nature machine intelligence

REVIEW ARTICLE

https://doi.org/10.1038/s42256-020-0217-y

Check for updates

Ensemble deep learning in bioinformatics

Yue Cao[1,2], Thomas Andrew Geddes[2,3], Jean Yee Hwa Yang[1,2] and Pengyi Yang[1,2,4] ✉

The remarkable flexibility and adaptability of ensemble methods and deep learning models have led to the proliferation of their application in bioinformatics research. Traditionally, these two machine learning techniques have largely been treated as independent methodologies in bioinformatics applications. However, the recent emergence of ensemble deep learning—wherein the two machine learning techniques are combined to achieve synergistic improvements in model accuracy, stability and reproducibility—has prompted a new wave of research and application. Here, we share recent key developments in ensemble deep learning and look at how their contribution has benefited a wide range of bioinformatics research from basic sequence analysis to systems biology. While the application of ensemble deep learning in bioinformatics is diverse and multifaceted, we identify and discuss the common challenges and opportunities in the context of bioinformatics research. We hope this Review Article will bring together the broader community of machine learning researchers, bioinformaticians and biologists to foster future research and development in ensemble deep learning, and inspire novel bioinformatics applications that are unattainable by traditional methods.

Bioinformatics, an interdisciplinary field of research, is at the centre of modern molecular biology, where computational methods are developed and utilized to transform biological data into knowledge and translate them for biomedical applications. Among the various computational methods utilized in bioinformatics research, machine learning, a branch of artificial intelligence characterized by data-driven model building, has been the key enabling computational technology[1]. At the forefront of machine learning, ensemble learning and deep learning have independently made a substantial impact on the field of bioinformatics through their widespread applications, from basic nucleotide and protein sequence analysis to systems biology[2,3].

Until recently, ensemble and deep learning models have largely been treated as independent methodologies in bioinformatics applications. The fast-growing synergy between these two popular techniques, however, has attracted a new wave of development and application of next-generation machine learning methods referred to as ensemble deep learning (Fig. 1a). The root of ensemble deep learning can be traced back two decades, when ensembles of neural networks were found to reduce generalization error[4]. However, the recent resurgence of ensemble deep learning models has brought about new ideas, algorithms, frameworks and architectures that substantially enrich the old paradigm. Through its novel application to a wide range of biological and biomedical research, ensemble deep learning is unleashing its power in dealing with key challenges, including small sample size, high-dimensionality, imbalanced class distribution, and noisy and heterogeneous data generated from diverse cellular and biological systems using an array of high-throughput omics technologies. These computational, methodological and technological undertakings and breakthroughs together are leading a phenomenal transformation of bioinformatics.

Both ensemble learning and deep learning methods have been extensively studied and reviewed in the context of bioinformatics applications[5,6]. However, the emergence of ensemble deep learning and its application in bioinformatics has yet to be documented. With the aim of providing a reference point to foster research in the increasingly popular field of ensemble deep learning and its application to various challenges in bioinformatics, in this Review Article we revisit the foundation of ensemble and deep learning, and summarize and categorize the latest developments in ensemble deep learning. This is followed by a survey of ensemble deep learning applications in bioinformatics. We then discuss the remaining challenges and opportunities that we hope will inspire future research and development across multiple disciplines.

Basics of ensemble and deep learning

Ensemble learning refers to a class of strategies where instead of building a single model, multiple 'base' models are combined to perform tasks such as supervised and unsupervised learning[7]. Classic ensemble methods for supervised learning fall into three categories: bagging-, boosting- and stacking-based methods. In bagging[8], individual base models are trained on subsets of data sampled randomly with replacement (Fig. 1b). In boosting[9], models are trained sequentially (Fig. 1c), where subsequent models focus on previous misclassified samples. In stacking, a meta-learner is trained to optimally combine the predictions made by base models[10]. Like supervised ensemble learning, conventional unsupervised ensemble learning, such as ensemble clustering[11], also relies on the generation and integration of base models (Fig. 1d). While their variants, including more advanced methods reviewed in the next section, have also been used in ensemble learning, a guiding principle in designing ensemble methods has been 'many heads are better than one'[12].

Deep learning, a branch of machine learning, is rooted in artificial neural networks[13]. The most fundamental architecture of deep learning models is the densely connected neural network (DNN), consisting of a series of layers of neurons; each of these is connected to all neurons in the previous layer[14]. More sophisticated models expand on the basic architectures. In convolutional neural networks (CNNs)[15], each layer comprises a series of filters that 'slide over' the output of the previous layer to extract local features across different parts of the input. In recurrent neural networks (RNNs)[16], circuits

[1]School of Mathematics and Statistics, University of Sydney, Sydney, New South Wales, Australia. [2]Charles Perkins Centre, University of Sydney, Sydney, New South Wales, Australia. [3]School of Environmental and Life Sciences, University of Sydney, Sydney, New South Wales, Australia. [4]Computational Systems Biology Group, Children's Medical Research Institute, University of Sydney, Westmead, New South Wales, Australia. ✉e-mail: pengyi.yang@sydney.edu.au

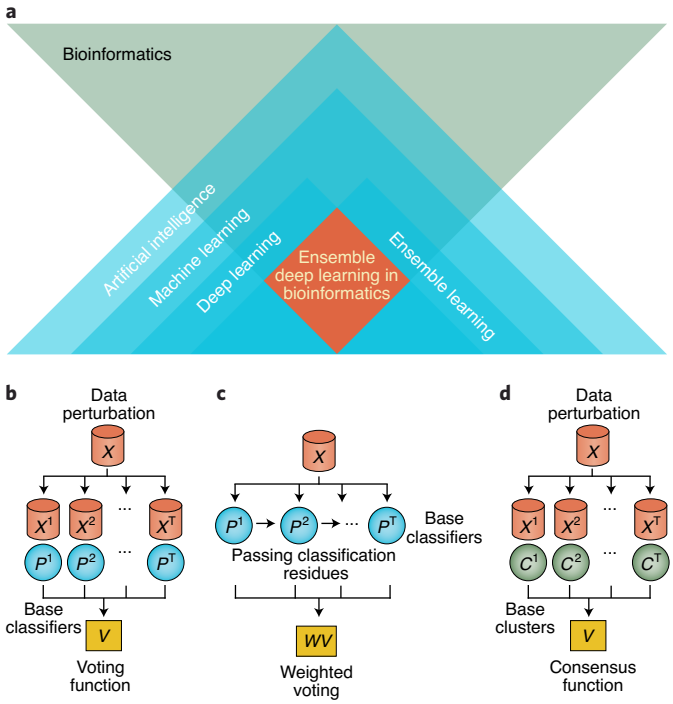

Fig. 1 | The focus of this Review Article and classic ensemble methods. **a**, Relationships of artificial intelligence, machine learning, deep learning, ensemble learning and bioinformatics. The red square denotes the focal point of this Review Article. **b**–**d**, Classic ensemble learning frameworks including bagging and its variants (**b**), boosting and its variants (**c**), and ensemble clustering based on data perturbation (**d**). *X*, input data.

are created to feed the output of a layer back into the same layer along with new input, allowing the model to act on dependencies between upstream and downstream values in a sequence. Variants of RNNs have been proposed to enable more effective learning in long-term dependency tasks, with the two most common ones being long short-term memory (LSTM)[17] and gated recurrent unit (GRU)[18]. In residual neural networks (ResNet)[19], shortcuts between upstream and downstream layers are introduced to improve the effectiveness of backpropagation in networks with many hidden layers. In autoencoders[20], networks are constructed with an encoder and a decoder that together learn a more efficient latent space representation of the original higher-dimensional data. Although the difference between traditional neural networks and deep learning may seem elusive, the latter is increasingly defined by its unique architectures and ability to learn complex data representations that are beyond the capacity of classic models[21].

