. Linear programming has been combined with other ML methods in CHO cells (180) for optimizing cell growth.

4.3 Optimizing and Engineering Proteins Is Much Easier With Machine Learning

Another opportunity for optimization is in the ingredients themselves. Growth factors are currently estimated to contribute over 95% of the total production cost (196). Altering proteins for use in cell culture media is a promising means to optimize the stability, price, or other desirable properties of these ingredients.

Recombinant production, a commonly used animal-free strategy to produce proteins, is a major contributor to media costs (197). Most proteins present in culture medium are highly specialized eukaryotic proteins, requiring post-translational modifications such as phosphorylation or glycosylation that require mammalian cell culture systems, rather than the far cheaper and more scalable bacterial production systems. Recent work developed an E. coli expression system to reduce the price of several relevant growth factors (197). Similar efforts are necessary to reduce the cost of other media ingredients.

Probably the biggest challenges when importing transgenic proteins are differential folding between species (i.e. different chaperones, different organism temperatures) and differential post-translational modification. There have been many efforts to use ML to predict post-translational glycosylation, with earlier efforts using random forests (198) and more recent efforts using multi-layer perceptrons (199) or ensemble models (200). As for differential folding, transformer-based models have been used to predict thermal protein stability (201). Combining transformers and GNNs has been used to predict protein subcellular localization (202). More recently, AlphaFold has been used to predict protein stability in a general sense (203), and ColabFold was developed to improve upon it (204).

Protein ingredients can also be engineered to optimize their stability in culture, reducing the total amount of protein needed for CM production (205). For example, FGF2 thermal stability was increased using point mutations in the protein (206). Similarly, Long R3 IGF-1 is a modified recombinant form of IGF that prevents inactivation by IGF binding factors, leading to x200 potency and x3 stability compared to standard insulin (207). Similar strategies could be applied to other expensive media ingredients to reduce the necessary concentrations or additions. RFdiffusion (based on a generative diffusion model) (208) and ProteinMPNN (209) are tools designed to solve this problem. ML is also used in directed evolution experiments that iterate upon a given protein's design (142,210).

For some proteins used in high concentrations, such as albumin, transferrin, and insulin, replacement with lower-cost alternatives, such as plant (146,211) or microbial (8,212) hydrolysates may be beneficial. For example, a protein similar to bovine insulin was found in cowpea (Vigna unguiculata)

and could be isolated for use as an insulin replacement (213,214). Screening plants or microbes for sequence homology to proteins of interest could accelerate this work. The traditional, non-ML tool for homology searches is protein BLAST (215), but BLAST is underpowered because proteins can have similar structures without having similar sequences. Services like Dali provide protein structure alignment but are constrained by the paucity of known protein structures and the computational complexity of three-dimensional comparisons (216). ML has changed how to approach this problem with FoldSeek (217), which uses autoencoder-based compression to simplify three-dimensional protein structural comparisons, using AlphaFold to infer protein structures when no empirical data is available (38).

5 Cell Culture Microscopy and Image Analysis

Microscopy is one of the foundational techniques of cell culture, providing information such as (i) the health of cells (e.g., whether they are mitotic, senescent, or apoptotic); (ii) the behavior of cells (e.g., whether they are invasive, contractile, or secretory); and (iii) the lineage of cells (e.g., whether they are stem cells, progenitor cells, or terminally differentiated cells). Many types of cell analysis rely on microscopy. For example, fusion index, the percentage of nuclei inside myotubes using immunostaining, is the predominant test used to quantify myogenic differentiation (64). Alternatively, inexpensive colorimetric staining has been shown as an alternative measure of myotube differentiation (218). Additionally, brightfield imaging, without the use of dyes, has been used to measure cell contractility, which is a measure of muscle cell maturity (219,220).

The relevance of microscopy during large-scale production will be especially dependent on its ability to serve as a low-cost, high-throughput tool. However, microscopy has been historically constrained by the complexity of its analysis. Microscopy analysis is routinely done manually by researchers with well-trained eyes, and doing it automatically requires systems that can incorporate many nuanced features of the image data. Unfortunately, appropriate systems tend to be nascent, poor, or non-existent in biological research. Furthermore, the use of dyes to improve image quality is undesirable due to the cost and time constraints of CM production. For CM production, cell segmentation and classification are fundamental and indispensable due to their multifaceted contributions - they are relevant to quality control, monitoring of cell culture health, and optimizing production. Utilizing ML approaches for automated cell segmentation and classification can lead to a reduction in the time, expenses, and errors involved in preparing the setup for manually analyzing the image data.

5.1 Automatic Image Analysis for Cell Segmentation using Machine Learning

Cell segmentation is the process of identifying and separating individual cells within an image. This is done in digital microscopy and histopathological imaging to study cell structure and function. The goal of cell segmentation is to pick out cells in an image, and segmentation is necessary to measure cell size, shape, and number. Segmentation also enables the tracking of individual cells over time, allowing researchers to assess changes in cell behavior and morphology, which can help in optimizing conditions for CM production. Furthermore, cell segmentation can also help in identifying contaminating debris or inappropriate cells from the culture, ensuring the quality and safety of the final product.

Recently ML has been used to improve cell segmentation for various applications (221–224). Manual identification of individual cells has poor reliability between and within human evaluators. Additionally, automating cell imaging tasks can free up researchers' time for higher value tasks and reduce errors from fatigue and subjectivity (225). However, there are several challenges that hinder

the effectiveness of ML-based cell segmentation. First, cells are highly variable in size, shape, and morphology. Second, segmentation may fail on images that are low-contrast, have uneven illumination, or are out-of-focus. Third, crowded and overlapping cells can make it difficult to distinguish individual cells using cell-segmentation algorithms. Fourth, training data can overfit on imaging artifacts, such as non-uniform illumination and background noise (226). These challenges are not unique to CM but must be taken into account prior to applying ML for cell segmentation in any field.

In order to overcome the above challenges and to develop robust algorithms for detecting and segmenting cells, the computer vision community requires access to large, diverse, and well-curated datasets with comprehensive annotations. While some public datasets for nuclei and cell segmentation have been released in the past (223,225,227–230), additional datasets specific to CM are necessary to develop such robust algorithms for accurate cell segmentation in microscopic images.

A classic method for cell segmentation is the watershed algorithm combined with optimal thresholding, which has been used in applications such as segmenting lymphocytes into nuclear and cytoplasmic regions (231). There has been growing interest in deep neural net architectures inspired by fully convolutional networks for cell segmentation (223,230,232–235). These architectures employ encoder-decoder blocks to transfer features from multiple scales and levels for efficient cell segmentation on histopathology and microscopy images. The U-Net model, a variant of the fully convolutional network architecture, has shown particular promise for this task (236,237). Unlike fully convolutional networks, U-Net incorporates skip connections that facilitate precise semantic segmentation by amalgamating features from diverse resolutions, enhancing the model's capability to capture intricate details. Similarly, U-Net++ employs advanced encoder-decoder structures and loss functions to enhance performance for cell segmentation (238). Such deep learning architectures have outperformed pathologists' performance for cell segmentation in various applications (239,240).

5.2 Automatic Image Analysis for Cell Classification using Machine Learning

Cell classification is another important task in image analysis. While cell segmentation is the process of separating individual cells from the background and from each other in an image, cell classification is the process of assigning labels to the segmented cells based on their morphology, phenotype, or function (241,242). The primary goal of cell classification is to assign each segmented cell to a specific category or label, such as cell type, state, or condition.

Manual analysis of thousands of microscopy images for cell classification is a tedious and error-prone task. Thus, there is a need for ML algorithms for automatic cell classification. Recently, ML has been used for the classification of cells within microscopy images, including the identification of anemia or blood disorders based on the shape, size, and optical properties of red blood cells (243), and for the analysis of cellular microenvironments to offer novel insights into biological mechanisms (244). However, the effectiveness of ML-based cell classification for microscopy faces several challenges. One challenge arises from the crowded or overlapping cells, making it difficult to distinguish individual cells. Differences in cell maturity and variations in cell shape resulting from diverse cultivation and treatment methods can introduce complexity into the analysis, especially when distinguishing between different types of cultured cells. Complex textures, patterns, and shapes in multi-modal microscopy further complicate cell classification, requiring the integration of multi-stream models that consider low-level cues such as edges and gradients (245). The intricate cell structures found in tissue images introduce additional complexities such as

heterogeneous cell populations and staining variations. These challenges necessitate the development of precise and efficient algorithms for cell classification.

In the domain of microscopy image analysis, many innovative architectures and methods have emerged for classifying cells. Among these, CNNs stand as a robust choice, with models like U-Net (237) and Mask R-CNN (246) excelling in cell classification tasks. RNNs, particularly long short-term memory and gated recurrent units, demonstrate their prowess when handling sequential data, making them valuable for tracking cell dynamics (247). For tasks involving complex relationships between cells, GNNs, such as graph convolutional networks, prove invaluable (248). Traditional approaches like random forests and decision trees, which are based on hand-picked image features, are still relevant for classification because they are computationally fast, easy to train, and resist overfitting (249,250). Transfer learning techniques harness pre-trained deep learning models like ResNet (251), while attention mechanisms and ensemble methods contribute to improved accuracy (252). With the ever-evolving landscape of microscopy, these versatile architectures continue to play pivotal roles in cell classification.

Following cell classification, cell phenotype analysis is pivotal in exploring cellular characteristics, encompassing physical and biochemical attributes such as size, shape, function, viability, proliferation, signaling, and morphological structure. This approach is particularly valuable in culture media optimization (253). By scrutinizing the physical attributes and functional behavior of cultured cells, researchers can ensure the consistency and quality of cell populations, evaluate metabolic activity, and assess cellular functionality within the culture. Moreover, the examination of cell morphology offers insights into cellular health and enables fine-tuning of culture media formulations (253,254). ML has been used to combine different forms of microscopy using transfer learning - such as between fluorescence and dye-free microscopy (255) or between traditional microscopy and mass spectrometry imaging (256). Thus, in the burgeoning field of CM, phenotype analysis contributes to the refinement of CM production, guiding the selection of cell strains and culture conditions to achieve better quality and desired growth attributes. Overall, cell phenotype analysis plays a pivotal role in enhancing cell culture processes and product quality, making it a valuable tool for CM production.

6 Bioprocess and Food Processing Optimization

Moving CM from lab bench scale to commercial scale requires efficient bioprocess design. This centers around the use of large bioreactors to produce a controlled environment for cell growth and differentiation that maximizes biomass and minimizes by-product yields. A variety of bioreactor types have been proposed for use in CM, which are reviewed elsewhere (257). One study estimated that producing 1kg of protein from muscle cells would require stirred tank bioreactors on the order of 5000 liters (258). This dwarfs research-scale mammalian cell culture and will require extensive optimization. In addition, after harvesting cells from the bioreactor, CM products will likely require food processing steps to create a final product. ML is well suited to increase the scale and efficiency of CM bioprocessing and food processing in a variety of ways.

6.1 Bioreactor Homeostasis Can be Maintained with Machine Learning

As the culture scale is increased, automation of closed systems will become important to increase efficiency and reduce contamination or other failure events (259). Real-time quality assurance and process monitoring will allow for adjusting culture conditions to optimize yield, monitoring for potential contamination, and reducing human errors. Automation in processes such as media recycling will further optimize the process and bring costs down.

Historically, mammalian bioreactor operation has been governed by systems such as proportional-integral-derivative controllers (260) or model predictive control (261). These systems, although mainstays of control theory, are designed to work in deterministic environments that can be well-described by linear differential equations. Although these systems can struggle with the vagaries and unpredictability of biological systems, they have been successfully employed in bioreactors to maximize antibody production (262), constrain overflow metabolism (263), and maintain glucose homeostasis (264). This type of modeling has ceded ground to ML-based modeling in recent years, probably because, compared to ML, these models are relatively fragile because of their dependence on their top-down mathematical models accurately describing reality.

The most straightforward application of ML to bioreactors is to use ML-based models to control the bioreactor's inputs. A wide gamut of ML models have been used to monitor and control bioreactors and industrial bioprocesses, especially supervised learning models like neural networks (265), random forests (266), and gradient boosting (267). The step beyond using ML to optimize models is to use ML to optimize the policies themselves that determine the models, which is done by reinforcement learning, with examples in bioreactors ranging across deep Q-networks (268), policy gradients (269), and probabilistic Bayesian optimization (270). These policy-based learning methods do not necessarily require a prior understanding of the biochemistry of the bioreactor.

6.2 The Unique Challenges of Structured Products Can Be Addressed by Machine Learning

As opposed to ground meat, structured tissues (i.e. a steak or fish filet) will require the formation of organized 3-dimensional tissues and bioreactor systems capable of supporting them. A structured product entails growing an organized and three-dimensional tissue, as opposed to growing cells in suspension - the difference between growing cells as a soup and growing cells as a steak. This type of engineered tissue has not been shown on any scale beyond a tissue for a single patient, making this a major whitespace in the field (259).

