

Deep Learning, Genomics, and Precision Medicine

For preliminary authorship information, see the [contributors](#) on GitHub.

Abstract

Abstract goes here.

Introduction

Biology and medicine are rapidly becoming data-intensive with respect to both research and practice. A recent comparison of genomics with social media, online videos and other data-intensive scientific disciplines suggested that the field of genomics alone would equal or surpass other fields in data generation and analysis within the next decade [1]. These data present new opportunities, but also new challenges. The data volume and complexity both indicate that automated algorithms will be needed to extract meaningful patterns and provide actionable knowledge allowing us to better treat, categorize, or study disease, all within data constrained and privacy critical environments.

Concurrent with this explosive growth in biomedical data, a new class of machine learning algorithm - artificial neural networks, also known as deep learning - is revolutionizing domains from X to Y (I dunno - playing chess? serving ads?). As recently applied to image analysis problems, these architectures are blowing away prior best-in-class results, and computer scientists are now building many-layered neural networks from collections of millions of images. In an early famous example, scientists from Google demonstrated that a neural network could learn to identify cats simply by watching online videos [2].

What if, more generally, deep learning could solve the challenges presented by the growth of data in biomedicine? Could these algorithms identify the "cats" hidden in our data - the patterns unknown to the researcher - and act on them? Deep learning has transformed image analysis, but what about biomedicine more broadly? In this review, we examine whether this is simply a matter of time or if there are unique challenges posed by biomedical data that render deep learning methods more challenging or less fruitful to apply.

What is deep learning?

Deep learning is built on a biologically-inspired approach from machine learning termed neural networks. Each neuron in a computational neural network, termed a node, has inputs, an activation function, and outputs. Each value from the inputs is usually multiplied by some weight and combined and summarized by the activation function. The value of the activation function is then multiplied by another set of weights to produce the output **TODO: we probably need a figure here - I see no way that we don't include this type of description in our paper, despite the fact that it's been done tons of times before. - I'm really partial to this nature review's explanation about making non-linear problems linear - figure 1 [3]** These neural networks are trained by identifying weights that produce a desired output from some specific input.

Neural networks can also be stacked. The outputs from one network can be used as inputs to another. This process produces a stacked, also known as a multi-layer, neural network. The multi-layer neural network techniques that underlie deep learning have a long history. Multi-layer methods have been discussed in the literature for more than five decades [4]. Given this context, it's challenging to consider "deep learning" as a new advance, though the term has only become widespread to describe analysis methods in the last decade. Much of the early history of neural networks has been extensively covered in a recent review [5]. For the purposes of this review, we identify deep learning approaches as those that use multi-layer neural networks to construct complex features.

We also identify a class of algorithms that we term "shallow learning" approaches. We do not use this as a pejorative term, but instead to denote algorithms which have all of the hallmarks of deep approaches except that they employ networks of limited depth. We found it valuable to include these as we sought to identify the current contributions of deep learning and to predict its future impact. Researchers may employ these shallow learning methods for a number of reasons including: 1) shallow networks provide a degree of interpretability that better matches their use case; 2) the available data are insufficient to support deeper architectures, however new datasets that will support deep methods are expected; 3) or as building blocks to be combined with other non-neural-network-based approaches at subsequent stages.

Will deep learning transform the study of human disease?

What would need to be true for deep learning to transform how we categorize, study, and treat individuals to maintain or restore health? With this review we set out to address this question, setting a high bar for "transform." Specifically, we sought to identify whether deep learning was a disruptive innovation that would induce a strategic inflection point on the practice of biology or medicine. There are numerous examples where deep learning has been applied to biological problems and produced somewhat improved results, and there are numerous reviews that have focused on general

applications of deep learning in biology [6]. We sought cases where deep learning was enabling researchers to solve challenges that were previously considered infeasible, or if it made difficult, tedious, and non-routine analyses routine.

Based on our guiding question, we focused on the application of deep learning to topics of biomedical importance. We divided the large range of topics into three broad classes: Disease and Patient Categorization, Fundamental Biological Study, and Patient Treatment. We briefly introduce the types of questions, approaches and data that are typical for each class in the application of deep learning.