Ensemble deep learning

Deep learning is well known for its power to approximate almost any function and increasingly demonstrates predictive accuracy that surpasses human experts. However, deep learning models are not without shortcomings: they often exhibit high variance and may fall into local loss minima during training. Indeed, empirical results of ensemble methods that combine the output of multiple deep learning models have been shown to achieve better generalizability than a single model[22]. In addition to simple ensemble approaches such as averaging output from individual models, combining heterogeneous models enables multifaceted abstraction of data, and may lead to better learning outcomes[23]. In this section, we categorize and summarize the most representative ensemble deep learning strategies for both supervised and unsupervised tasks.

Supervised ensemble deep learning. In this section, we summarize the key ensemble deep learning frameworks for supervised tasks.

Ensemble across multiple models. The aggregation of multiple and often independent deep learning models is the most straightforward application of ensemble deep learning to classify (Fig. 2a). As diversity of individual networks is an essential characteristic of a good ensemble model[24], a variety of strategies exist to promote diversity of base networks. One approach is to encourage negative correlation in the classification error of base models[25]. The key motivation behind promoting negative correlation among base models is to encourage complementary learning of the training data to achieve better generalizability of the ensemble. An alternative approach to increase base model diversity is through multiple choice learning in which each network is 'specialized' on a particular subset of data during the training step[26].

An issue associated with training and storing multiple models is the computational and storage demand involved. To address this, methods that perform knowledge distillation have become increasingly popular[27]. One such implementation is based on the concept of a teacher–student network framework, where the teacher networks are selected from a pool of pre-trained networks and the student network distils knowledge of multiple teachers into a single and often simpler network[28,29]. The testing phase is storage and computationally efficient, as the samples only need to pass through a single student network.

Ensemble within a single model. Ensemble strategies described in the previous section require training of multiple models. Deep learning models are often computationally costly to train and may take days or even weeks depending on the scale of the dataset and model. Effort has been made to develop 'implicit ensembles' where a single neural network could achieve an effect similar to integrating multiple network models. To this end, a group of techniques focuses on random deactivation of neurons and layers during the training process of a single model. This leads to an implicit ensemble of networks with different architectures (Fig. 2b). For example, the random deactivation of neurons, termed dropout, originally proposed as a regularization strategy[30] for addressing model overfitting is now widely known as an implicit ensemble strategy[31,32]. This has inspired follow-up works on random deactivation of building blocks, termed ResBlocks, in ResNets[33] and the combined random deactivation of neurons and layers[34]. Besides random deactivation-based methods, alternative strategies have also been explored. One popular approach is the snapshot ensemble technique, where the key idea is to save multiple versions of a single model during the training process for forming an ensemble[35]. In a snapshot ensemble, a cyclic learning rate scheduler is utilized, where the learning rate is abruptly changed every few epochs to perturb the network and thus may lead to diversity in the snapshots of the model.

Ensemble with model branching. Single-model ensemble approaches greatly reduce training cost compared to ensembles of multiple models. However, such a reduction in computational demand comes potentially at a cost in base model diversity. Since the information captured by the lower layers of neural networks is likely to be similar across models, a group of techniques has emerged with a focus on sharing lower layers followed by 'branching' of additional layers[36]. These model branching approaches introduce diversity while also enjoying the reduction of time and computation of training multiple models (Fig. 2c). Besides reducing computational cost, model branching has also been adapted to address other challenges in training an ensemble. For example, the gradient can be propagated over a shorter path in a branching network, mitigating the vanishing gradient problem[37]. In the knowledge distillation frame-