Imposing organization upon cells is a classic, challenging problem from the field of tissue engineering, and solutions often involve ML. Random forests (271) and neural networks (272) have been used to predict optimal parameters for the extrusion printing of hydrogel scaffolds. Gradient boosting has been used to predict the self-assembly of dipeptide-based hydrogel scaffolds (273). In theory, ML could be used to design the structures of tissue scaffolds, but this avenue appears to be unexplored at this time.

Tissue self-organization is a powerful principle for structured meat production. Tissue scaffolds or extrusion printing are currently used to create top-down structured tissues because of a lack of understanding of bottom-up self-organization of tissue in vitro. However, breakthroughs in bottom-up tissue self-organization would accelerate the scale-up of CM, eliminating the costs of scaffolds. In the context of CM, relevant types of self-organization are the alignment of fibers (especially myofibrils and collagen cables), the production of functional vasculature networks, maintaining ratios of meat-relevant cell types (myocytes, adipocytes, fibroblasts), and the structural determinants of texture and mouthfeel (274). Attempts have been made to model tissue self-organization using differential adhesion (275), cellular Potts models (276), and agent-based modeling (277). Accurate models of tissue self-organization may provide design principles for making tissues with defined structure in CM.

A family of methods that may help with structured tissue construction in CM is spatial transcriptomics, which correlates gene expression in situ to physical coordinates within a tissue section (278). Spatial transcriptomics has been used for understanding specific aspects of tissue

structure and cell organization that rely on spatial context, such as extracellular forces and gradients of signaling molecules (136). GNNs have been successfully used with spatial transcriptomics data to model cellular communication (279–281) and the deep learning model Tangram has also been used to resolve cell types and decrease imputation error from spatial data (282). Although spatial transcriptomics has not yet been used in the context of CM, it could prove useful in dissecting the complex tissue architectures involved in structured products.

Structured products will also require bioreactor systems capable of perfusion and harvesting of large intact tissues. The nature of a structured product bioreactor bears similarity to the fluidized and/or packed bed bioreactors that are used industrially in the wastewater and mineral extraction industries. Numerical fluid dynamical models are classically used to predict the behavior of these systems, but ML has, in recent years, been used for the more unpredictable aspects of these systems (283,284). In particular, gradient boosting has been used both to predict bed expansion of fluidized bed bioreactors (285) and mass transfer in packed bed bioreactors (286). Similar methods may be useful for the particular challenge of CM structured product bioreactors.

6.3 Real-Time Sensory Prediction and Control Could be Applied with Reinforcement Learning

Unlike tissue engineering for medical treatments, the sensory properties of CM, such as flavor and texture, are critical to its commercial success. These may be generated during the cell culture from flavor or texture components of cells, media, or scaffolds, or after harvest using food processing or additives. ML is playing an important role in enhancing the analysis of the flavors and textures of other food products through the analysis of diverse data types and these techniques are likely to play a role in CM development as well.

During product development, flavor can be measured in a variety of ways. Earlier ML models used data from gas chromatography-mass spectrometry, which is an analytical chemistry method used to separate and fingerprint substances from complex mixtures (287,288). Later, researchers focused on electronic noses that mimic the olfactory capability of humans through different sensors. Since an electronic nose can be used to collect real-time sensor data during production, ML can also be applied in real-time for quality and flavor control (289–291). However, it should be taken into consideration that many compounds important for the aroma of meat are generated during cooking (292). The latest research efforts concentrate on utilizing the molecular structure and physicochemical attributes of flavor compounds to predict flavors, including taste or smell (293–297). These characteristics are quantified into molecular descriptors, numerical representations that encapsulate the properties of the molecules involved. These descriptors then serve as inputs for ML models, which are trained to predict flavor profiles and odor characteristics with greater objectivity.

Since most of these data are tabular, a wide range of traditional ML approaches, such as support vector machines, random forests, k-nearest neighbour, and AdaBoost Tree, have been employed (295,297,298). Additionally, deep learning approaches, such as CNNs (287) and multilayer perceptrons (288,289,291), and unsupervised learning approaches, such as cluster analysis by using principal component analysis (289,291), have also been applied to identify or predict flavors. These methodologies have collectively demonstrated that ML can significantly contribute to the enhancement of the sensory properties of a range of food products.

AI approaches, particularly RL, offer significant potential for enhancing efficiency in food processing operations (269,299,300). Food-processing facilities are typically equipped with a variety of sensors,

including those for temperature, pressure, moisture, and pH levels. These sensors play a crucial role in ensuring precise ingredient measurements, which are fundamental for achieving the desired taste and aroma profiles of food products. However, the challenge arises when new products are developed, as determining the optimal mixture of ingredients often involves extensive trial and error. Deploying an RL agent in this context can effectively manage and adjust the various sensor readings, thereby ensuring that the food consistently meets the specific standards and requirements set by the food processor. This approach not only streamlines the development process but also enhances the precision and quality of the final product. RL can also be valuable for sensory prediction, in particular the texture prediction of finished CM products (301).

7 Discussion

In the last decade, the field of CM has made strides toward lower-cost and more efficient production processes but must progress significantly further to effectively rival traditional meat. ML offers great promise in improving every stage of CM production, from cell line development to the final product's sensory characteristics. With the current surge in both public and private research and development for both the ML and CM fields, and the already successful integration of ML into numerous life sciences fields, the integration of ML into CM research is timely. Despite the numerous opportunities, there are only a handful of peer-reviewed, publicly accessible studies that describe the use of ML in CM production, and an equally small group of researchers versed in both ML and CM. This review aims to bridge the gap between the ML and CM fields and create a starting point for scientists to better understand how to apply ML to CM research.

Since ML has been successfully employed in many other bioinformatics sectors, many of the existing methods can be adopted to CM. However, the key limitation is the availability of sufficient data. Creating ML models requires large training and validation datasets, and ML models are only as robust and reliable as the amount and quality of data used to develop them. The scarcity of public data in CM complicates the development of models or even the assessment of potential model types. There are some publicly available dataset repositories, such as the Gene Expression Omnibus (GEO) for RNA-seq data[1], GenBank for sequenced genome data[2], and Uniprot for protein sequences[3]. However, compared to species used in medical studies, few CM-relevant datasets are reported and many lack adequate descriptions, data annotations, or samples and replicates to be considered for statistical analysis. As an example, a query for “stem cell” in the GEO DataSets generates 71327 datasets for Homo sapiens (human) and only 61 for Bos taurus (cattle) (as of Feb 1, 2024). A survey of datasets with potential relevance to CM research is included in Supplementary Table 1 (hosted at https://zenodo.org/doi/10.5281/zenodo.11092402). Efforts to generate properly annotated data and incentives for the sharing of data from ongoing experiments would greatly accelerate the application of ML to CM research. The Cultivated Meat Modeling Consortium[4] offers a model for community-generated and shared data, while protecting intellectual property, to accelerate computational models.

Transfer learning might be able to partially make up for the scarcity of CM-relevant data, by using models from data-rich biomedical species, such as humans and mice, to inform models for data-poor CM-relevant species. Cross-species graph-based transfer learning has previously been applied in a non-CM context on RNA-seq data for cell-type identification (302,303). Challenges to this approach

[1] https://www.ncbi.nlm.nih.gov/geo/
[2] https://www.ncbi.nlm.nih.gov/genbank/
[3] https://www.uniprot.org/
[4] https://thecmmc.org/

include biological heterogeneity from differing sets of genes or differing functions for genes across species, which can potentially be mitigated with gene orthology information (304). As a starting point, researchers may look into fine-tuning large scale models that have been successful in achieving improved performance on biological tasks relevant to CM with limited task-specific data, including gene network analysis and cell type annotation, such as Geneformer and scGPT (305,306).

Another challenge relates to the scale of existing studies. Most ML research related to CM has been limited to laboratory environments, which might not mirror the conditions of mass production. Solutions could include either verifying these lab models at a commercial level or using process simulations to adapt them for larger operations.

From the perspective of biologists, an important way to speed up ML work is to produce the biological models that are used to generate data for training ML models. For example, the production and dissemination of more high-quality cell lines from agriculturally relevant animals is needed to aid the generation of omics and microscopy data. Furthermore, biological scientists can make efforts to contribute to the body of existing data by publishing any quality data that results from their experiments, whether or not they are directly used in their own studies, including negative results. Publication of transcriptomic and epigenomic data has become more commonplace, however complete microscopy image sets, in particular, are rarely published. There is also a need for more proteomic and metabolomic data, especially in understudied food-relevant species, to address numerous use cases including metabolic modeling, flavor profiling, and bioreactor scaling concerns (307). Moreover, datasets should be properly annotated - this first includes metadata describing the samples the data came from, how the data was generated, and any processing that was done to the data. Ideally, data would include all raw data, which ML scientists could use to regenerate the original dataset. Similarly, properly labeled data (labeling of individual data points) is important for supervised learning models, which remain the most popular and widely used. Finally, the quality and consistency of datasets are also critical for their use in training ML models. Variations in sample handling and data collection can lead to large and varied systematic errors that make it challenging to usefully combine multiple datasets or for a model trained on one dataset to apply to another.

Efforts to make ML more accessible to researchers, such as infrastructure, frameworks, benchmarks, and libraries, could help facilitate the application of ML to CM. Currently, a variety of ML models are freely available from online resources like Paperswithcode[5] and Hugging Face[6]. Additionally, libraries such as Python's scikit-learn (308) or frameworks such as PyTorch Lightning[7] offer a good starting point for coding ML. Similarly, accessible web interfaces could make ML tools more accessible to biologists, following the examples of AlphaFold and Foldseek, which both have interfaces integrated into the widely used online protein database UniProt.

Overall, there is an enormous opportunity for CM researchers to incorporate ML techniques and for ML professionals to explore the CM field. The ML field has been moving at unprecedented speed over the past few years, and CM researchers could make gains simply by porting over what is, in effect, yesterday's news in ML. This review aims to be an introductory resource for researchers eager to explore this cross-disciplinary method, which could help to establish CM as a viable and sustainable protein source in our diets.

[5] https://paperswithcode.com/
[6] https://huggingface.co/
[7] https://lightning.ai/

Abbreviations

AI: artificial intelligence

CHO: Chinese hamster ovary

CM: cultured meat

CNN: convolutional neural network

FAO: Food and Agriculture Organization

GAN: generative adversarial network

GNI: gene network inference

GNN: graph neural network

ML: machine learning

NLP: natural language processing

OECD: Organization for Economic Co-operation

RL: reinforcement learning

RNN: recurrent neural network

VAE: Variational Autoencoder

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

BD, MET, JS, and RV were funded by a grant from the Schmidt Futures.

Acknowledgments

The authors would like to thank Zachary Consenza, Roy Nadler, Amin Nikkhah, Dave Staszak, Charlie Taylor, Sofia Giampaoli, Animesh Acharjee, Ana Velez Rueda, Jessica Carballido, and Leandro Sommese for reviewing early drafts of this manuscript and providing valuable insights. We thank Isha Datar and Cam Linke for their assistance in finding funding for the project. We would also like to thank Evan Rapoport for his facilitative leadership on this project.