Disease and Patient Categorization

A key challenge in biomedicine is the accurate classification of diseases and disease subtypes. In oncology, current "gold standard" approaches involve histology, requiring manual human expertise for quantification, or small panel of molecular markers, such as cell surface receptors or genes' expression. One example is the PAM50 approach to classifying breast cancer where the expression of 50 marker genes divides breast cancer patients into four subtypes. Significant heterogeneity still remains within these four subtypes [12]. Given the increasing wealth of molecular data available, a more comprehensive subtyping seems possible.

Several studies have used deep learning methods in order to better categorize breast cancer patients. For example, Tan et al. applied denoising autoencoders (DA), an unsupervised approach, in order to cluster breast cancer patients [14]. Ciresan et al. utilized convolutional neural networks (CNN) to count mitotic divisions in histological images; a feature that is highly correlated with disease outcome [15]. Despite these recent advances, a number of challenges exist in this area of research, such as the integration of disparate types of data, including electronic health records (EHR), imaging and histology data, and molecular omics data.

Fundamental Biological Study

Deep learning can be applied to answer more fundamental biological questions, and is especially suited to leveraging large amounts of data from high throughput omics studies. One classic biological problem where machine learning has been extensively applied is the prediction of molecular targets. Recent advances using deep learning have shown higher accuracy in determining molecular targets. For example, Lee et al. used deep recurrent neural networks (RNN) to predict gene targets of micro-RNAs [16]. Wang et al. used a residual CNN to predict protein-protein contact on a genome-wide scale [17]. Other biological questions that have been investigated include the prediction of protein secondary structure based on sequence data [18], recognition of functional genomic elements such as enhancers and promoters [20], predicting the deleterious effects of nucleotide polymorphisms[23], etc.

Patient Treatment

Although the application of deep learning to patient treatment is just beginning, we expect a dramatic increase in methods aiming to recommend patient treatment, predict treatment outcome, and guide future development of new therapies. Specifically, effort in this area aims to identify drug targets, identify drug interactions or predict drug response. One recent approach for predicting drug response is the use of protein structure to predict drug interactions and drug bioactivity through CNN [24]. Since CNNs leverage spatial relationships within the data, this particular deep learning framework is well suited to the problem. Drug discovery and drug "repurposing" are two other hot topics. Aliper et al. used transcriptomic data to predict which drugs might be repurposed for other diseases through deep fully connected neural networks [25]. In a similar vein, Wang et al. used restricted boltzman machines (RBM) to predict drug molecular targets [26].

Does deep learning create a strategic inflection point in how we categorize individuals with respect to health and disease?

Focus for the purpose of this review - within a health care context.

We currently categorize individuals using relatively *ad hoc* categories. These are divided, to an extent, by organ (e.g. cancers by tumor site), perhaps to an extent symptom, and to an extent by immediate cause. This system undergoes continual refinement (e.g. new subtypes of disease) as our understanding improves.

Would deep learning enable us to do this automatically in some principled way? Are there reasons to believe that this would be advantageous? Would it be positive to have disease categories changed by data, or would the changing definition (i.e. as more data are accumulated) actually be harmful? What impacts would this have on the training of physicians?

What are the major challenges in this space, and does deep learning enable us to tackle any of them? Are there example approaches whereby deep learning is already having a transformational impact? I (Casey) have added some sections below where I think we could contribute to the field with our discussion.

Major areas of existing contributions

There are a number of major challenges in this space. How do we get data together from multiple distinct systems? How do we find biologically meaningful patterns in that data? How do we store and compute on this data at scale? How do we share these data while respecting privacy? I've made a section for each of these. Feel free to add more. I see each section as something on the order of 1-2 paragraphs in our context.