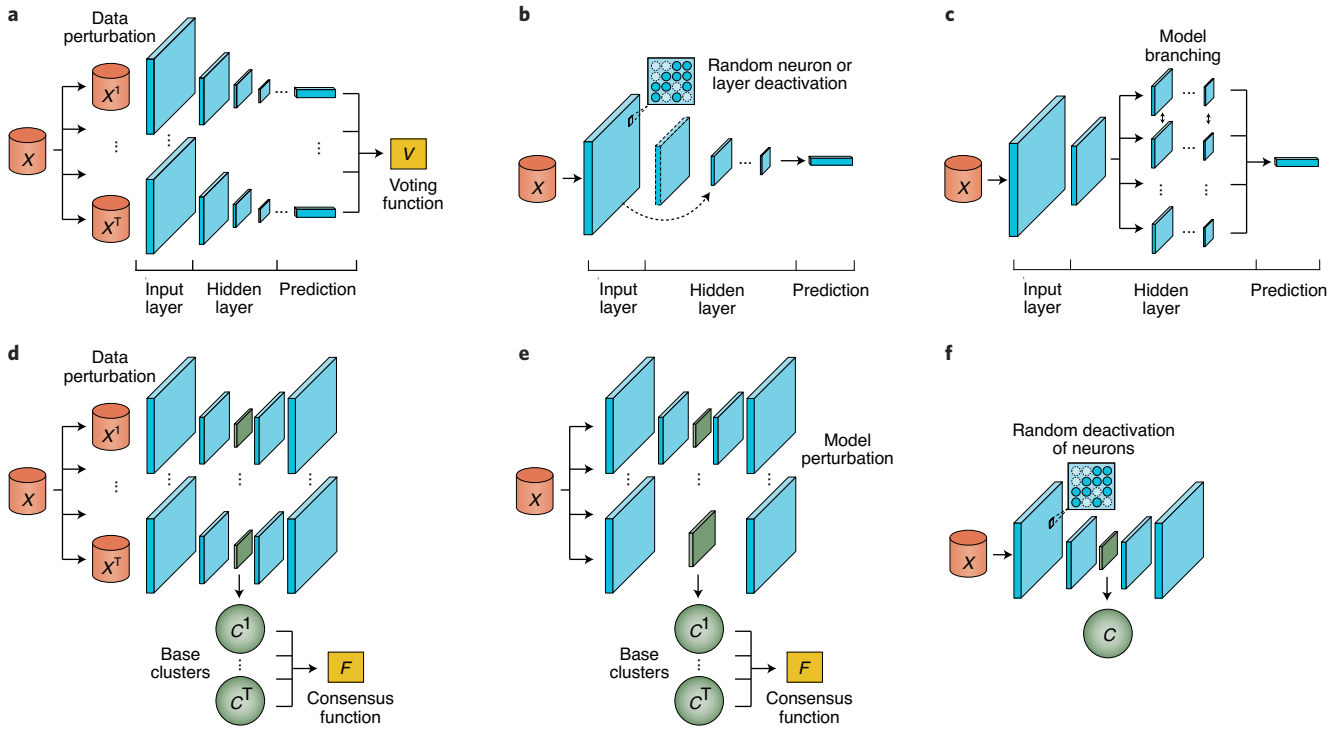

Fig. 2 | Typical ensemble deep learning frameworks in supervised and unsupervised learning. a, Ensemble across multiple models. Each neural network is trained separately on the dataset, usually perturbed to allow the network to learn from diverse training samples. **b**, Ensemble within a single model. Common strategies for creating intrinsic variants of the network include randomly deactivating and bypassing layers (indicated by the curved arrow) and randomly deactivating neurons (indicated by the close-up). **c**, Ensemble by model branching. Common strategies include sharing lower layers and branching out to learn different higher-level features with or without weight sharing. **d**, Unsupervised ensemble by data perturbation. Each autoencoder is trained with a perturbed dataset such as bootstrapping. The latent representations are extracted for clustering and combined through a consensus function. **e**, Model perturbation-based unsupervised ensemble. Multiple autoencoders each with a different model architecture can be used to learn diverse representation of the original data. **f**, Unsupervised ensemble within a single model. Similar to the supervised case, random deactivation of neurons can be used to create intrinsic variants of the network.

work, each branch acts as a student model, ensembled to form a teacher model on the fly to reduce the computationally intensive process of pre-training the teacher model[38]. The key commonality between these model branching network ensembles is that by sharing information, the base networks avoid parameter search from scratch and can converge faster.

Unsupervised ensemble deep learning. In this section, we summarize the key ensemble deep learning frameworks for unsupervised tasks.

Ensemble across multiple models. Most unsupervised ensemble deep learning methods employ autoencoders, a popular unsupervised network architecture. Similar to the supervised approach, unsupervised ensemble methods can be categorized into those that generate and combine multiple models through data and model perturbation, and those that achieve implicit ensemble within a single model.

For methods based on data perturbation, strategies akin to bagging in supervised learning are widely used (Fig. 2d). For example, Geddes et al. used random feature projection of the input data to train a set of autoencoders to create a cluster ensemble[39]. Training a series of unsupervised networks with different hyper-parameters is a common ensemble strategy for methods based on model perturbation (Fig. 2e). An example extending this approach uses different activation functions and a weighting scheme to improve model accuracy[40]. An alternative to data and model perturbation is to use multi-view clustering when such data are available. Representative examples include multi-view representation learning using deep canonically correlated autoencoders[41] and multi-view spectral clustering where multiple embedding networks were used to represent the original data from different feature sets[42].

Ensemble within a single model. The power of autoencoders in data dimension reduction has motivated research around creating better data representations that are robust to noise in the input data. For example, a denoising autoencoder architecture was introduced in ref. [43], where values of a random subset of neurons are masked (that is, changed to zero) during each training epoch, forcing the network to overcome noise introduced to the data. The concept of randomly masking neurons in denoising autoencoders is analogous to the dropout method used in the supervised approach, and hence can be considered as an implicit ensemble within a single model, or 'pseudo-ensemble'[44], for unsupervised deep learning (Fig. 2f). In this line of research, a recent study exploits the flexibility of the dropout algorithm and embeds it in a more advanced variational autoencoder architecture[45]. The proposed algorithm employs a novel strategy to learn the dropout parameter, thus alleviating the need for manual tuning. Another extension in this direction is the 'stacked' denoising autoencoder that uses multiple layers of denoising autoencoders for improving data representation[46]. The data representation learned from such stacked denoising autoencoders led to substantially improved classification accuracy compared with using raw input data.

Theoretical advances for ensemble deep learning. While early works on the bias–variance trade-off framework have laid the theoretical foundation for neural network ensembles[47], recent research on ensemble deep learning mostly relies on empirical experiments due to the increasingly specialized ensemble methodologies and complex neural network architectures. Nevertheless, efforts have been made to advance the theoretical foundation of this fast-growing field[48]. Studies have shown the existence of multiple local minima in training neural networks, where some enjoy better generalizability than others[49]. This has inspired ensemble techniques such as snapshot methods that take advantage of the diversity of multiple local minima[35]. Theoretical justification for dropout as a form of averaging has been discussed in ref. [31], where the expectation of the gradient with dropout was shown to be the gradient of the regularized ensemble error. A recent mathematical framework provided a new perspective of dropout by relating it to a form of data augmentation[50].

Bioinformatics applications of ensemble deep learning

This section categorizes representative works in different areas of bioinformatics applications (Table 1) and identifies their benefits, such as improving model accuracy, reproducibility, interpretability and model inference.

Sequence analysis. Biological sequence analysis represents one of the fundamental applications of computational methods in molecular biology. RNN and its variants (for example, LSTM and GRU) are well-suited to sequential data. For example, an LSTM/CNN multi-model was trained to extract distinct features to predict pathogenic potential of DNA sequences[51]. Compared to DNA sequences, RNA sequences offer an additional layer of information where instructions encoded in genes are transcribed. While traditional methods rely on various manually curated RNA sequence features, ensemble deep learning enables automatic learning from raw data. One example is in predicting localization of long non-coding RNAs, where multiple sub-networks were used to integrate distinct feature sets to maximize model performance[52]. In another work, a CNN/RNN ensemble was used to integrate features and raw sequence data to predict different types of translation initiation sites[53], overcoming the generalizability issue of traditional methods that can only predict a specific type of translational initiation sites.

Following transcription, messenger RNAs (mRNAs) are further translated into proteins that carry out various functions. Similar to RNA sequence analysis, methods relying on ensembles of multiple sub-networks were used to integrate information from multiple features sets to predict DNA binding sites[54] and post-translational modification (PTM) sites[55] on protein sequences. The study on PTM site prediction has further demonstrated that features learned by ensemble models are 'transferable' for predicting different types of PTMs, a key property for tackling the issue of small sample size in training data.

Genome analysis. While sequence analysis has led to many biological discoveries, alone it cannot capture the full repertoire of information encoded in the genome. Additional layers of genetic information including structural variants[56] (for example, copy number variations (CNVs)) and epigenetic modifications[57] of the genome bring important insight to the understanding of biological systems, populations and complex diseases.