Figures

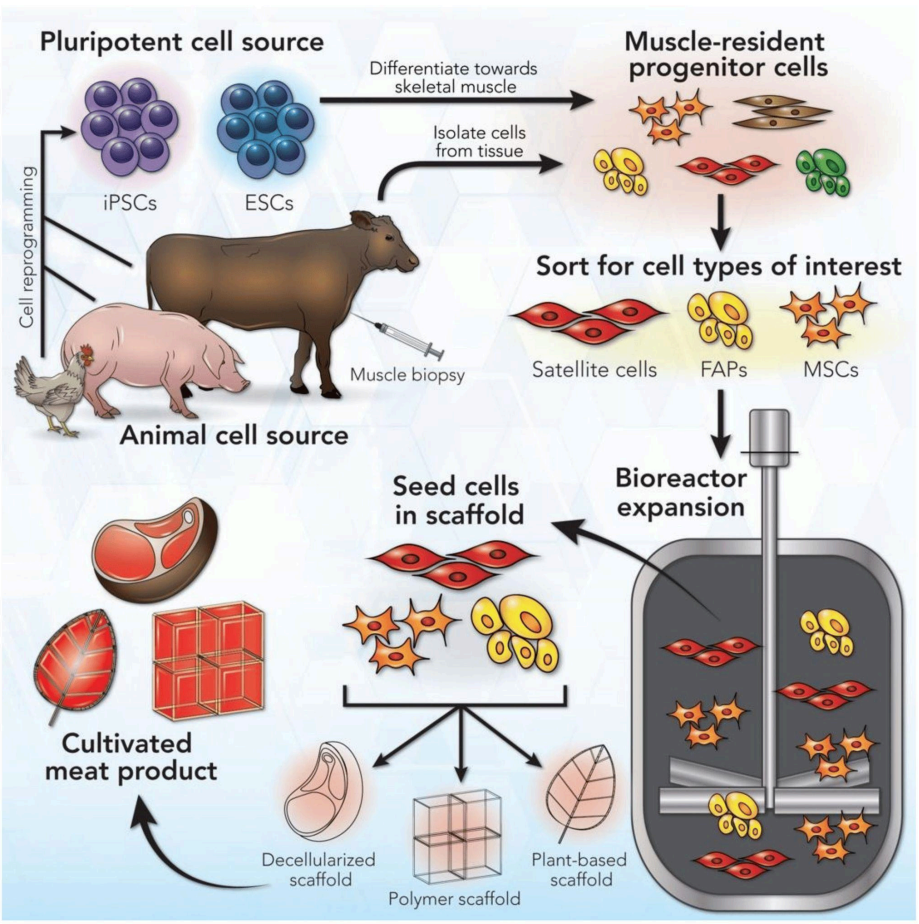

Figure 1. Cultured meat manufacturing process. Reprinted from Reiss, J. et.al., 2021. Creative Commons Attribution (CC BY) license. (29)

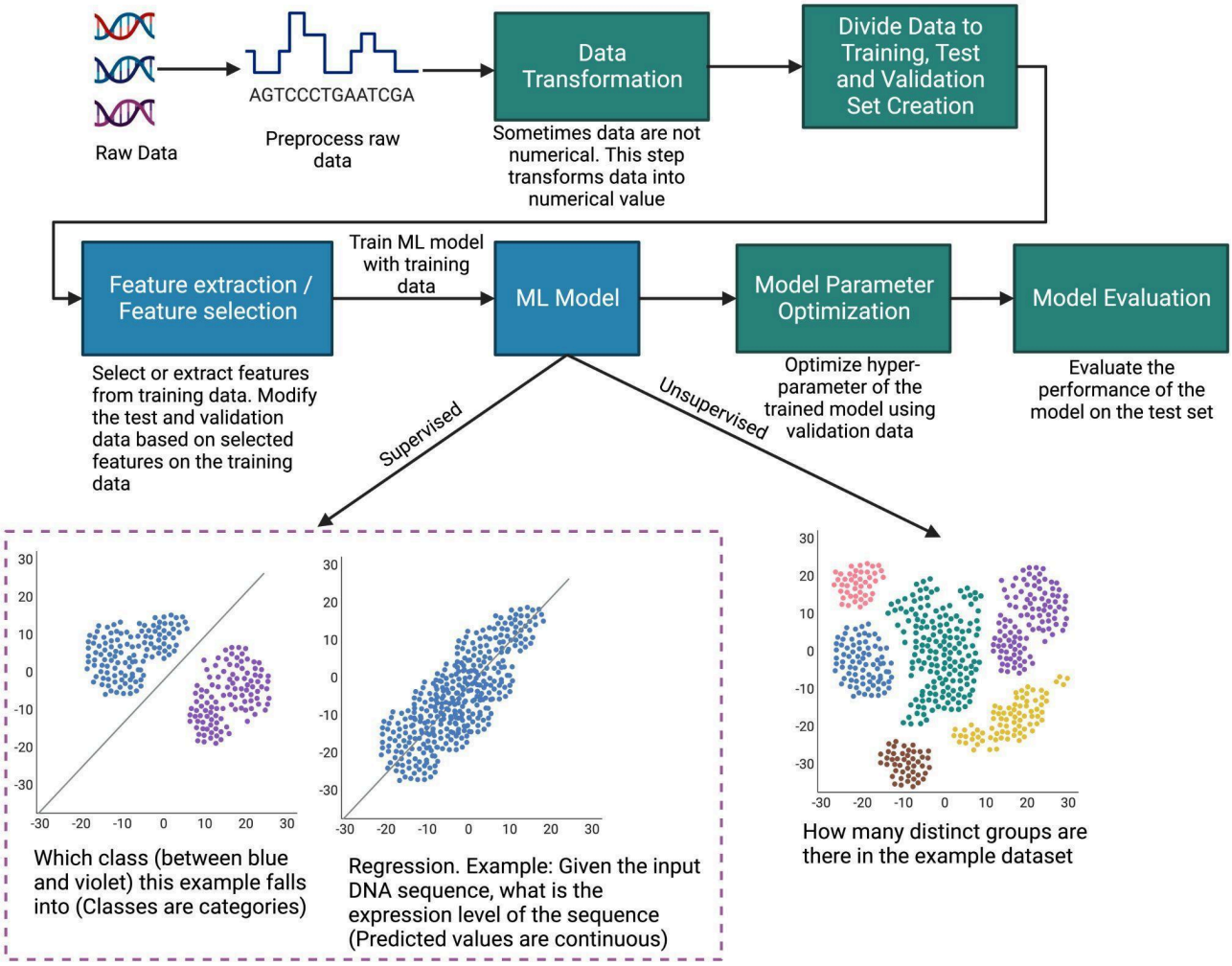

Figure 2. Illustration of the typical steps of how machine learning can be applied to biological data.

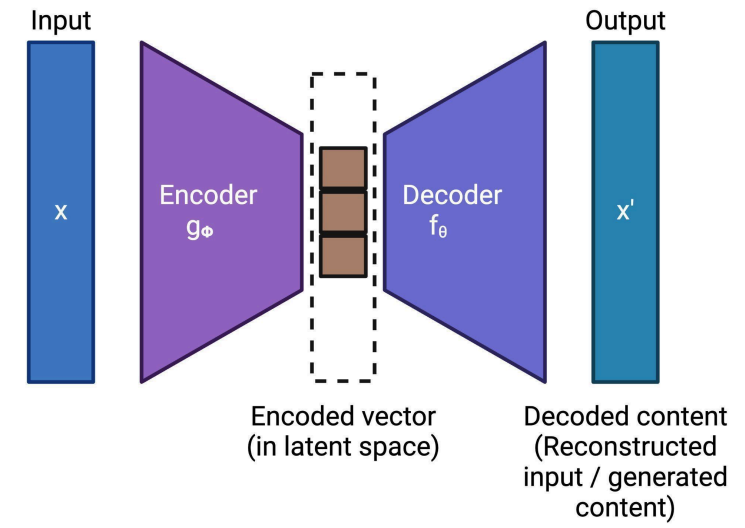

Figure 3. General architecture of a Variational Autoencoder (VAE).

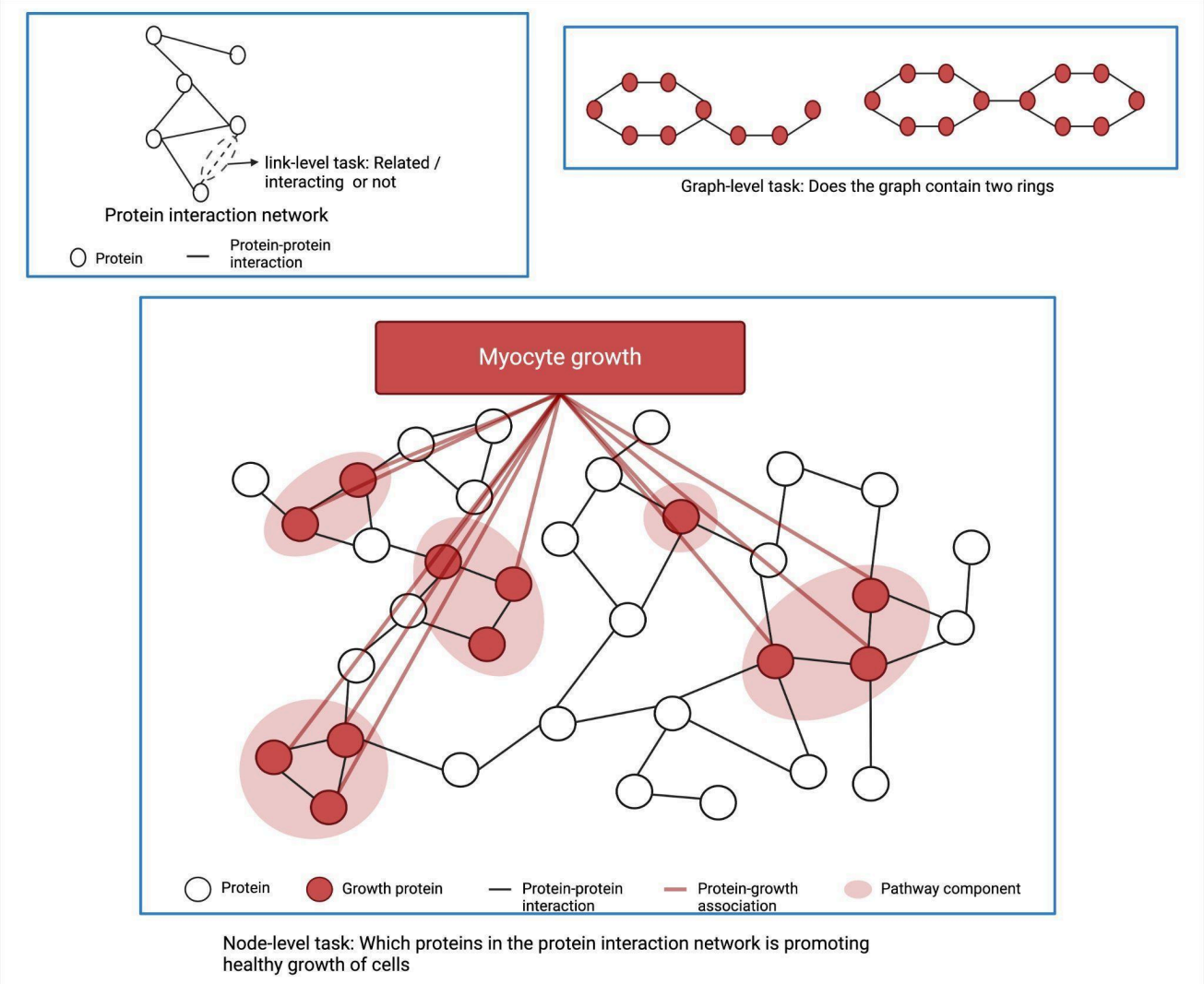

Figure 4. Typical machine learning tasks for network analysis.

References

1. Crippa M, Solazzo E, Guizzardi D, Monforti-Ferrario F, Tubiello FN, Leip A. Food systems are responsible for a third of global anthropogenic GHG emissions. Nat Food. 2021 Mar;2(3):198–209.
2. Steinfeld H, Gerber P, Wassenaar T, Castel V, Rosales M, de Haan C. Livestock's long shadow. 2006 [cited 2023 May 19]; Available from: https://www.fao.org/3/a0701e/a0701e00.htm
3. Poore J, Nemecek T. Reducing food's environmental impacts through producers and consumers. Science. 2018 Jun;360(6392):987–92.
4. Michele P. OECD-FAO Agricultural Outlook 2021-2030. 2021; Available from: https://www.fao.org/documents/card/en?details=cb5332en
5. Clark MA, Domingo NGG, Colgan K, Thakrar SK, Tilman D, Lynch J, et al. Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets. Science. 2020 Nov 6;370(6517):705–8.
6. Ivanovich CC, Sun T, Gordon DR, Ocko IB. Future warming from global food consumption. Nat Clim Change. 2023 Mar;13(3):297–302.
7. Sinke P, Swartz E, Sanctorum H, van der Giesen C, Odegard I. Ex-ante life cycle assessment of commercial-scale cultivated meat production in 2030. Int J Life Cycle Assess. 2023 Mar 1;28(3):234–54.
8. Tuomisto HL, Teixeira de Mattos MJ. Environmental Impacts of Cultured Meat Production. Environ Sci Technol. 2011 Jul 15;45(14):6117–23.
9. Risner D, Li F, Fell JS, Pace SA, Siegel JB, Tagkopoulos I, et al. Preliminary Techno-Economic Assessment of Animal Cell-Based Meat. Foods. 2021 Jan;10(1):3.
10. Business Wire. Triplebar and Umami Meats Initiate Technology Collaboration To Optimize Cell Lines for Sustainable Seafood [Internet]. 2023 [cited 2024 Apr 2]. Available from: https://www.businesswire.com/news/home/20230302005187/en/Triplebar-and-Umami-Meats-Initiate-Technology-Collaboration-To-Optimize-Cell-Lines-for-Sustainable-Seafood
11. Ho S. This Cambridge Startup's 'Renaissance Farm' Uses AI To Make Cell-Based Meat - Green Queen [Internet]. Green Queen. 2021 [cited 2023 May 17]. Available from: https://www.greenqueen.com.hk/animal-alternative-technologies-renaissance-farm-cell-based-meat/
12. Leach C. Science Entrepreneur Club. 2024 [cited 2023 Nov 30]. Alt Atlas: Artificial Intelligence powering real meat of the future. Available from: https://www.science-entrepreneur.com/insights/artificial-intelligence-powering-real-meat-of-the-future
13. Marston J. AFN. 2022 [cited 2023 May 17]. Brief: Prolific Machines raises $42m from Breakthrough, Mayfield to slash cultivated meat costs. Available from: https://agfundernews.com/prolific-machines-raises-42m-to-slash-cultivated-meat-manufacturing-costs
14. Penarredonda JL. BBC. 2017 [cited 2023 May 17]. Could AI help to create a meat-free world? Available from: https://www.bbc.com/future/article/20171214-could-ai-help-create-a-meat-free-world
15. Protein Report. Protein Report. 2022 [cited 2023 May 17]. Hoxton Farms Raises a $22m Series A to Solve the Big Fat Problem for Meat Alternatives. Available from: https://www.proteinreport.org/newswire/hoxton-farms-raises-22m-series-solve-big-fat-problem-meat-alternatives