Clinical care

Imaging applications in health care

One of the general areas where deep learning methods have had substantial success has been in image analysis. Applications in areas of medicine that use imaging extensively are also emerging. Mammography has been one area with numerous contributions [27]. In all of this work, the researchers must work around a specific challenge - the limited number of well annotated training images. To expand the number and diversity of images, the researchers have employed approaches where they employ adversarial examples [30] or first train towards human-created features before subsequent fine tuning [27]. The presence of a large bank of well-annotated mammography images would aid in the application of deep neural networks to this area. Though this strategy has not yet been employed in this domain, large collections of unlabeled images might first be used in an unsupervised context to construct high-quality feature detectors. Then the small number of labeled examples could be used for subsequent training. Similar strategies have been employed for EHR data where high-quality labeled examples are also difficult to obtain [31].

In addition to radiographic images, histology slides are also being analyzed with deep learning approaches. Ciresan et al. [15] developed one of the earliest examples, winning the 2012 International Conference on Pattern Recognition's Contest on Mitosis Detection while achieving human competitive accuracy. Their approach uses what has become a standard convolutional neural network architecture trained on public data. In more recent work, Wang et al.[32] analyzed stained slides to identify cancers within slides of lymph node slices. The approach provided a probability map for each slide. On this task a pathologist has about a 3% error rate. The pathologist did not produce any false positives, but did have a number of false negatives. Their algorithm had about twice the error rate of a pathologist. However, their algorithms errors were not strongly correlated with the pathologist. Theoretically, combining both could reduce the error rate to under 1%. In this area, these algorithms may be ready to incorporate into existing tools to aid pathologists. The authors' work suggests that this could reduce the false negative rate of such evaluations. This theme of an ensemble between deep learning algorithm and human expert may help overcome some of the challenges presented by data limitations.

One source of training examples with rich clinical annotations is the electronic health record. Recently Lee et al.[33] developed an approach to distinguish individuals with Age-related Macular Degeneration from control individuals. They extracted approximately 100,000 images from structured electronic health records, which they used to train and evaluate a deep neural network. Combining this data resource with standard deep learning techniques, the authors reach greater than 93% accuracy. One item that is important to note with regards to this work is that the authors used their test set for evaluating when training had concluded. In other domains, this has resulted in a minimal change in the estimated accuracy [34]. However, there is not yet a single accepted standard within the field of biomedical research for such evaluations. We recommend the use of an independent test set wherever it is feasible. Despite this minor limitation, the work clearly illustrates the potential that can be unlocked from images stored in electronic health records.

TODO: Potential remaining topics: #122 & #151 looked interesting from an early glance. - Do we want to make the point that most of the imaging exam

Electronic health records

EHR data include substantial amounts of free text, which remains challenging to approach[35]. Often, researchers developing algorithms that perform well on specific tasks must design and implement domain- specific features [36]. These features capture unique aspects of the literature being processed. Deep learning methods are natural feature constructors. In recent work, the authors evaluated the extent to which deep learning methods could be applied on top of generic features for domain-specific concept extraction [37]. They found that performance was in line with, but did not exceed, existing state of the art methods. The deep learning method had performance lower than the best performing domain-specific method in their evaluation [37]. This highlights the challenge of predicting the eventual impact of deep learning on the field. This provides support that deep learning may impact the field by reducing the researcher time and cost required to develop specific solutions, but it may not lead to performance increases.

In recent work, Yoon et al.[38] analyzed simple features using deep neural networks and found that the patterns recognized by the algorithms could be re-used across tasks. Their aim was to analyze the free text portions of pathology reports to identify the primary site and laterality of tumors. The only features the authors supplied to the algorithms that they evaluated were unigrams and bigrams. These are the counts for single words and two-word combinations in a free text document. They subset the full set of words and word combinations to the 400 most commonly used ones. The machine learning algorithms that they employed (naïve Bayes, logistic regression, and deep neural networks) all performed relatively similarly on the task of identifying the primary site. However, when the authors evaluated the more challenging task, i.e. evaluating the laterality of each tumor, the deep neural network outperformed the other methods. Of particular interest, when the authors first trained a neural network to predict primary site and then

repurposed those features as a component of a secondary neural network trained to predict laterality, the performance was higher than a laterality-trained neural network. This indicates a potential strength of deep methods. It may be possible to repurpose features from task to task, improving overall predictions as the field tackles new challenges.