A number of ensemble deep learning methods have been developed on this front, such as classifying cancer types using CNV data and a snapshot ensemble model comprising CNNs, LSTMs and convolutional autoencoders[58]. The use of supervised CNN and LSTM models allows both global and local sequential features to be captured, and further integration with unsupervised convolutional autoencoders enables unsupervised pre-training, an effective component for handling small sample size[59]. Beyond combining different network architectures, studies have also integrated different genomic data modalities to capture distinct and complementary information. In one study, DNA sequences and their neighbouring cytosine–guanine dinucleotide (CpG) states were used as input into two sub-networks of an ensemble to explore their relationship in predicting DNA methylation states[60]. This has led to the identification of sequence motifs related to DNA methylation and the effect of their mutation on CpG methylation. In another study, an ensemble network that takes input data either from DNA sequences alone or with the addition of epigenetic information extracted from chromatin immunoprecipitation (ChIP) and deoxyribonuclease (DNase) sequencing were used to predict human immunodeficiency virus type 1 (HIV-1) integration sites[61]. The ensemble network, comprising CNNs with attention layers[62], enabled the discovery of DNA sequence motifs that are important for HIV-1 integration.

Gene expression. Gene expression data including microarray, RNA-sequencing (RNA-seq) and, recently, single-cell RNA-seq (scRNA-seq)[63–65], has been studied extensively to better understand complex diseases and to identify biomarkers that can guide therapeutic decision making. A recent study on cancer type classification demonstrated how ensemble deep learning can serve as a potential strategy to address the key challenge of reproducibility in biomarker research[66]. The use of a DNN ensemble in this work allowed the derivation of important genes through consensus ranking across multiple models, resulting in a robust set of biomarkers. Due to the difficulty of obtaining patient samples, especially for rare diseases and cancer types, another common challenge in analysing gene expression data from cancers and diseases is the small sample size. The use of ensemble learning to mitigate this issue is exemplified by ref. [67], where the authors applied a multi-model approach to generate initial predictions from RNA-seq gene expression profiles of cancer samples and integrated these predictions using a DNN to produce the final ensemble prediction.

In addition to its role in medical research, ensemble deep learning has been used in a wide range of applications to improve understanding of basic biological mechanisms from gene expression data. An example is the use of a DNN ensemble to explore the embryonic to fetal transition process, a defining stage where cells lose the potential for regeneration[68]. A benefit of training multiple networks is that the prediction scores from each network can be further used to generate an integrative score to determine the transition state of a sample between the embryonic and adult state, a strategy that is not possible with a single model. The utility of unsupervised ensemble deep learning has also been demonstrated on the extraction of biological pathway signatures[69]. By integrating signatures across 100 autoencoders through consensus clustering, the ensemble model detected more biological pathways with higher significance than a single model. Unsupervised deep learning ensembles have also been applied to cell type identification in single-cell research. In ref. [39], an ensemble of autoencoders was used to generate a diverse set of latent representations of scRNA-seq data for subsequent analysis.

Structural bioinformatics. Proteins are the key products of genes, and their functions and mechanisms are largely governed by protein structures encoded in amino acid sequences. Therefore, modelling and characterizing proteins from their primary amino acid sequences to secondary and tertiary structures is essential for understanding and predicting their functions[70]. RNN and its architectural variants are specifically designed to capture long- and short-range interactions between sequences, and are hence well-suited to decoding the relationship between amino acid sequences and the protein structures they encode. Extending on the use of a single RNN, the ensemble of variants of RNNs with CNNs is a common hybrid architecture in recent applications that seeks to combine the

Table 1 | Categorization of recent ensemble deep learning methods in bioinformatics application

Type of learning	Ensemble technique	Deep learning architecture	Sequence analysis	Genome analysis	Gene expression	Structural bio-informatics	Proteomics	Systems biology	Multi-omics	Bioimage informatics
Supervised	Multiple models	DNN			Refs. [66–68]		Ref. [80]	Ref. [83]	Ref. [93]	Ref. [98]
		CNN	Ref. [54]	Ref. [61]		Ref. [75]		Refs. [82,85]		
		CNN + RNN	Refs. [51,53]	Ref. [60]		Refs. [71,72]	Ref. [79]	Ref. [84]	Ref. [90]	
		CNN + RNN + ResNet	Ref. [55]			Refs. [73,74,76]				
		Others	Ref. [52]	Ref. [58]						Ref. [95]
	Within single model	CNN + RNN		Ref. [58]						
	Model branching	CNN								Refs. [96,97]
		CNN + ResNet								Ref. [94]
Unsupervised	Multiple models	Autoencoder			Refs. [39,69]				Ref. [91,92]	
		Others							Ref. [89]	
	Within single model	Autoencoder		Ref. [58]						

power of RNN in analysing sequential data and CNN on extracting local features[71,72]. The replacement of CNN with ResNet[73] as well as the addition of residual connections between GRU and CNN[74] were also explored to facilitate feature propagation for improved modelling of long-range dependencies between amino acids. In these works, ensemble deep learning not only improved generalizability on independent datasets but also led to the discovery of novel features associated with protein structures.

Besides predicting protein structures, many studies have focused on directly predicting protein functions. An example of ensemble deep learning application in this domain is illustrated by the work of Zacharaki[75], who used an ensemble of CNNs for protein enzymatic function prediction. Specifically, the ensemble is a fusion of two CNNs trained separately on protein properties and amino acid features for extracting complementary information. In another example, Singh et al.[76] built an ensemble deep learning model to identify residue conformation crucial to protein folding and function. While the dataset used for model training has an extreme class imbalance (1.4:1,000), the ensemble model, consisting of ResNet and LSTM modules, yielded robust performance on independent test sets without manual generation of a balanced dataset.

Proteomics. While protein structure and function prediction are essential tasks for characterizing individual proteins, technological advances in quantitative mass spectrometry (MS) have now enabled global profiling of the entire proteome in cells, tissues and species[77]. Computational analysis of such large volume datasets is transforming our understanding of proteome dynamics in complex systems and diseases[78].

Ensemble deep learning has been used as a key technique for addressing various aspects of proteomics data analysis. The work of Zohora et al.[79] exemplifies the application of ensemble deep learning to peptide identification from a liquid chromatography-MS (LC-MS) map, a critical step for identifying and quantifying protein abundance. Specifically, a hybrid network architecture comprising both CNN and RNN modules was designed to detect sequential features along the axes during the scan of an MS map. The final model, an ensemble of multiple networks with different parameters, was shown to achieve state-of-the-art results for protein quantification. Another study proposed an ensemble of DNNs for learning from data-independent acquisition (DIA) MS data[80]. While conventional MS runs select only a few important peptides based on their signal levels (that is, data-dependent acquisition) for subsequent quantification, the DIA approach fragments every single peptide for improved proteome coverage. However, the DIA approach may lead to an increase in co-eluted peptides and therefore higher interference in the data. The ensemble framework was able to quantify the amount of interference between multiple peptides mapped to the same point, thereby removing interference and improving peptide identification confidence and quantification accuracy.