16. Shieber J. This Y Combinator startup is taking lab-grown meat upscale with elk, lamb and Wagyu beef cell lines [Internet]. TechCrunch. 2021 [cited 2023 May 17]. Available from: https://techcrunch.com/2021/03/23/this-y-combinator-startup-is-taking-lab-grown-meat-upscale-with-elk-lamb-and-wagyu-beef-cell-lines/
17. Southey F. foodnavigator.com. 2023 [cited 2023 May 17]. 'World's first' growth media factory in the works for cultivated meat, milk. Available from: https://www.foodnavigator.com/Article/2023/01/16/World-s-first-growth-media-factory-in-the-works-for-cultivated-meat-milk
18. Cosenza Z, Block DE, Baar K, Chen X. Multi-objective Bayesian algorithm automatically discovers low-cost high-growth serum-free media for cellular agriculture application. Eng Life Sci. 2023;23(8):e2300005.
19. Cosenza Z, Astudillo R, Frazier PI, Baar K, Block DE. Multi-information source Bayesian optimization of culture media for cellular agriculture. Biotechnol Bioeng. 2022;119(9):2447–58.
20. Cosenza Z, Block DE, Baar K. Optimization of muscle cell culture media using nonlinear design of experiments. Biotechnol J. 2021;16(11):2100228.
21. Nikkhah A, Rohani A, Zarei M, Kulkarni A, Batarseh FA, Blackstone NT, et al. Toward sustainable culture media: Using artificial intelligence to optimize reduced-serum formulations for cultivated meat. Sci Total Environ. 2023 Oct 10;894:164988.
22. Post MJ, Levenberg S, Kaplan DL, Genovese N, Fu J, Bryant CJ, et al. Scientific, sustainability and regulatory challenges of cultured meat. Nat Food. 2020 Jul;1(7):403–15.
23. Beheshtizadeh N, Gharibshahian M, Pazhouhnia Z, Rostami M, Zangi AR, Maleki R, et al. Commercialization and regulation of regenerative medicine products: Promises, advances and challenges. Biomed Pharmacother. 2022 Sep 1;153:113431.
24. Battle M, Bomkamp C, Carter M, Colley Clarke J, Fathman L, Gertner D, et al. 2023 State of the Industry Report: Cultivated meat and seafood [Internet]. Good Food Institute; 2024. Available from: https://gfi.org/resource/cultivated-meat-and-seafood-state-of-the-industry-report/
25. Kupferschmidt K. Lab Burger Adds Sizzle to Bid for Research Funds. Science. 2013 Aug 9;341(6146):602–3.
26. Poinski M. Future Meat Technologies raises $347M and makes a $1.70 cell-based chicken breast. Food Dive [Internet]. 2021 Dec 20 [cited 2023 Apr 12]; Available from: https://www.fooddive.com/news/future-meat-technologies-raises-347m-and-makes-a-170-cell-based-chicken/611712/
27. FAO. World Food and Agriculture - Statistical Yearbook 2020 [Internet]. Rome, Italy: FAO; 2020 [cited 2023 Apr 27]. 366 p. (FAO Statistical Yearbook – World Food and Agriculture). Available from: https://www.fao.org/documents/card/en/c/cb1329en
28. Humbird D. Scale-Up Economics for Cultured Meat: Techno-Economic Analysis and Due Diligence [Internet]. Engineering Archive; 2020 [cited 2023 Nov 30]. Available from: https://engrxiv.org/preprint/view/1438
29. Reiss J, Robertson S, Suzuki M. Cell Sources for Cultivated Meat: Applications and Considerations throughout the Production Workflow. Int J Mol Sci. 2021 Jan;22(14):7513.
30. Bishop CM. Pattern Recognition and Machine Learning [Internet]. 2006 [cited 2024 Feb 23]. Available from: https://link.springer.com/book/9780387310732
31. Li Z, Liu F, Yang W, Peng S, Zhou J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Trans Neural Netw Learn Syst. 2022 Dec;33(12):6999–7019.
32. Lipton ZC, Berkowitz J, Elkan C. A Critical Review of Recurrent Neural Networks for Sequence Learning [Internet]. arXiv; 2015 [cited 2024 Apr 16]. Available from:

http://arxiv.org/abs/1506.00019
33. Alpaydin E. Introduction to Machine Learning, fourth edition. MIT Press; 2020. 709 p.
34. Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, et al. Machine learning in bioinformatics. Brief Bioinform. 2006 Mar;7(1):86–112.
35. Parasa NA, Namgiri JV, Mohanty SN, Dash JK. Introduction to unsupervised learning in bioinformatics. Data Anal Bioinforma Mach Learn Perspect. 2021;35–49.
36. Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT press; 2018.
37. Angermueller C, Dohan D, Belanger D, Deshpande R, Murphy K, Colwell L. Model-based reinforcement learning for biological sequence design. 2020;
38. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9.
39. Neftci EO, Averbeck BB. Reinforcement learning in artificial and biological systems. Nat Mach Intell. 2019;1(3):133–43.
40. Jafari R, Javidi MM, Kuchaki Rafsanjani M. Using deep reinforcement learning approach for solving the multiple sequence alignment problem. SN Appl Sci. 2019 May 18;1(6):592.
41. Barozet A, Molloy K, Vaisset M, Siméon T, Cortés J. A reinforcement-learning-based approach to enhance exhaustive protein loop sampling. Bioinformatics. 2020 Feb 15;36(4):1099–106.
42. Menden K, Marouf M, Oller S, Dalmia A, Magruder DS, Kloiber K, et al. Deep learning–based cell composition analysis from tissue expression profiles. Sci Adv. 2020;6(30):eaba2619.
43. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436–44.
44. Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw. 2015;61:85–117.
45. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2017 [cited 2024 Apr 16]. Available from: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
46. Li H, Tian S, Li Y, Fang Q, Tan R, Pan Y, et al. Modern deep learning in bioinformatics. J Mol Cell Biol. 2020 Jun;12(11):823–7.
47. Li Y. Deep Reinforcement Learning: An Overview [Internet]. arXiv; 2018 [cited 2023 Dec 20]. Available from: http://arxiv.org/abs/1701.07274
48. Kingma DP, Welling M. An Introduction to Variational Autoencoders. Found Trends® Mach Learn. 2019;12(4):307–92.
49. Singh A, Ogunfunmi T. An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications. Entropy. 2022 Jan;24(1):55.
50. Wei R, Mahmood A. Recent Advances in Variational Autoencoders With Representation Learning for Biomedical Informatics: A Survey. IEEE Access. 2021;9:4939–56.
51. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative Adversarial Networks: An Overview. IEEE Signal Process Mag. 2018 Jan;35(1):53–65.
52. Zhang Y, Yang L, Chen J, Fredericksen M, Hughes DP, Chen DZ. Deep Adversarial Networks for Biomedical Image Segmentation Utilizing Unannotated Images. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. Medical Image Computing and Computer Assisted Intervention − MICCAI 2017. Cham: Springer International Publishing; 2017. p. 408–16.

53. Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. Bioinformatics. 2021;37(15):2112–20.
54. Zhang D, Zhang W, Zhao Y, Zhang J, He B, Qin C, et al. DNAGPT: A Generalized Pre-trained Tool for Multiple DNA Sequence Analysis Tasks [Internet]. bioRxiv; 2024 [cited 2024 Apr 16]. p. 2023.07.11.548628. Available from: https://www.biorxiv.org/content/10.1101/2023.07.11.548628v3
55. Bran AM, Schwaller P. Transformers and Large Language Models for Chemistry and Drug Discovery [Internet]. arXiv; 2023 [cited 2024 Apr 16]. Available from: http://arxiv.org/abs/2310.06083
56. Nguyen E, Poli M, Durrant MG, Thomas AW, Kang B, Sullivan J, et al. Sequence modeling and design from molecular to genome scale with Evo [Internet]. bioRxiv; 2024 [cited 2024 Apr 16]. p. 2024.02.27.582234. Available from: https://www.biorxiv.org/content/10.1101/2024.02.27.582234v1
57. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The Graph Neural Network Model. IEEE Trans Neural Netw. 2009 Jan;20(1):61–80.
58. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: A review of methods and applications. AI Open. 2020 Jan 1;1:57–81.
59. Kumar A, Singh SS, Singh K, Biswas B. Link prediction techniques, applications, and performance: A survey. Phys Stat Mech Its Appl. 2020 Sep 1;553:124289.
60. Li H, Sun Y, Hong H, Huang X, Tao H, Huang Q, et al. Inferring transcription factor regulatory networks from single-cell ATAC-seq data based on graph neural networks. Nat Mach Intell. 2022 Apr;4(4):389–400.
61. Muzio G, O'Bray L, Borgwardt K. Biological network analysis with deep learning. Brief Bioinform. 2021 Mar 1;22(2):1515–30.
62. Nelson W, Zitnik M, Wang B, Leskovec J, Goldenberg A, Sharan R. To Embed or Not: Network Embedding as a Paradigm in Computational Biology. Front Genet [Internet]. 2019 [cited 2024 Jan 24];10. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2019.00381
63. Listrat A, Lebret B, Louveau I, Astruc T, Bonnet M, Lefaucheur L, et al. How Muscle Structure and Composition Influence Meat and Flesh Quality. Sci World J. 2016 Feb 28;2016:e3182746.
64. Ben-Arye T, Levenberg S. Tissue Engineering for Clean Meat Production. Front Sustain Food Syst. 2019 Jun 18;3:46.
65. Reiss J, Robertson S, Suzuki M. Cell Sources for Cultivated Meat: Applications and Considerations throughout the Production Workflow. Int J Mol Sci. 2021 Jul 13;22(14):7513.
66. Smith-Uchotsk R, Wanjiru P. Identification of hazards in meat products manufactured from cultured animal cells: executive summary [Internet]. Risk Assessment Unit Science, Evidence and Research Division, Food Standards Agency; 2023 Mar [cited 2023 Oct 10]. Available from: https://www.food.gov.uk/research/identification-of-hazards-in-meat-products-manufactured-from-cultured-animal-cells-executive-summary
67. Stout AJ, Mirliani AB, Soule-Albridge EL, Cohen JM, Kaplan DL. Engineering carotenoid production in mammalian cells for nutritionally enhanced cell-cultured foods. Metab Eng. 2020 Nov;62:126–37.
68. Cong YS, Wright WE, Shay JW. Human Telomerase and Its Regulation. Microbiol Mol Biol Rev. 2002 Sep;66(3):407–25.
69. Hayflick L. The limited in vitro lifetime of human diploid cell strains. Exp Cell Res. 1965 Mar 1;37(3):614–36.