Systems biology. Systems biology aims to map interactions of molecule species, regulatory relationships and mechanisms to understand complex biological systems as a whole[81]. One key aspect of systems biology is the understanding of what and how biological molecules interact. In recent times, ensemble deep learning has been applied on this front to predict interactions among different biological molecules and entities. The application of an interpretable ensemble of CNN models for predicting binding affinity between peptides and major histocompatibility complex is an example of ensemble deep learning in this domain[82] and has important implication in clinics. The model demonstrated good generalizability across 30 independent datasets and uncovered binding motifs with literature support. In predicting protein–protein interactions, an ensemble of DNNs trained on *S. cerevisiae* achieved more accurate results than other machine learning methods[83]. Subsequently, the model was applied to other datasets generated from different organisms and the relative accuracy on each dataset was shown to be a good indicator of the evolutionary relationships of those organisms.

Systems biology also extends to the interaction between biological molecules and chemical compounds. In particular, the study of protein and chemical compound interaction in drug development has seen a growing number of ensemble deep learning applications. For example, Karimi et al. proposed an ensemble model that comprised various network modules for compound–protein affinity prediction[84]. To overcome the limited availability of labelled datasets, the model exploited abundant unlabelled compound and protein data through unsupervised pre-training. This was followed by interaction prediction on labelled data using CNN and RNN modules in the ensemble. In another work on predicting drug and

target protein interactions, a CNN-based ensemble model was used to score the likelihood of interaction of randomly selected drug–protein pairs[85]. The trained model revealed that drugs with similar structures bind to similar target proteins, suggesting potential similarity in the effects of these drugs.

Multi-omics. Multi-omics analysis is a topic closely related to systems biology, where integrative methods are used to understand biological regulation by combining an array of omics data. There is a growing interest in multi-omic studies as it is increasingly recognized that a single type of omics data does not capture the entire landscape of the complex biological networks[86].

Many conventional machine learning methods have been proposed to utilize the complementary information present across multiple modalities of omics data[87,88]. Most conventional approaches, however, do not account for the relationships among different omics layers. To this end, Liang et al. proposed to use an ensemble of deep belief networks to encode gene expression, miRNA expression and DNA methylation data into multiple layers of hidden variables for integrative clustering[89], thereby actively exploring regulation across different omics layers. Ensembles of different deep learning architectures have also been utilized to take advantage of the unique characteristics of different data types. Using an ensemble of CNNs and LSTMs, both genomic sequences and their secondary structures can now be integrated for alternative polyadenylation site prediction on pre-mRNAs[90]. This addressed the gap where existing models overlooked RNA secondary structures, despite these being important features to the polyadenylation process. Another application in multi-omics was the use of a novel ensemble of autoencoders wherein a coupling cost was used to encourage the base autoencoders to learn from each other[91]. This unsupervised model allowed the integration of two vastly different data types—single-cell transcriptomics and electrophysiological profiles—and to identify common and unique cell types across datasets.

High dimensionality and heterogeneity are both issues associated with the large number of molecular features in multi-omics datasets. The application of autoencoders is popular in dealing with these challenges. In one instance, an ensemble of autoencoders was used to extract lower dimension and integrate over 450,000 features in pan-cancer classification[92]. Stacking multiple deep learning models, each handling a different modality of omics data[93], is another approach that avoids feature concatenation that might otherwise exacerbate the issue of high dimensionality in datasets potentially containing tens of thousands of features.

Bioimage informatics. Traditionally, analysis of bioimages is often performed manually by field experts. With the growing number of computer vision applications demonstrating their superior performance over human experts, automatic analysis has become an increasing focus in bioinformatics studies. A primary application of ensemble deep learning in bioimage informatics is the detection of diseases such as cancers in patient images. For instance, to improve classification of glioma from magnetic resonance images, Lu et al. embedded a branching module into ResNet for integrating multi-scale information obtained from different receptive fields of the original ResNet[94]. Codella et al. proposed an ensemble model that combined network architectures, including ResNet, CNN and U-Net, to segment and classify skin lesions from dermoscopic images[95]. It is noteworthy that the proposed model achieved a segmentation result with 95% accuracy, surpassing that of human experts who exhibit an accuracy of around 91%. To segment cervical cell images, Song et al. performed multi-resolution extraction and colour space transformation of the images to generate diverse feature sets, leading to enhanced segmentation accuracy[96].

Besides improving classification and segmentation accuracy, ensemble deep learning methods have also been explored in addressing various other challenges in bioimage analysis. For example, an ensemble network with knowledge distillation and a branching strategy was used to reduce the number of parameters in the model and therefore lower the likelihood of overfitting on small datasets[97]. To deal with the problem of class imbalance, Yuan et al.[98] introduced an iterative regularization approach that, for a given iteration, penalizes misclassification of samples that were correctly classified in previous iterations. This method alleviated the problem of bias in favour of majority classes and preserved correctly classified minority examples.

Challenges and opportunities

The applications we have reviewed here reveal various challenges and opportunities surrounding ensemble deep learning in bioinformatics research. In the following sections, we highlight several key directions in which ensemble deep learning is likely to have increasingly important impacts.

Small sample size. Deep learning is known for its exceptional performance on data with large sample size. While modern omics technologies have enabled the profiling of tens of thousands of molecular species and biological events in a single experiment, the number of samples available is usually small owing to the cost in time and labour. Hence, bioinformatics applications are often confronted with the issue of limited sample size, causing unstable predictions and thus low reproducibility in results.

Fortunately, one essential property of ensemble methods is stability. Leveraging this key property, a number of ensemble deep learning methods were proposed to specifically address small sample size challenges, opening up the opportunity to utilize deep learning in bioinformatics. While the most popular approach so far has been using pre-trained models, more specialized methods have also been explored. Examples include extracting intermediate features learned by the network to generate additional output for integration and thus stabilizing the ensemble prediction[99]; and encouraging cooperation among individual models through a pairwise loss, thereby reducing the variance caused by small sample size[100]. These methods represent promising strategies that can be explored in future lines of research.

High-dimensionality and class imbalance. Omics data are well-known for their high-dimensionality, as biological features (for example, genes, proteins) frequently outnumber samples. This is further exacerbated by the issue of small sample size already mentioned. The problem, widely known as the 'curse of dimensionality', has been identified as one of the main causes of overfitting in deep learning models due to the large number of parameters that needs to be fitted[101]. While deep learning models seem to be particularly susceptible to the high-dimensionality of omics data, the combination of deep learning with ensemble methods such as model averaging[39] and the implicit ensemble through dropout[30] has been demonstrated to be an effective approach for handling this issue.

Imbalanced class distribution is another common issue in many bioinformatics applications[102] where, for example, a biological event of interest is only present in a small proportion of the data. Ensemble deep learning is found to be an effective remedy for dealing with this challenge. Bioinformatics applications include the use of bootstrap-sampling- and random-sampling-based ensemble deep learning for dealing with class imbalance in DNA and protein sequence analyses[53,54]. Due to the increasing use of high-throughput technologies, ensemble deep learning strategies that are capable of dealing with these challenges will remain an active research direction in bioinformatics.

Data noise and heterogeneity. Biological systems are inherently heterogeneous and noisy. This is further confounded by technical

noise from various sources including experimental protocol and omics platform. A key characteristic of ensemble methods is their robustness to data noise[103], which can facilitate the reproducible extraction of biological signals from noisy and heterogeneous data. The application of methods such as denoising autoencoders also strengthens model robustness[43]. The integration of ensemble and deep learning methods therefore provides an opportunity to address noise and heterogeneity in biological data.