70. Soice E, Johnston J. Immortalizing Cells for Human Consumption. Int J Mol Sci. 2021 Oct 28;22(21):11660.
71. Béjar J, Hong Y, Alvarez MC. An ES-like cell line from the marine fish Sparus aurata: characterization and chimaera production. Transgenic Res. 2002 Jun;11(3):279–89.
72. Buonocore F, Libertini A, Prugnoli D, Mazzini M, Scapigliati G. Production and Characterization of a Continuous Embryonic Cell Line from Sea Bass (Dicentrarchus labrax L.). Mar Biotechnol. 2006 Jan 1;8(1):80–5.
73. Gignac SJ, Vo NTK, Mikhaeil MS, Alexander JAN, MacLatchy DL, Schulte PM, et al. Derivation of a continuous myogenic cell culture from an embryo of common killifish, Fundulus heteroclitus. Comp Biochem Physiol A Mol Integr Physiol. 2014 Sep 1;175:15–27.
74. Goswami M, Pinto N, Yashwanth BS, Sathiyanarayanan A, Ovissipour R. Development of a cell line from skeletal trunk muscle of the fish Labeo rohita. Cytotechnology. 2023 Aug 1;75(4):349–61.
75. Krishnan S, Ulagesan S, Cadangin J, Lee JH, Nam TJ, Choi YH. Establishment and Characterization of Continuous Satellite Muscle Cells from Olive Flounder (Paralichthys olivaceus): Isolation, Culture Conditions, and Myogenic Protein Expression. Cells. 2023 Jan;12(18):2325.
76. Li N, Guo L, Guo H. Establishment, characterization, and transfection potential of a new continuous fish cell line (CAM) derived from the muscle tissue of grass goldfish (Carassius auratus). Vitro Cell Dev Biol - Anim. 2021 Oct 1;57(9):912–31.
77. Parameswaran V, Shukla R, Bhonde R, Hameed ASS. Development of a Pluripotent ES-like Cell Line from Asian Sea Bass (Lates calcarifer)—An Oviparous Stem Cell Line Mimicking Viviparous ES Cells. Mar Biotechnol. 2007 Dec 1;9(6):766–75.
78. Parton A, Forest D, Kobayashi H, Dowell L, Bayne C, Barnes D. Cell and molecular biology of SAE, a cell line from the spiny dogfish shark, Squalus acanthias. Comp Biochem Physiol Part C Toxicol Pharmacol. 2007 Feb 1;145(1):111–9.
79. Potter G, Smith AST, Vo NTK, Muster J, Weston W, Bertero A, et al. A More Open Approach Is Needed to Develop Cell-Based Fish Technology: It Starts with Zebrafish. One Earth. 2020 Jul 24;3(1):54–64.
80. Saad MK, Yuen JSK, Joyce CM, Li X, Lim T, Wolfson TL, et al. Continuous fish muscle cell line with capacity for myogenic and adipogenic-like phenotypes. Sci Rep. 2023 Mar 29;13(1):5098.
81. Rubio N, Datar I, Stachura D, Kaplan D, Krueger K. Cell-Based Fish: A Novel Approach to Seafood Production and an Opportunity for Cellular Agriculture. Front Sustain Food Syst [Internet]. 2019 [cited 2023 May 19];3. Available from: https://www.frontiersin.org/articles/10.3389/fsufs.2019.00043
82. Chen ZH, You ZH, Guo ZH, Yi HC, Luo GX, Wang YB. Prediction of Drug–Target Interactions From Multi-Molecular Network Based on Deep Walk Embedding Model. Front Bioeng Biotechnol [Internet]. 2020 [cited 2024 Jan 24];8. Available from: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00338
83. Perozzi B, Al-Rfou R, Skiena S. DeepWalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2014 [cited 2024 Jan 24]. p. 701–10. (KDD '14). Available from: https://doi.org/10.1145/2623330.2623732
84. Yang L, Li LP, Yi HC. DeepWalk based method to predict lncRNA-miRNA associations via lncRNA-miRNA-disease-protein-drug graph. BMC Bioinformatics. 2022 Feb 25;22(12):621.
85. Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. LINE: Large-scale Information

Network Embedding. In: Proceedings of the 24th International Conference on World Wide Web [Internet]. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee; 2015 [cited 2024 Jan 24]. p. 1067–77. (WWW '15). Available from: https://doi.org/10.1145/2736277.2741093
86. Wang N, Zeng M, Li Y, Wu F xiang, Li M. Essential Protein Prediction Based on node2vec and XGBoost. J Comput Biol. 2021 Jul;28(7):687–700.
87. Yue Y, Ye C, Peng PY, Zhai HX, Ahmad I, Xia C, et al. A deep learning framework for identifying essential proteins based on multiple biological information. BMC Bioinformatics. 2022 Aug 4;23(1):318.
88. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space [Internet]. arXiv; 2013 [cited 2024 Jan 24]. Available from: http://arxiv.org/abs/1301.3781
89. Jha K, Saha S, Singh H. Prediction of protein–protein interaction using graph neural networks. Sci Rep. 2022 May 19;12(1):8360.
90. Yang F, Fan K, Song D, Lin H. Graph-based prediction of Protein-protein interactions with attributed signed graph embedding. BMC Bioinformatics. 2020 Jul 21;21(1):323.
91. Huang K, Xiao C, Glass LM, Zitnik M, Sun J. SkipGNN: predicting molecular interactions with skip-graph networks. Sci Rep. 2020 Dec 3;10(1):21092.
92. Kang C, Zhang H, Liu Z, Huang S, Yin Y. LR-GNN: a graph neural network based on link representation for predicting molecular associations. Brief Bioinform. 2022 Jan 1;23(1):bbab513.
93. Sun F, Sun J, Zhao Q. A deep learning method for predicting metabolite–disease associations via graph neural network. Brief Bioinform. 2022 Jul 1;23(4):bbac266.
94. Ciortan M, Defrance M. GNN-based embedding for clustering scRNA-seq data. Bioinformatics. 2021 Nov;38(4):1037–44.
95. Sun C, Xuan P, Zhang T, Ye Y. Graph Convolutional Autoencoder and Generative Adversarial Network-Based Method for Predicting Drug-Target Interactions. IEEE/ACM Trans Comput Biol Bioinform. 2022 Jan;19(1):455–64.
96. Zhang J, Jiang Z, Hu X, Song B. A novel graph attention adversarial network for predicting disease-related associations. Methods. 2020 Jul 1;179:81–8.
97. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer [Internet]. bioRxiv; 2022 [cited 2024 Apr 16]. p. 2021.10.04.463034. Available from: https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2
98. Wang T, Shao W, Huang Z, Tang H, Zhang J, Ding Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nat Commun. 2021 Jun 8;12(1):3445.
99. Lee BK, Mayhew EJ, Sanchez-Lengeling B, Wei JN, Qian WW, Little KA, et al. A principal odor map unifies diverse tasks in olfactory perception. Science. 2023 Sep;381(6661):999–1006.
100. Ramalingam V, Song Z, Hwang I. The potential role of secondary metabolites in modulating the flavor and taste of the meat. Food Res Int. 2019 Aug 1;122:174–82.
101. Huang Y, Jiang D, Ren G, Yin Y, Sun Y, Liu T, et al. De Novo Production of Glycyrrhetic Acid 3-O-mono-β-D-glucuronide in Saccharomyces cerevisiae. Front Bioeng Biotechnol [Internet]. 2021 [cited 2023 Sep 26];9. Available from: https://www.frontiersin.org/articles/10.3389/fbioe.2021.709120
102. Li B, Wang L, Wu YJ, Xia ZY, Yang BX, Tang YQ. Improving Acetic Acid and Furfural Resistance of Xylose-Fermenting Saccharomyces cerevisiae Strains by Regulating Novel Transcription Factors Revealed via Comparative Transcriptomic Analysis. Appl Environ

Microbiol. 2021 Apr 27;87(10):e00158-21.
103. Chen Y, Li Y, Narayan R, Subramanian A, Xie X. Gene expression inference with deep learning. Bioinformatics. 2016 Jun 15;32(12):1832–9.
104. Li X, Wang K, Lyu Y, Pan H, Zhang J, Stambolian D, et al. Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. Nat Commun. 2020;11(1):2338.
105. Lin E, Mukherjee S, Kannan S. A deep adversarial variational autoencoder model for dimensionality reduction in single-cell RNA sequencing analysis. BMC Bioinformatics. 2020;21(1):1–11.
106. Alvarez M, Rahmani E, Jew B, Garske KM, Miao Z, Benhammou JN, et al. Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM. Sci Rep. 2020;10(1):11019.
107. Messmer T, Dohmen RGJ, Schaeken L, Melzener L, Hueber R, Godec M, et al. Single-cell analysis of bovine muscle-derived cell types for cultured meat production. Front Nutr [Internet]. 2023 Sep 13 [cited 2024 Apr 2];10. Available from: https://www.frontiersin.org/articles/10.3389/fnut.2023.1212196
108. Melzener L, Schaeken L, Fros M, Messmer T, Raina D, Kiessling A, et al. Optimisation of cell fate determination for cultured muscle differentiation [Internet]. bioRxiv; 2024 [cited 2024 Mar 29]. p. 2023.09.06.556523. Available from: https://www.biorxiv.org/content/10.1101/2023.09.06.556523v2
109. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, et al. Mapping transcriptomic vector fields of single cells. Cell. 2022 Feb 17;185(4):690-711.e45.
110. Brendel M, Su C, Bai Z, Zhang H, Elemento O, Wang F. Application of Deep Learning on Single-cell RNA Sequencing Data Analysis: A Review. Genomics Proteomics Bioinformatics [Internet]. 2022; Available from: https://www.sciencedirect.com/science/article/pii/S1672022922001668
111. Petegrosso R, Li Z, Kuang R. Machine learning and statistical methods for clustering single-cell RNA-sequencing data. Brief Bioinform. 2019 Jun;21(4):1209–23.
112. Svensson V, Gayoso A, Yosef N, Pachter L. Interpretable factor models of single-cell RNA-seq via variational autoencoders. Bioinformatics. 2020;36(11):3418–21.
113. Tian T, Wan J, Song Q, Wei Z. Clustering single-cell RNA-seq data with a model-based deep learning approach. Nat Mach Intell. 2019;1(4):191–8.
114. Wang D, Gu J. VASC: dimension reduction and visualization of single-cell RNA-seq data by deep variational autoencoder. Genomics Proteomics Bioinformatics. 2018;16(5):320–31.
115. Alharbi F, Vakanski A. Machine Learning Methods for Cancer Classification Using Gene Expression Data: A Review. Bioengineering. 2023 Feb;10(2):173.
116. Gan Y, Huang X, Zou G, Zhou S, Guan J. Deep structural clustering for single-cell RNA-seq data jointly through autoencoder and graph neural network. Brief Bioinform [Internet]. 2022 Feb;23(2). Available from: https://doi.org/10.1093/bib/bbac018
117. Rao J, Zhou X, Lu Y, Zhao H, Yang Y. Imputing single-cell RNA-seq data by combining graph convolution and autoencoder neural networks. iScience. 2021;24(5):102393.
118. Shan Y, Yang J, Li X, Zhong X, Chang Y. GLAE: A graph-learnable auto-encoder for single-cell RNA-seq analysis. Inf Sci. 2023;621:88–103.
119. Wang J, Ma A, Chang Y, Gong J, Jiang Y, Qi R, et al. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. Nat Commun. 2021;12(1):1882.
120. Wen H, Ding J, Jin W, Wang Y, Xie Y, Tang J. Graph Neural Networks for Multimodal Single-Cell Data Integration. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA:

Association for Computing Machinery; 2022. p. 4153–63. (KDD ’22). Available from: https://doi-org.uml.idm.oclc.org/10.1145/3534678.3539213
121. Aromolaran O, Beder T, Oswald M, Oyelade J, Adebiyi E, Koenig R. Essential gene prediction in Drosophila melanogaster using machine learning approaches based on sequence and functional features. Comput Struct Biotechnol J. 2020;18:612–21.
122. Chuai G, Ma H, Yan J, Chen M, Hong N, Xue D, et al. DeepCRISPR: optimized CRISPR guide RNA design by deep learning. Genome Biol. 2018;19:1–18.
123. He Y, Yuan H, Wu C, Xie Z. DISC: a highly scalable and accurate inference of gene expression and structure for single-cell transcriptomes using semi-supervised deep learning. Genome Biol. 2020;21(1):1–28.
124. Lotfollahi M, Susmelj AK, De Donno C, Ji Y, Ibarra IL, Wolf FA, et al. Learning interpretable cellular responses to complex perturbations in high-throughput screens. bioRxiv [Internet]. 2021; Available from: https://www.biorxiv.org/content/early/2021/05/18/2021.04.14.439903
125. Osorio D, Zhong Y, Li G, Xu Q, Yang Y, Tian Y, et al. scTenifoldKnk: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. Patterns. 2022;3(3):100434.
126. Tian D, Wenlock S, Kabir M, Tzotzos G, Doig AJ, Hentges KE. Identifying mouse developmental essential genes using machine learning. Dis Model Mech. 2018;11(12):dmm034546.
127. Bernstein NJ, Fong NL, Lam I, Roy MA, Hendrickson DG, Kelley DR. Solo: Doublet Identification in Single-Cell RNA-Seq via Semi-Supervised Deep Learning. Cell Syst. 2020;11(1):95-101.e5.
128. Baek S, Lee I. Single-cell ATAC sequencing analysis: From data preprocessing to hypothesis generation. Comput Struct Biotechnol J. 2020 Jun 12;18:1429–39.
129. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017 Sep;14(9):865–8.
130. Huang X, Song C, Zhang G, Li Y, Zhao Y, Zhang Q, et al. scGRN: a comprehensive single-cell gene regulatory network platform of human and mouse. Nucleic Acids Res. 2023 Oct 27;52(D1):D293–303.
131. Kim D, Tran A, Kim HJ, Lin Y, Yang JYH, Yang P. Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. Npj Syst Biol Appl. 2023 Oct 19;9(1):1–13.
132. Melzener L, Ding S, Hueber R, Messmer T, Zhou G, Post MJ, et al. Comparative analysis of cattle breeds as satellite cell donors for cultured beef [Internet]. bioRxiv; 2022 [cited 2024 Apr 17]. p. 2022.01.14.476358. Available from: https://www.biorxiv.org/content/10.1101/2022.01.14.476358v2
133. Meßmer T. Decoding cultured meat production: The transcriptomic landscape of bovine satellite cells in proliferation and differentiation. Maastricht: Maastricht University; 2023.
134. Metzger K, Tuchscherer A, Palin MF, Ponsuksili S, Kalbe C. Establishment and validation of cell pools using primary muscle cells derived from satellite cells of pig skeletal muscle. In Vitro Cell Dev Biol Anim. 2020 Mar;56(3):193–9.
135. Athaya T, Ripan RC, Li X, Hu H. Multimodal deep learning approaches for single-cell multi-omics data integration. Brief Bioinform. 2023 Sep 1;24(5):bbad313.
136. Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L, et al. Best practices for single-cell analysis across modalities. Nat Rev Genet. 2023 Aug;24(8):550–72.
137. Yu D, Cai Z, Li D, Zhang Y, He M, Yang Y, et al. Myogenic Differentiation of Stem