The development of multi-omics technologies further contributed to heterogeneity within datasets in that different molecular species measured across omics platforms must be combined and analysed integratively to understand biological systems holistically. Ensemble deep learning methods such as multi-model approaches reviewed previously have been demonstrated to be highly effective in combining different omics data for joint inference[89] and classification[90]. Given these intrinsic properties of data generated from biological systems, we expect ensemble deep learning methods to play an increasingly important role in omics data analysis and in integrating large-scale multi-omics data.

Model interpretability. A common criticism of deep learning models is their lack of interpretability. Besides building an accurate model, gaining insight from the model is also critical in bioinformatics applications, since having an interpretable model of a biological system may lead to testable hypotheses that can be validated through experiments.

Several studies reviewed in previous sections have already made notable progress in this direction. For example, attention layers in ensemble networks were used to identify motifs of HIV integration sites[61] and drug binding sites[84]. The stability and reproducibility offered by ensemble methods such as in feature selection[104] are also making a substantial impact in biomarker discovery[105]. This is evident from the application of ensemble deep learning methods to identifying molecular markers for the diagnosis of primary and metastatic cancers[66] and to provide insights into normal development and cancers[68]. As we move from predictive to preventive biomedical research, models that offer biological insight into data will become increasingly desirable.

Choice of network architecture. The choice of network architecture is crucial for achieving optimal performance in a specific domain and application. For example, many studies choose to employ variants of the RNN such as the LSTM, which is suitable for learning sequential information in biological sequences[53,72]. DNN and CNN architectures, on the other hand, are shown to be suitable for biological applications that handle high-dimensional input[61,66].

The use of multi-model ensembles makes it possible to exploit the power of hybrid architectures or to combine heterogeneous data types in multi-omics. Examples reviewed include the ResNet/RNN hybrid used to capture the relationship between each layer of features in RNA secondary structure prediction[73], and the CNN/LSTM hybrid to learn both RNA sequences and secondary structures for joint prediction of alternative polyadenylation sites on pre-mRNAs[90]. While these studies demonstrate the importance and the application of specialized network architectures in bioinformatics, the exponential growth of new network architectures proposed in the computer science literature is likely to lead to many more novel applications in bioinformatics in the coming years.

Computational expense. Deep learning models typically contain large numbers of parameters and the computational burden of generating an ensemble of multiple deep learning models could be extremely high especially when working with large-scale omics data. Nevertheless, recent developments in ensemble deep learning have made use of the modularity of deep learning architectures and provided a panel of ensemble strategies and algorithms to enable more efficient model fitting with a substantial reduction in training time. The improvement of computer hardware and technological advances in computing methods such as distributed and federated deep learning[106,107] also facilitate the application and deployment of ensemble deep learning on large-scale omics data. Given that the size and complexity of biological data are only expected to soar as technology progresses, the development of more efficient ensemble deep learning algorithms and architectures will be another crucial direction in both machine learning and bioinformatics research.

Future outlook

While the ensemble of neural networks has existed long before the deep learning era, the recent development of ensemble deep learning has substantially enriched the field with novel architectures and ensemble strategies that greatly improve model accuracy, reliability and efficiency. These innovations, together with properties such as robustness to small sample size, high-dimensionality and data noise, have transformed ensemble deep learning into a new force, leading to remarkable and widespread breakthroughs across different fields of bioinformatics applications. Nonetheless, many of the advanced ensemble techniques that harness the power of recent deep learning architectures remain under-explored in their application to bioinformatics. In addition, the development and application of models that enable interpretation of biological systems are still in their infancy. We hope this Review Article has sparked thoughts on ensemble deep learning across multiple disciplines, and will inspire future research and applications that embraces the myriad of ensemble deep learning strategies to revolutionize biological and biomedical research.

Received: 21 March 2020; Accepted: 14 July 2020;
Published online: 17 August 2020

References

1. Larranaga, P. et al. Machine learning in bioinformatics. *Briefings Bioinform.* **7**, 86–112 (2006).
2. Eraslan, G., Avsec, Ž., Gagneur, J. & Theis, F. J. Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* **20**, 389–403 (2019).
3. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-generation machine learning for biological networks. *Cell* **173**, 1581–1592 (2018).
4. Hansen, L. K. & Salamon, P. Neural network ensembles. *IEEE Trans. Pattern Anal. Mach.* **12**, 993–1001 (1990).
5. Yang, P., Hwa Yang, Y., Zhou, B. B. & Zomaya, A. Y. A review of ensemble methods in bioinformatics. *Curr. Bioinform.* **5**, 296–308 (2010).
6. Min, S., Lee, B. & Yoon, S. Deep learning in bioinformatics. *Briefings Bioinform.* **18**, 851–869 (2017).
7. Dietterich, T. G. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems* 1–15 (Springer, 2000).
8. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
9. Schapire, R. E., Freund, Y., Bartlett, P. & Lee, W. S. Boosting the margin: a new explanation for the effectiveness of voting methods. *Ann. Stat.* **26**, 1651–1686 (1998).
10. Wolpert, D. H. Stacked generalization. *Neural Netw.* **5**, 241–259 (1992).
11. Vega-Pons, S. & Ruiz-Shulcloper, J. A survey of clustering ensemble algorithms. *Int. J. Pattern Recogn.* **25**, 337–372 (2011).
12. Altman, N. & Krzywinski, M. Points of significance: ensemble methods: bagging and random forests. *Nat. Methods* **14**, 933–935 (2017).
13. Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015).
14. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
15. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Proc. 26th Int. Conf. Advances in Neural Information Processing Systems* 1097–1105 (NIPS, 2012).
16. Williams, R. J. & Zipser, D. A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.* **1**, 270–280 (1989).
17. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).

18. Cho, K. et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proc. 2014 Conf. Empirical Methods in Natural Language Processing* 1724–1734 (EMNLP, 2014).
19. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
20. Baldi, P. Autoencoders, unsupervised learning, and deep architectures. In *Proc. ICML Workshop on Unsupervised and Transfer learning* 37–49 (ICML, 2012).
21. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
22. Ju, C., Bibaut, A. & van der Laan, M. The relative performance of ensemble methods with deep convolutional neural networks for image classification. *J. Appl. Stat.* **45**, 2800–2818 (2018).
23. Lee, S., Purushwalkam, S., Cogswell, M., Crandall, D. & Batra, D. Why *M* heads are better than one: training a diverse ensemble of deep networks. Preprint at https://arxiv.org/abs/1511.06314 (2015).
24. Granitto, P. M., Verdes, P. F. & Ceccatto, H. A. Neural network ensembles: evaluation of aggregation algorithms. *Artif. Intell.* **163**, 139–162 (2005).
25. Liu, Y. & Yao, X. Ensemble learning via negative correlation. *Neural Netw.* **12**, 1399–1404 (1999).
26. Lee, S. et al. Stochastic multiple choice learning for training diverse deep ensembles. In *Proc. 30th Int. Conf. Advances in Neural Information Processing Systems* 2119–2127 (NIPS, 2016).
27. Hinton, G., Vinyals, O. & Dean, J. Distilling the knowledge in a neural network. Preprint at http://arxiv.org/abs/1503.02531 (2015).
28. Shen, Z., He, Z. & Xue, X. Meal: multi-model ensemble via adversarial learning. In *Proc. AAAI Conf. Artificial Intelligence* Vol. 33 4886–4893 (AAAI, 2019).
29. Parisotto, E., Ba, J. & Salakhutdinov, R. Actor-mimic: deep multitask and transfer reinforcement learning. In *Proc. Int. Conf. Learning Representations* (ICLR, 2016).
30. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
31. Baldi, P. & Sadowski, P. J. Understanding dropout. In *Proc. 27th Int. Conf. Advances in Neural Information Processing Systems* 2814–2822 (NIPS, 2013).
32. Hara, K., Saitoh, D. & Shouno, H. Analysis of dropout learning regarded as ensemble learning. In *Proc. 25th Int. Conf. Artificial Neural Networks* 72–79 (ICANN, 2016).
33. Huang, G., Sun, Y., Liu, Z., Sedra, D. & Weinberger, K. Q. Deep networks with stochastic depth. In *14th European Conf. Computer Vision* 646–661 (Springer, 2016).
34. Singh, S., Hoiem, D. & Forsyth, D. Swapout: learning an ensemble of deep architectures. In *Proc. 30th Int. Conf. Advances in Neural Information Processing Systems* 28–36 (NIPS, 2016).
35. Huang, G. et al. Snapshot ensembles: train 1, get *M* for free. Preprint at https://arxiv.org/abs/1704.00109 (2017).
36. Han, B., Sim, J. & Adam, H. Branchout: regularization for online ensemble tracking with convolutional neural networks. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition* 3356–3365 (IEEE, 2017).
37. Wang, X., Bao, A., Cheng, Y. & Yu, Q. Multipath ensemble convolutional neural network. *IEEE Trans. Emerg. Topics Comput.* https://doi.org/10.1109/TETCI.2018.2877154 (2018).
38. Zhu, X., Gong, S. et al. Knowledge distillation by on-the-fly native ensemble. In *Proc. 32nd Int. Conf. Advances in Neural Information Processing Systems* 7517–7527 (NIPS, 2018).
39. Geddes, T. A. et al. Autoencoder-based cluster ensembles for single-cell RNA-seq data analysis. *BMC Bioinform.* **20**, 660 (2019).
40. Shao, H., Jiang, H., Lin, Y. & Li, X. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mech. Syst. Signal Process.* **102**, 278–297 (2018).
41. Wang, W., Arora, R., Livescu, K. & Bilmes, J. On deep multi-view representation learning. In *Proc. 32nd Int. Conf. International Conference on Machine Learning* 1083–1092 (ICML, 2015).
42. Huang, Z. et al. Multi-view spectral clustering network. In *Proc. 28th Int. Joint Conf. Artificial Intelligence* 2563–2569 (IJCAI, 2019).
43. Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. In *Proc. 25th Int. Conf. Machine Learning* 1096–1103 (ICML, 2008).
44. Bachman, P., Alsharif, O. & Precup, D. Learning with pseudo-ensembles. In *Proc. 28th Int. Conf. Advances in Neural Information Processing Systems* 3365–3373 (NIPS, 2014).
45. Antelmi, L., Ayache, N., Robert, P. & Lorenzi, M. Sparse multi-channel variational autoencoder for the joint analysis of heterogeneous data. In *Proc. 36th Int. Conf. Machine Learning* 302–311 (ICML, 2019).
46. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. & Manzagol, P.-A. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11**, 3371–3408 (2010).
47. Geman, S., Bienenstock, E. & Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **4**, 1–58 (1992).
48. Bengio, Y. Learning deep architectures for AI. *Found. Trends Mach. Learn.* **2**, 1–127 (2009).
49. Keskar, N. S., Nocedal, J., Tang, P. T. P., Mudigere, D. & Smelyanskiy, M. On large-batch training for deep learning: generalization gap and sharp minima. In *Proc. 5th Int. Conf. Learning Representations* (ICLR, 2017).
50. Zhao, D., Yu, G., Xu, P. & Luo, M. Equivalence between dropout and data augmentation: a mathematical check. *Neural Netw.* **115**, 82–89 (2019).
51. Bartoszewicz, J. M., Seidel, A., Rentzsch, R. & Renard, B. Y. Deepac: predicting pathogenic potential of novel DNA with reverse-complement neural networks. *Bioinformatics* **36**, 81–89 (2020).
52. Cao, Z., Pan, X., Yang, Y., Huang, Y. & Shen, H.-B. The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics* **34**, 2185–2194 (2018).
53. Zhang, S., Hu, H., Jiang, T., Zhang, L. & Zeng, J. TITER: predicting translation initiation sites by deep learning. *Bioinformatics* **33**, i234–i242 (2017).
54. Zhang, Y., Qiao, S., Ji, S. & Zhou, J. Ensemble-CNN: predicting DNA binding sites in protein sequences by an ensemble deep learning method. In *Proc. 14th Int. Conf. Intelligent Computing* 301–306 (ICIC, 2018).
55. He, F. et al. Protein ubiquitylation and sumoylation site prediction based on ensemble and transfer learning. In *Proc. 2019 IEEE Int. Conf. Bioinformatics and Biomedicine* 117–123 (IEEE, 2019).
56. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
57. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–1068 (2010).
58. Karim, M. R., Rahman, A., Jares, J. B., Decker, S. & Beyan, O. A snapshot neural ensemble method for cancer-type prediction based on copy number variations. *Neural Comput. Appl.* https://doi.org/10.1007/s00521-019-04616-9 (2019).
59. Erhan, D. et al. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res* **11**, 625–660 (2010).
60. Angermueller, C., Lee, H. J., Reik, W. & Stegle, O. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biol.* **18**, 67 (2017).
61. Hu, H. et al. Deephint: understanding HIV-1 integration via deep learning with attention. *Bioinformatics* **35**, 1660–1667 (2019).
62. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. Preprint at https://arxiv.org/abs/1409.0473 (2014).
63. Yang, Y. H. & Speed, T. Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* **3**, 579–588 (2002).
64. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98 (2011).
65. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
66. Grewal, J. K. et al. Application of a neural network whole transcriptome-based pan-cancer method for diagnosis of primary and metastatic cancers. *JAMA Netw. Open* **2**, e192597 (2019).
67. Xiao, Y., Wu, J., Lin, Z. & Zhao, X. A deep learning-based multi-model ensemble method for cancer prediction. *Comput. Methods Prog. Biomed.* **153**, 1–9 (2018).
68. West, M. D. et al. Use of deep neural network ensembles to identify embryonic-fetal transition markers: repression of *COX7A1* in embryonic and cancer cells. *Oncotarget* **9**, 7796–7811 (2018).
69. Tan, J. et al. Unsupervised extraction of stable expression signatures from public compendia with an ensemble of neural networks. *Cell Syst.* **5**, 63–71 (2017).
70. Lee, D., Redfern, O. & Orengo, C. Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell Biol.* **8**, 995–1005 (2007).
71. Li, Z. & Yu, Y. Protein secondary structure prediction using cascaded convolutional and recurrent neural networks. In *Proc. 25th Int. Joint Conf. Artificial Intelligence* 2560–2567 (AAAI, 2016).
72. Torrisi, M., Kaleel, M. & Pollastri, G. Deeper profiles and cascaded recurrent and convolutional neural networks for state-of-the-art protein secondary structure prediction. *Sci. Rep.* **9**, 12374 (2019).
73. Singh, J., Hanson, J., Paliwal, K. & Zhou, Y. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat. Commun.* **10**, 5407 (2019).
74. Zhang, B., Li, J. & Lü, Q. Prediction of 8-state protein secondary structures by a novel deep learning architecture. *BMC Bioinform.* **19**, 293 (2018).
75. Zacharaki, E. I. Prediction of protein function using a deep convolutional neural network ensemble. *PeerJ Comput. Sci.* **3**, e124 (2017).
76. Singh, J. et al. Detecting proline and non-proline cis isomers in protein structures from sequences using deep residual ensemble learning. *J. Chem. Inf. Model.* **58**, 2033–2042 (2018).