Cells for Skeletal Muscle Regeneration. Stem Cells Int. 2021 Feb 4;2021:8884283.
138. Liongue C, Ward AC. Evolution of Class I cytokine receptors. BMC Evol Biol. 2007 Jul 18;7(1):120.
139. Charoensawan V, Adryan B, Martin S, Söllner C, Thisse B, Thisse C, et al. The Impact of Gene Expression Regulation on Evolution of Extracellular Signaling Pathways. Mol Cell Proteomics MCP. 2010 Dec;9(12):2666–77.
140. Antuofermo E, Orioles M, Murgia C, Burrai GP, Penati M, Gottardi C, et al. Exploring Immunohistochemistry in Fish: Assessment of Antibody Reactivity by Western Immunoblotting. Anim Open Access J MDPI. 2023 Sep 15;13(18):2934.
141. Bennett NR, Watson JL, Ragotte RJ, Borst AJ, See DL, Weidle C, et al. Atomically accurate de novo design of single-domain antibodies [Internet]. bioRxiv; 2024 [cited 2024 Apr 16]. p. 2024.03.14.585103. Available from: https://www.biorxiv.org/content/10.1101/2024.03.14.585103v1
142. Hie BL, Shanker VR, Xu D, Bruun TUJ, Weidenbacher PA, Tang S, et al. Efficient evolution of human antibodies from general protein language models. Nat Biotechnol. 2024 Feb;42(2):275–83.
143. Makowski EK, Kinnunen PC, Huang J, Wu L, Smith MD, Wang T, et al. Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. Nat Commun. 2022 Jul 1;13(1):3788.
144. Gayoso A, Steier Z, Lopez R, Regier J, Nazor KL, Streets A, et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. Nat Methods. 2021 Mar;18(3):272–82.
145. Kim J, McFee M, Fang Q, Abdin O, Kim PM. Computational and artificial intelligence-based methods for antibody development. Trends Pharmacol Sci. 2023 Mar 1;44(3):175–89.
146. Humbird D. Scale‐up economics for cultured meat. Biotechnol Bioeng. 2021 Aug;118(8):3239–50.
147. Stout AJ, Zhang X, Letcher SM, Rittenberg ML, Shub M, Chai KM, et al. Engineered autocrine signaling eliminates muscle cell FGF2 requirements for cultured meat production. Cell Rep Sustain [Internet]. 2024 Jan 26 [cited 2024 Apr 3];1(1). Available from: https://www.cell.com/cell-reports-sustainability/abstract/S2949-7906(23)00009-5
148. Danilevicz MF, Gill M, Tay Fernandez CG, Petereit J, Upadhyaya SR, Batley J, et al. DNABERT-based explainable lncRNA identification in plant genome assemblies. bioRxiv. 2022;2022–02.
149. Levy B, Xu Z, Zhao L, Kremling K, Altman R, Wong P, et al. FloraBERT: cross-species transfer learning withattention-based neural networks for geneexpression prediction. 2022;
150. Wang Y, Wang H, Wei L, Li S, Liu L, Wang X. Synthetic promoter design in Escherichia coli based on a deep generative network. Nucleic Acids Res. 2020 Jul 9;48(12):6403–12.
151. Zrimec J, Fu X, Muhammad AS, Skrekas C, Jauniskis V, Speicher NK, et al. Controlling gene expression with deep generative design of regulatory DNA. Nat Commun. 2022 Aug 30;13(1):5099.
152. Kotopka BJ, Smolke CD. Model-driven generation of artificial yeast promoters. Nat Commun. 2020 Apr 30;11(1):2113.
153. Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. Genome Biol. 2020 Feb 3;21(1):22.
154. Akiyama M, Sakakibara Y. Informative RNA base embedding for RNA structural alignment and clustering by deep representation learning. NAR Genomics Bioinforma. 2022;4(1):lqac012.

155. Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods. 2021;18(10):1196–203.
156. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Training language models to follow instructions with human feedback. Adv Neural Inf Process Syst. 2022 Dec 6;35:27730–44.
157. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal Policy Optimization Algorithms [Internet]. arXiv; 2017 [cited 2024 Jan 24]. Available from: http://arxiv.org/abs/1707.06347
158. O'Neill EN, Cosenza ZA, Baar K, Block DE. Considerations for the development of cost-effective cell culture media for cultivated meat production. Compr Rev Food Sci Food Saf. 2021;20(1):686–709.
159. Simsa R, Yuen J, Stout A, Rubio N, Fogelstrand P, Kaplan DL. Extracellular Heme Proteins Influence Bovine Myosatellite Cell Proliferation and the Color of Cell-Based Meat. Foods. 2019 Oct;8(10):521.
160. Hubalek S, Post MJ, Moutsatsou P. Towards resource-efficient and cost-efficient cultured meat. Curr Opin Food Sci. 2022 Oct 1;47:100885.
161. Messmer T, Klevernic I, Furquim C, Ovchinnikova E, Dogan A, Cruz H, et al. A serum-free media formulation for cultured meat production supports bovine satellite cell differentiation in the absence of serum starvation. Nat Food. 2022 Jan;3(1):74–85.
162. Pasitka L, Cohen M, Ehrlich A, Gildor B, Reuveni E, Ayyash M, et al. Spontaneous immortalization of chicken fibroblasts generates stable, high-yield cell lines for serum-free production of cultured meat. Nat Food. 2023 Jan;4(1):35–50.
163. Ham RG. Clonal growth of mammalian cells in a chemically defined, synthetic medium*. Proc Natl Acad Sci. 1965 Feb;53(2):288–93.
164. Brewer GJ, Torricelli JR, Evege EK, Price PJ. Optimized survival of hippocampal neurons in B27-supplemented neurobasal™, a new serum-free medium combination. J Neurosci Res. 1993;35(5):567–76.
165. Stout AJ, Rittenberg ML, Shub M, Saad MK, Mirliani AB, Kaplan DL. A Beefy-R culture medium: replacing albumin with rapeseed protein isolates [Internet]. bioRxiv; 2022 [cited 2023 Apr 12]. p. 2022.09.02.506409. Available from: https://www.biorxiv.org/content/10.1101/2022.09.02.506409v1
166. Kuo HH, Gao X, DeKeyser JM, Fetterman KA, Pinheiro EA, Weddle CJ, et al. Negligible-Cost and Weekend-Free Chemically Defined Human iPSC Culture. Stem Cell Rep. 2020 Feb 11;14(2):256–70.
167. Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: State of the art. Chem Eng Sci. 2008 Oct 1;63(19):4846–72.
168. Kolkmann AM, Van Essen A, Post MJ, Moutsatsou P. Development of a Chemically Defined Medium for in vitro Expansion of Primary Bovine Satellite Cells. Front Bioeng Biotechnol [Internet]. 2022 [cited 2023 Apr 12];10. Available from: https://www.frontiersin.org/articles/10.3389/fbioe.2022.895289
169. Freddi A, Salmon M. Introduction to the Taguchi Method. In: Freddi A, Salmon M, editors. Design Principles and Methodologies: From Conceptualization to First Prototyping with Examples and Case Studies [Internet]. Cham: Springer International Publishing; 2019 [cited 2023 May 17]. p. 159–80. (Springer Tracts in Mechanical Engineering). Available from: https://doi.org/10.1007/978-3-319-95342-7_7
170. Hashizume T, Ozawa Y, Ying BW. Employing active learning in the optimization of culture medium for mammalian cells [Internet]. bioRxiv; 2022 [cited 2023 May 17]. p. 2022.12.24.521878. Available from:

https://www.biorxiv.org/content/10.1101/2022.12.24.521878v1
171. Gamble C, Bryant D, Carrieri D, Bixby E, Dang J, Marshall J, et al. Machine Learning Optimization of Photosynthetic Microbe Cultivation and Recombinant Protein Production [Internet]. bioRxiv; 2021 [cited 2023 May 17]. p. 2021.08.06.453272. Available from: https://www.biorxiv.org/content/10.1101/2021.08.06.453272v1
172. Kanda GN, Tsuzuki T, Terada M, Sakai N, Motozawa N, Masuda T, et al. Robotic search for optimal cell culture in regenerative medicine. Méndez-Ferrer S, Zaidi M, Méndez-Ferrer S, Sebastian S, Monville C, editors. eLife. 2022 Jun 28;11:e77007.
173. Golovin D, Solnik B, Moitra S, Kochanski G, Karro J, Sculley D. Google Vizier: A Service for Black-Box Optimization. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. [cited 2023 May 17]. Available from: https://dl.acm.org/doi/10.1145/3097983.3098043
174. Balandat M, Karrer B, Jiang DR, Daulton S, Letham B, Wilson AG, et al. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization [Internet]. arXiv; 2020 [cited 2024 Jan 24]. Available from: http://arxiv.org/abs/1910.06403
175. Katoch S, Chauhan SS, Kumar V. A review on genetic algorithm: past, present, and future. Multimed Tools Appl. 2021 Feb 1;80(5):8091–126.
176. Havel J, Link H, Hofinger M, Franco-Lara E, Weuster-Botz D. Comparison of genetic algorithms for experimental multi-objective optimization on the example of medium design for cyanobacteria. Biotechnol J. 2006;1(5):549–55.
177. Delahaye D, Chaimatanan S, Mongeau M. Simulated Annealing: From Basics to Applications. In: Gendreau M, Potvin JY, editors. Handbook of Metaheuristics [Internet]. Cham: Springer International Publishing; 2019 [cited 2023 May 17]. p. 1–35. (International Series in Operations Research & Management Science). Available from: https://doi.org/10.1007/978-3-319-91086-4_1
178. Wijayaningrum VN, Mahmudy WF, Natsir MH. Optimization of Poultry Feed Composition Using Hybrid Adaptive Genetic Algorithm and Simulated Annealing. J Telecommun Electron Comput Eng JTEC. 2017 Sep 1;9(2–8):183–7.
179. Vié A. Qualities, Challenges and Future of Genetic Algorithms [Internet]. Rochester, NY; 2020 [cited 2023 Sep 6]. Available from: https://papers.ssrn.com/abstract=3726035
180. Schinn SM, Morrison C, Wei W, Zhang L, Lewis NE. A genome-scale metabolic network model and machine learning predict amino acid concentrations in Chinese Hamster Ovary cell cultures. Biotechnol Bioeng. 2021;118(5):2118–23.
181. Suchocki T, Wojdak-Maksymiec K, Szyda J. Using gene networks to identify genes and pathways involved in milk production traits in Polish Holstein dairy cattle. Czech J Anim Sci. 2016 Nov 30;61(11):526–38.
182. Taussat S, Boussaha M, Ramayo-Caldas Y, Martin P, Venot E, Cantalapiedra-Hijar G, et al. Gene networks for three feed efficiency criteria reveal shared and specific biological processes. Genet Sel Evol. 2020 Nov 10;52(1):67.
183. Huynh-Thu VA, Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. Sci Rep. 2018 Feb 21;8(1):3384.
184. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, et al. Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe. Cell Syst. 2020 Mar 25;10(3):265-274.e11.
185. Kim J, T. Jakobsen S, Natarajan KN, Won KJ. TENET: gene network reconstruction using transfer entropy reveals key regulatory factors from single cell transcriptomic data. Nucleic Acids Res. 2021 Jan 11;49(1):e1.
186. Liu Z, Gao J, Li T, Jing Y, Xu C, Zhu Z, et al. A novel approach GRNTSTE to reconstruct gene regulatory interactions applied to a case study for rat pineal rhythm gene. Sci