77. Walther, T. C. & Mann, M. Mass spectrometry-based proteomics in cell biology. *J. Cell Biol.* **190**, 491–500 (2010).
78. Cox, J. & Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80**, 273–299 (2011).
79. Zohora, F. T. et al. DeepIso: a deep learning model for peptide feature detection from LC-MS map. *Sci. Rep.* **9**, 17168 (2019).
80. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).
81. Kitano, H. Computational systems biology. *Nature* **420**, 206–210 (2002).
82. Hu, Y. et al. ACME: pan-specific peptide–MHC class I binding prediction through attention-based deep neural networks. *Bioinformatics* **35**, 4946–4954 (2019).
83. Zhang, L., Yu, G., Xia, D. & Wang, J. Protein–protein interactions prediction based on ensemble deep neural networks. *Neurocomputing* **324**, 10–19 (2019).
84. Karimi, M., Wu, D., Wang, Z. & Shen, Y. DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics* **35**, 3329–3338 (2019).
85. Hu, S. et al. Predicting drug-target interactions from drug structure and protein sequence using novel convolutional neural networks. *BMC Bioinform.* **20**, 689 (2019).
86. Yang, P. et al. Multi-omic profiling reveals dynamics of the phased progression of pluripotency. *Cell Syst.* **8**, 427–445 (2019).
87. Kim, H. J. et al. Transcriptional network dynamics during the progression of pluripotency revealed by integrative statistical learning. *Nucl. Acids Res.* **48**, 1828–1842 (2020).
88. Ramazzotti, D., Lal, A., Wang, B., Batzoglou, S. & Sidow, A. Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nat. Commun.* **9**, 4453 (2018).
89. Liang, M., Li, Z., Chen, T. & Zeng, J. Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **12**, 928–937 (2014).
90. Arefeen, A., Xiao, X. & Jiang, T. DeepPasta: deep neural network based polyadenylation site analysis. *Bioinformatics* **35**, 4577–4585 (2019).
91. Gala, R. et al. A coupled autoencoder approach for multi-modal analysis of cell types. In *Proc. 33st Int. Conf. Advances in Neural Information Processing Systems* 9263–9272 (NIPS, 2019).
92. Zhang, X. et al. Integrated multi-omics analysis using variational autoencoders: application to pan-cancer classification. In *Proc. 2019 IEEE Int. Conf. Bioinformatics and Biomedicine* 765–769 (IEEE, 2019).
93. Sharifi-Noghabi, H., Zolotareva, O., Collins, C. C. & Ester, M. MOLI: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* **35**, i501–i509 (2019).
94. Lu, Z. et al. The classification of gliomas based on a pyramid dilated convolution resnet model. *Pattern Recognit. Lett.* **133**, 173–179 (2020).
95. Codella, N. C. F. et al. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM J. Res. Dev.* **61**, 5 (2017).
96. Song, Y. et al. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. *IEEE Trans. Biomed. Eng.* **62**, 2421–2433 (2015).
97. Rasti, R., Teshnehlab, M. & Phung, S. L. Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks. *Pattern Recognit.* **72**, 381–390 (2017).
98. Yuan, X., Xie, L. & Abouelenien, M. A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. *Pattern Recognit.* **77**, 160–172 (2018).
99. Xie, J., Xu, B. & Chuang, Z. Horizontal and vertical ensemble with deep representation for classification. Preprint at https://arxiv.org/abs/1306.2759 (2013).
100. Dvornik, N., Schmid, C. & Mairal, J. Diversity with cooperation: ensemble methods for few-shot classification. In *Proc. IEEE Int. Conf. Computer Vision* 3723–3731 (IEEE, 2019).
101. Bzdok, D., Nichols, T. E. & Smith, S. M. Towards algorithmic analytics for large-scale datasets. *Nat. Mach. Intell.* **1**, 296–306 (2019).
102. Yang, P. et al. Sample subset optimization techniques for imbalanced and ensemble learning problems in bioinformatics applications. *IEEE Trans. Cybern.* **44**, 445–455 (2014).
103. Yang, P. et al. AdaSampling for positive-unlabeled and label noise learning with bioinformatics applications. *IEEE Trans. Cybern.* **49**, 1932–1943 (2019).
104. Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P. & Saeys, Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* **26**, 392–398 (2010).
105. Pusztai, L., Hatzis, C. & Andre, F. Reproducibility of research and preclinical validation: problems and solutions. *Nat. Rev. Clin. Oncol.* **10**, 720–724 (2013).
106. Dean, J. et al. Large scale distributed deep networks. In *Proc. 26th Int. Conf. Advances in Neural Information Processing Systems* 1223–1231 (NIPS, 2012).
107. Smith, V., Chiang, C.-K., Sanjabi, M. & Talwalkar, A. S. Federated multi-task learning. In *Proc. 31th Int. Conf. Advances in Neural Information Processing Systems* 4424–4434 (NIPS, 2017).

Acknowledgements

P.Y. was supported by an Australian Research Council (ARC) Discovery Early Career Researcher Award (DE170100759) and a National Health and Medical Research Council Investigator Grant (1173469). J.Y.H.Y. and P.Y. were supported by an ARC Discovery Project (DP170100654). Y.C. was supported by a University of Sydney Postgraduate Award. T.A.G. was supported by a postgraduate scholarship from Research Training Program.

Author contributions

P.Y. conceptualized this work. Y.C and P.Y. reviewed the literature and drafted the manuscript. All authors wrote and edited the Review Article.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to P.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2020