Rep. 2022 Jun 17;12(1):10227.
187. Hale B, Ratnayake S, Flory A, Wijeratne R, Schmidt C, Robertson AE, et al. Gene regulatory network inference in soybean upon infection by Phytophthora sojae [Internet]. bioRxiv; 2022 [cited 2023 May 17]. p. 2022.10.19.512983. Available from: https://www.biorxiv.org/content/10.1101/2022.10.19.512983v2
188. Huang J, Zheng J, Yuan H, McGinnis K. Distinct tissue-specific transcriptional regulation revealed by gene regulatory networks in maize. BMC Plant Biol. 2018 Jun 7;18(1):111.
189. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010 Mar;28(3):245–8.
190. Ravindra N, Sehanobish A, Pappalardo JL, Hafler DA, van Dijk D. Disease State Prediction from Single-Cell Data Using Graph Attention Networks. In: Proceedings of the ACM Conference on Health, Inference, and Learning [Internet]. New York, NY, USA: Association for Computing Machinery; 2020. p. 121–30. (CHIL '20). Available from: https://doi.org/10.1145/3368555.3384449
191. Vinodkumar PK, Ozcinar C, Anbarjafari G. Prediction of SGRNA off-target activity in crispr/cas9 gene editing using graph convolution network. Entropy. 2021;23(5):608.
192. Hagrot E, Oddsdóttir HÆ, Hosta JG, Jacobsen EW, Chotteau V. Poly-pathway model, a novel approach to simulate multiple metabolic states by reaction network-based model – Application to amino acid depletion in CHO cell culture. J Biotechnol. 2017 Oct 10;259:235–47.
193. Hong JK, Yeo HC, Lakshmanan M, Han S hyuk, Cha HM, Han M, et al. In silico model-based characterization of metabolic response to harsh sparging stress in fed-batch CHO cell cultures. J Biotechnol. 2020 Jan 20;308:10–20.
194. Martínez-Monge I, Albiol J, Lecina M, Liste-Calleja L, Miret J, Solà C, et al. Metabolic flux balance analysis during lactate and glucose concomitant consumption in HEK293 cell cultures. Biotechnol Bioeng. 2019;116(2):388–404.
195. Széliová D, Ruckerbauer DE, Galleguillos SN, Petersen LB, Natter K, Hanscho M, et al. What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. Metab Eng. 2020 Sep 1;61:288–300.
196. Specht L. An analysis of culture medium costs and production volumes for cultivated meat. Good Food Institute; 2020 Feb.
197. Venkatesan M, Semper C, Skrivergaard S, Di Leo R, Mesa N, Rasmussen MK, et al. Recombinant production of growth factors for application in cell culture. iScience. 2022 Oct 21;25(10):105054.
198. Hamby SE, Hirst JD. Prediction of glycosylation sites using random forests. BMC Bioinformatics. 2008 Nov 27;9(1):500.
199. Pakhrin SC, Aoki-Kinoshita KF, Caragea D, Kc DB. DeepNGlyPred: A Deep Neural Network-Based Approach for Human N-Linked Glycosylation Site Prediction. Molecules. 2021 Jan;26(23):7314.
200. Ching-Hsuan C, Chi-Chang C, Shih-Huan L, Chi-Wei C, Zong-Han C, Yen-Wei C. N-GlycoGo: Predicting Protein N-Glycosylation Sites on Imbalanced Data Sets by Using Heterogeneous and Comprehensive Strategy. IEEE Access. 2020;8:165944–50.
201. Jung F, Frey K, Zimmer D, Mühlhaus T. DeepSTABp: A Deep Learning Approach for the Prediction of Thermal Protein Stability. Int J Mol Sci. 2023 Jan;24(8):7444.
202. Dubourg-Felonneau G, Abbasi A, Akiva E, Lee L. Improving Protein Subcellular Localization Prediction with Structural Prediction & Graph Neural Networks [Internet]. bioRxiv; 2022 [cited 2024 Jan 24]. p. 2022.11.29.518403. Available from: https://www.biorxiv.org/content/10.1101/2022.11.29.518403v1

203. Pak MA, Markhieva KA, Novikova MS, Petrov DS, Vorobyev IS, Maksimova ES, et al. Using AlphaFold to predict the impact of single mutations on protein stability and function. PLOS ONE. 2023 Mar 16;18(3):e0282689.
204. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022 Jun;19(6):679–82.
205. Goldenzweig A, Fleishman SJ. Principles of Protein Stability and Their Application in Computational Design. Annu Rev Biochem. 2018 Jun 20;87:105–29.
206. Dvorak P, Bednar D, Vanacek P, Balek L, Eiselleova L, Stepankova V, et al. Computer-assisted engineering of hyperstable fibroblast growth factor 2. Biotechnol Bioeng. 2018;115(4):850–62.
207. Voorhamme D, Yandell CA. LONG R3IGF-I as a more potent alternative to insulin in serum-free culture of HEK293 cells. Mol Biotechnol. 2006 Oct;34(2):201–4.
208. Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models [Internet]. bioRxiv; 2022 [cited 2023 Sep 24]. p. 2022.12.09.519842. Available from: https://www.biorxiv.org/content/10.1101/2022.12.09.519842v2
209. Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, et al. Robust deep learning–based protein sequence design using ProteinMPNN. Science. 2022 Oct 7;378(6615):49–56.
210. Saito Y, Oikawa M, Sato T, Nakazawa H, Ito T, Kameda T, et al. Machine-Learning-Guided Library Design Cycle for Directed Evolution of Enzymes: The Effects of Training Data Composition on Sequence Space Exploration. ACS Catal. 2021 Dec 3;11(23):14615–24.
211. Okamoto Y, Haraguchi Y, Yoshida A, Takahashi H, Yamanaka K, Sawamura N, et al. Proliferation and differentiation of primary bovine myoblasts using Chlorella vulgaris extract for sustainable production of cultured meat. Biotechnol Prog. 2022;38(3):e3239.
212. Jeong Y, Choi WY, Park A, Lee YJ, Lee Y, Park GH, et al. Marine cyanobacterium Spirulina maxima as an alternate to the animal cell culture medium supplement. Sci Rep. 2021 Mar 1;11(1):4906.
213. Mj B, Fo U, Cc U. Influence of Cowpea (Vigna unguiculata) Peptides on Insulin Resistance. J Nutr Health Food Sci [Internet]. 2015 May 5 [cited 2023 May 17];3(2). Available from: https://symbiosisonlinepublishing.com/nutritionalhealth-foodscience/
214. Venâncio TM, Oliveira AEA, Silva LB, Machado OLT, Fernandes KVS, Xavier-Filho J. A protein with amino acid sequence homology to bovine insulin is present in the legume Vigna unguiculata (cowpea). Braz J Med Biol Res. 2003 Sep;36:1167–73.
215. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.
216. Holm L, Laakso LM. Dali server update. Nucleic Acids Res. 2016 Jul 8;44(W1):W351–5.
217. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nat Biotechnol. 2023 May 8;1–4.
218. Veliça P, Bunce CM. A quick, simple and unbiased method to quantify C2C12 myogenic differentiation. Muscle Nerve. 2011;44(3):366–70.
219. Furuhashi M, Morimoto Y, Shima A, Nakamura F, Ishikawa H, Takeuchi S. Formation of contractile 3D bovine muscle tissue for construction of millimetre-thick cultured steak. Npj Sci Food. 2021 Mar 2;5(1):6.
220. Ribeiro AJS, Schwab O, Mandegar MA, Ang YS, Conklin BR, Srivastava D, et al. Multi-Imaging Method to Assay the Contractile Mechanical Output of Micropatterned Human

iPSC-Derived Cardiac Myocytes. Circ Res. 2017 May 12;120(10):1572–83.
221. Al-Kofahi Y, Zaltsman A, Graves R, Marshall W, Rusu M. A deep learning-based algorithm for 2-D cell segmentation in microscopy images. BMC Bioinformatics. 2018 Oct 3;19(1):365.
222. Durkee MS, Abraham R, Clark MR, Giger ML. Artificial Intelligence and Cellular Segmentation in Tissue Microscopy Images. Am J Pathol. 2021 Oct 1;191(10):1693–701.
223. Kumar N, Verma R, Anand D, Zhou Y, Onder OF, Tsougenis E, et al. A Multi-Organ Nucleus Segmentation Challenge. IEEE Trans Med Imaging. 2020 May;39(5):1380–91.
224. Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. Nat Methods. 2022 Dec;19(12):1634–41.
225. Verma R, Kumar N, Patil A, Kurian NC, Rane S, Graham S, et al. MoNuSAC2020: A Multi-Organ Nuclei Segmentation and Classification Challenge. IEEE Trans Med Imaging. 2021 Dec;40(12):3413–23.
226. Zinchuk V, Grossenbacher-Zinchuk O. Machine Learning for Analysis of Microscopy Images: A Practical Guide and Latest Trends. Curr Protoc. 2023;3(7):e819.
227. Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nat Methods. 2019 Dec;16(12):1247–53.
228. Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat Biotechnol. 2022 Apr;40(4):555–65.
229. Kaimal J, Thul P, Xu H, Ouyang W, Lundberg E. HPA Cell Image Segmentation Dataset [Internet]. Zenodo; 2021 [cited 2023 Apr 25]. Available from: https://zenodo.org/record/4665863
230. Naylor P, Laé M, Reyal F, Walter T. Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map. IEEE Trans Med Imaging. 2019 Feb;38(2):448–59.
231. Mohammed EA, Mohamed MMA, Naugler C, Far BH. Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding. In: 2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). 2013. p. 1–5.
232. Gómez-de-Mariscal E, García-López-de-Haro C, Ouyang W, Donati L, Lundberg E, Unser M, et al. DeepImageJ: A user-friendly environment to run deep learning models in ImageJ. Nat Methods. 2021 Oct;18(10):1192–5.
233. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation [Internet]. arXiv; 2015 [cited 2023 Apr 25]. Available from: http://arxiv.org/abs/1411.4038
234. Wienert S, Heim D, Saeger K, Stenzinger A, Beil M, Hufnagl P, et al. Detection and Segmentation of Cell Nuclei in Virtual Microscopy Images: A Minimum-Model Approach. Sci Rep. 2012 Jul 11;2(1):503.
235. Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun. 2016 Aug 16;7(1):12474.
236. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-Net: deep learning for cell counting, detection, and morphometry. Nat Methods. 2019 Jan;16(1):67–70.
237. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [Internet]. arXiv; 2015 [cited 2023 Apr 25]. Available from: http://arxiv.org/abs/1505.04597
238. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation [Internet]. arXiv; 2018 [cited 2023 Apr 25].

Available from: http://arxiv.org/abs/1807.10165
239. Franklin MM, Schultz FA, Tafoya MA, Kerwin AA, Broehm CJ, Fischer EG, et al. A Deep Learning Convolutional Neural Network Can Differentiate Between Helicobacter Pylori Gastritis and Autoimmune Gastritis With Results Comparable to Gastrointestinal Pathologists. Arch Pathol Lab Med. 2021 Apr 15;146(1):117–22.
240. Hekler A, Utikal JS, Enk AH, Solass W, Schmitt M, Klode J, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. Eur J Cancer. 2019 Sep 1;118:91–6.
241. Dursun G, Bijelić D, Ayşit N, Vatandaşlar BK, Radenović L, Çapar A, et al. Combined segmentation and classification-based approach to automated analysis of biomedical signals obtained from calcium imaging. PLOS ONE. 2023 Feb 6;18(2):e0281236.
242. Huynh HT, Dat VVT, Anh HB. White Blood Cell Segmentation and Classification Using Deep Learning Coupled with Image Processing Technique. In: Dang TK, Küng J, Chung TM, Takizawa M, editors. Future Data and Security Engineering Big Data, Security and Privacy, Smart City and Industry 40 Applications. Singapore: Springer; 2021. p. 399–410. (Communications in Computer and Information Science).
243. Belashov AV, Zhikhoreva AA, Belyaeva TN, Salova AV, Kornilova ES, Semenova IV, et al. Machine Learning Assisted Classification of Cell Lines and Cell States on Quantitative Phase Images. Cells. 2021 Oct;10(10):2587.
244. Winfree S. User-Accessible Machine Learning Approaches for Cell Segmentation and Analysis in Tissue. Front Physiol [Internet]. 2022 [cited 2023 Sep 6];13. Available from: https://www.frontiersin.org/articles/10.3389/fphys.2022.833333
245. Lou W, Yu X, Liu C, Wan X, Li G, Liu S, et al. Multi-stream Cell Segmentation with Low-level Cues for Multi-modality Images. 2022;
246. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In 2017 [cited 2024 Jan 25]. p. 2961–9. Available from: https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html
247. Ghojogh B, Ghodsi A. Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey [Internet]. arXiv; 2023 [cited 2024 Jan 25]. Available from: http://arxiv.org/abs/2304.11461
248. Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, et al. Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis. IEEE Trans Med Imaging. 2022 Apr;41(4):757–70.
249. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological Image Analysis: A Review. IEEE Rev Biomed Eng. 2009;2:147–71.
250. Kumar N, Verma R, Chen C, Lu C, Fu P, Willis J, et al. Computer-extracted features of nuclear morphology in hematoxylin and eosin images distinguish stage II and IV colon tumors. J Pathol. 2022;257(1):17–28.
251. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition [Internet]. arXiv; 2015 [cited 2024 Jan 25]. Available from: http://arxiv.org/abs/1512.03385
252. Marzahl C, Aubreville M, Voigt J, Maier A. Classification of Leukemic B-Lymphoblast Cells from Blood Smear Microscopic Images with an Attention-Based Deep Learning Method and Advanced Augmentation Techniques. In: Gupta A, Gupta R, editors. ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging. Singapore: Springer; 2019. p. 13–22. (Lecture Notes in Bioengineering).
253. Zhou T, Reji R, Kairon RS, Chiam KH. A review of algorithmic approaches for cell culture media optimization. Front Bioeng Biotechnol [Internet]. 2023 [cited 2024 Jan 25];11. Available from: https://www.frontiersin.org/articles/10.3389/fbioe.2023.1195294

254. Grzesik P, Warth SC. One-Time Optimization of Advanced T Cell Culture Media Using a Machine Learning Pipeline. Front Bioeng Biotechnol [Internet]. 2021 [cited 2024 Jan 25];9. Available from: https://www.frontiersin.org/articles/10.3389/fbioe.2021.614324
255. Jang J, Wang C, Zhang X, Choi HJ, Pan X, Lin B, et al. A deep learning-based segmentation pipeline for profiling cellular morphodynamics using multiple types of live cell microscopy. Cell Rep Methods. 2021 Nov 22;1(7):100105.
256. Race AM, Sutton D, Hamm G, Maglennon G, Morton JP, Strittmatter N, et al. Deep Learning-Based Annotation Transfer between Molecular Imaging Modalities: An Automated Workflow for Multimodal Data Integration. Anal Chem. 2021 Feb 16;93(6):3061–71.
257. Allan SJ, De Bank PA, Ellis MJ. Bioprocess Design Considerations for Cultured Meat Production With a Focus on the Expansion Bioreactor. Front Sustain Food Syst [Internet]. 2019 [cited 2023 Apr 26];3. Available from: https://www.frontiersin.org/articles/10.3389/fsufs.2019.00044
258. Stephens N, Di Silvio L, Dunsford I, Ellis M, Glencross A, Sexton A. Bringing cultured meat to market: Technical, socio-political, and regulatory challenges in cellular agriculture. Trends Food Sci Technol. 2018 Aug 1;78:155–66.
259. Specht EA, Welch DR, Rees Clayton EM, Lagally CD. Opportunities for applying biomedical production and manufacturing methods to the development of the clean meat industry. Biochem Eng J. 2018 Apr 15;132:161–8.
260. Synoground BF, McGraw CE, Elliott KS, Leuze C, Roth JR, Harcum SW, et al. Transient ammonia stress on Chinese hamster ovary (CHO) cells yield alterations to alanine metabolism and IgG glycosylation profiles. Biotechnol J. 2021;16(7):2100098.
261. Sarna S, Patel N, Corbett B, McCready C, Mhaskar P. Process-aware data-driven modelling and model predictive control of bioreactor for the production of monoclonal antibodies. Can J Chem Eng. 2023;101(5):2677–92.
262. Kiparissides A, Pistikopoulos EN, Mantalaris A. On the model-based optimization of secreting mammalian cell (GS-NS0) cultures. Biotechnol Bioeng. 2015;112(3):536–48.
263. Bogaerts Ph, Mhallem Gziri K, Richelle A. From MFA to FBA: Defining linear constraints accounting for overflow metabolism in a macroscopic FBA-based dynamical model of cell cultures in bioreactor. J Process Control. 2017 Dec 1;60:34–47.
264. Craven S, Whelan J, Glennon B. Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. J Process Control. 2014 Apr 1;24(4):344–57.
265. Zavala-Ortiz DA, Denner A, Aguilar-Uscanga MG, Marc A, Ebel B, Guedon E. Comparison of partial least square, artificial neural network, and support vector regressions for real-time monitoring of CHO cell culture processes using in situ near-infrared spectroscopy. Biotechnol Bioeng. 2022;119(2):535–49.
266. Vaitkus V, Brazauskas K, Repšytė J. Soft-sensors based on Black-box Models for Bioreactors Monitoring and State Estimation. In: Proceedings of the 2020 12th International Conference on Bioinformatics and Biomedical Technology [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2023 Aug 14]. p. 43–7. (ICBBT 2020). Available from: https://doi.org/10.1145/3405758.3405780
267. Zhang W, Li J, Liu T, Leng S, Yang L, Peng H, et al. Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae. Bioresour Technol. 2021 Dec 1;342:126011.
268. Oh TH, Park HM, Kim JW, Lee JM. Integration of reinforcement learning and model predictive control to optimize semi-batch bioreactor. AIChE J. 2022;68(6):e17658.
269. Petsagkourakis P, Sandoval IO, Bradford E, Zhang D, del Rio-Chanona EA. Reinforcement learning for batch bioprocess optimization. Comput Chem Eng. 2020 Feb

2;133:106649.
270. Luna M, Martínez E. A Bayesian Approach to Run-to-Run Optimization of Animal Cell Bioreactors Using Probabilistic Tendency Models. Ind Eng Chem Res. 2014 Nov 5;53(44):17252–66.
271. Conev A, Litsa EE, Perez MR, Diba M, Mikos AG, Kavraki LE. Machine Learning-Guided Three-Dimensional Printing of Tissue Engineering Scaffolds. Tissue Eng Part A [Internet]. 2020 Dec 16 [cited 2023 Sep 25]; Available from: https://www.liebertpub.com/doi/10.1089/ten.tea.2020.0191
272. Bone JM, Childs CM, Menon A, Póczos B, Feinberg AW, LeDuc PR, et al. Hierarchical Machine Learning for High-Fidelity 3D Printed Biopolymers. ACS Biomater Sci Eng. 2020 Dec 14;6(12):7021–31.
273. Li F, Han J, Cao T, Lam W, Fan B, Tang W, et al. Design of self-assembly dipeptide hydrogels and machine learning via their chemical features. Proc Natl Acad Sci. 2019 Jun 4;116(23):11259–64.
274. Nishimura T. The role of intramuscular connective tissue in meat texture. Anim Sci J. 2010;81(1):21–7.
275. Cerchiari AE, Garbe JC, Jee NY, Todhunter ME, Broaders KE, Peehl DM, et al. A strategy for tissue self-organization that is robust to cellular heterogeneity and plasticity. Proc Natl Acad Sci. 2015 Feb 17;112(7):2287–92.
276. Libby ARG, Briers D, Haghighi I, Joy DA, Conklin BR, Belta C, et al. Automated Design of Pluripotent Stem Cell Self-Organization. Cell Syst. 2019 Nov 27;9(5):483-495.e10.
277. Wang M, Tsanas A, Blin G, Robertson D. Predicting pattern formation in embryonic stem cells using a minimalist, agent-based probabilistic model. Sci Rep. 2020 Oct 1;10(1):16209.
278. Tian L, Chen F, Macosko EZ. The expanding vistas of spatial transcriptomics. Nat Biotechnol. 2023 Jun;41(6):773–82.
279. Fischer DS, Schaar AC, Theis FJ. Modeling intercellular communication in tissues using spatial graphs of cells. Nat Biotechnol. 2023 Mar;41(3):332–6.
280. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021 Nov;18(11):1342–51.
281. Tanevski J, Flores ROR, Gabor A, Schapiro D, Saez-Rodriguez J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. Genome Biol. 2022 Apr 14;23(1):97.
282. Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nat Methods. 2021 Nov;18(11):1352–62.
283. Koerich DM, Lopes GC, Rosa LM. Investigation of phases interactions and modification of drag models for liquid-solid fluidized bed tapered bioreactors. Powder Technol. 2018 Nov 1;339:90–101.
284. Ouyang Y, Zou HK, Gao XY, Chu GW, Xiang Y, Chen JF. Computational fluid dynamics modeling of viscous liquid flow characteristics and end effect in rotating packed bed. Chem Eng Process - Process Intensif. 2018 Jan 1;123:185–94.
285. Peng J, Sun W, Zhou G, Xie L, Han H, Xiao Y. The accurate prediction and analysis of bed expansion characteristics in liquid–solid fluidized bed based on machine learning methods. Chem Eng Sci. 2022 Oct 12;260:117841.
286. Guo S, Ao X, Ma X, Cheng S, Men C, Harada H, et al. Machine-learning-aided application of high-gravity technology to enhance ammonia recovery of fresh waste leachate. Water Res. 2023 May 15;235:119891.

287. Bi K, Zhang D, Qiu T, Huang Y. GC-MS Fingerprints Profiling Using Machine Learning Models for Food Flavor Prediction. Processes. 2020 Jan;8(1):23.
288. Zhu W, Benkwitz F, Kilmartin PA. Volatile-Based Prediction of Sauvignon Blanc Quality Gradings with Static Headspace–Gas Chromatography–Ion Mobility Spectrometry (SHS–GC–IMS) and Interpretable Machine Learning Techniques. J Agric Food Chem. 2021 Mar 17;69(10):3255–65.
289. Gonzalez Viejo C, Tongson E, Fuentes S. Integrating a Low-Cost Electronic Nose and Machine Learning Modelling to Assess Coffee Aroma Profile and Intensity. Sensors. 2021 Jan;21(6):2016.
290. Gonzalez Viejo C, Fuentes S, Godbole A, Widdicombe B, Unnithan RR. Development of a low-cost e-nose to assess aroma profiles: An artificial intelligence application to assess beer quality. Sens Actuators B Chem. 2020 Apr 1;308:127688.
291. Tian H, Liu H, He Y, Chen B, Xiao L, Fei Y, et al. Combined application of electronic nose analysis and back-propagation neural network and random forest models for assessing yogurt flavor acceptability. J Food Meas Charact. 2020 Feb 1;14(1):573–83.
292. Khan MI, Jo C, Tariq MR. Meat flavor precursors and factors influencing flavor precursors—A systematic review. Meat Sci. 2015 Dec 1;110:278–84.
293. Bouysset C, Belloir C, Antonczak S, Briand L, Fiorucci S. Novel scaffold of natural compound eliciting sweet taste revealed by machine learning. Food Chem. 2020 Sep 15;324:126864.
294. Wiltschko AB. Learning to Smell: Using Deep Learning to Predict the Olfactory Properties of Molecules [Internet]. 2019 [cited 2024 Feb 21]. Available from: https://blog.research.google/2019/10/learning-to-smell-using-deep-learning.html
295. Lee J, Song SB, Chung YK, Jang JH, Huh J. BoostSweet: Learning molecular perceptual representations of sweeteners. Food Chem. 2022 Jul 30;383:132435.
296. Tuwani R, Wadhwa S, Bagler G. BitterSweet: Building machine learning models for predicting the bitter and sweet taste of small molecules. Sci Rep. 2019 May 9;9(1):7155.
297. Wang YT, Yang ZX, Piao ZH, Xu XJ, Yu JH, Zhang YH. Prediction of flavor and retention index for compounds in beer depending on molecular structure using a machine learning method. RSC Adv. 2021 Nov 10;11(58):36942–50.
298. Ji H, Pu D, Yan W, Zhang Q, Zuo M, Zhang Y. Recent advances and application of machine learning in food flavor prediction and regulation. Trends Food Sci Technol. 2023 Aug 1;138:738–51.
299. Aljaafreh A. Agitation and mixing processes automation using current sensing and reinforcement learning. J Food Eng. 2017 Jun 1;203:53–7.
300. Bi S, Zhang B, Mu L, Ding X, Wang J. Optimization of tobacco drying process control based on reinforcement learning. Dry Technol. 2020 Jul 1;38(10):1291–9.
301. Kircali Ata S, Shi JK, Yao X, Hua XY, Haldar S, Chiang JH, et al. Predicting the Textural Properties of Plant-Based Meat Analogs with Machine Learning. Foods. 2023 Jan;12(2):344.
302. Liu X, Shen Q, Zhang S. Cross-species cell-type assignment from single-cell RNA-seq data by a heterogeneous graph neural network. Genome Res. 2023 Jan 1;33(1):96–111.
303. Wang YM, Sun Y, Wang B, Wu Z, He XY, Zhao Y. Transfer learning for clustering single-cell RNA-seq data crossing-species and batch, case on uterine fibroids. Brief Bioinform. 2024 Jan 1;25(1):bbad426.
304. Park Y, Muttray NP, Hauschild AC. Species-agnostic transfer learning for cross-species transcriptomics data integration without gene orthology. Brief Bioinform. 2024 Mar 1;25(2):bbae004.

305. Cui H, Wang C, Maan H, Pang K, Luo F, Duan N, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods. 2024 Feb 26;1–11.
306. Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, et al. Transfer learning enables predictions in network biology. Nature. 2023 Jun;618(7965):616–24.
307. Nissa MU, Reddy PJ, Pinto N, Sun Z, Ghosh B, Moritz RL, et al. The PeptideAtlas of a widely cultivated fish Labeo rohita: A resource for the Aquaculture Community. Sci Data. 2022 Apr 13;9(1):171.
308. Grisel O, Mueller A, Lars, Gramfort A, Louppe G, Fan TJ, et al. scikit-learn/scikit-learn: Scikit-learn 1.4.0 [Internet]. Zenodo; 2024 [cited 2024 Feb 6]. Available from: https://zenodo.org/records/10532824