Identifying consistent subgroups of individuals and individual health trajectories from clinical tests is also an active area of research. Approaches inspired by deep learning have been used for both unsupervised feature construction and supervised prediction. Early work by Lasko et al. [39], combined sparse autoencoders and Gaussian processes to distinguish gout from leukemia from uric acid sequences. Later work showed that unsupervised feature construction of many features via denoising autoencoder neural networks could dramatically reduce the number of labeled examples required for subsequent supervised analyses [40]. In addition, it pointed towards learned features being useful for subtyping within a single disease. A concurrent large-scale analysis of an electronic health records system found that a deep denoising autoencoder architecture applied to the number and co-occurrence of clinical test events, though not the results of those tests, constructed features that were more useful for disease prediction than other existing feature construction methods [41]. Taken together, these results support the potential of unsupervised feature construction in this domain. However, numerous challenges including data integration (patient demographics, family history, laboratory tests, text-based patient records, image analysis, genomic data) and better handling of streaming temporal data with many features, will need to be overcome before we can fully assess the potential of deep learning for this application area.

Still, recent work has also revealed domains in which deep networks have proven superior to traditional methods. Survival analysis models the time leading to an event of interest from a shared starting point, and in the context of EHR data, often associates these events to subject covariates. Exploring this relationship is difficult, however, given that EHR data types are often heterogeneous, covariates are often missing, and conventional approaches require the covariate-event relationship be linear and aligned to a specific starting point [42]. Early approaches, such as the Faraggi-Simon feed-forward network, aimed to relax the linearity assumption, but performance gains were lacking [43]. Katzman et al. in turn developed a deep implementation of the Faraggi-Simon network that, in addition to outperforming Cox regression, was capable of comparing the risk between a given pair of treatments, thus potentially acting as recommender system [44]. To overcome the remaining difficulties, researchers have turned to deep exponential families, a class of latent generative models that are constructed from any type of exponential family distributions [45]. The result was a deep survival analysis model capable of overcoming challenges posed by missing data and heterogeneous data types, while uncovering nonlinear relationships between covariates and failure time. They showed their model more accurately stratified patients as a function of disease risk score compared the current clinical implementation. TODO: @sw1: Are there specific challenges with using a deep neural network as opposed to current standard methods?

Opportunities

However, significant work needs to be done to move these from conceptual advances to practical game-changers.

- Large data resources (see sample # issues that mammography researchers are working around)
- Semi-supervised methods to take advantage of large number of unlabeled examples
- Transfer learning.

Unique challenges

Additionally, unique barriers exist in this space that may hinder progress in this field.

Standardization/integration

EHRs are designed and optimized primarily for patient care and billing purposes, meaning research is at most a tertiary priority. This presents significant challenges to EHR based research in general, and particularly to data intensive deep learning research. EHRs are used differently even within the same health care system [46]. Individual users have unique usage patterns, and different departments have different priorities which introduce missing data in a non-random fashion. Just et al. demonstrated that even the most basic task of matching patients can be challenging due to data entry issues [48]. This is before considering challenges caused by system migrations and health care system expansions through acquisitions. Replication between hospital systems requires controlling for both these systematic biases as well as for population and demographic effects. Historically, rules-based algorithms have been popular in EHR-based research but because these are developed at a single institution and trained with a specific patient population they do not transfer easily to other populations [49]. Wiley et al. [50] showed that warfarin dosing algorithms often under perform in African Americans, illustrating that some of these issues are unsolved even at a treatment best practices level. This may be a promising application of deep learning, as rules-based algorithms were also the standard in most natural language processing but have been superseded by machine learning and in particular deep learning methods [51].

Temporal Patient Trajectories

Traditionally, physician training programs justified long training hours by citing increased continuity of care and learning

by following the progression of a disease over time, despite the known consequences of decreased mental and quality of life [52]. Yet, a common practice in EHR-based research is to take a point in time snapshot and convert patient data to a traditional vector for machine learning and statistical analysis. This results in significant signal losses as timing and order of events provide insight into a patient's disease and treatment. Efforts to account for the order of events have shown promise [56] but require exceedingly large patient sizes due to discrete combinatorial bucketing.

Lasko et al. [39] used autoencoders on longitudinal sequences of serum urine acid measurements to identify population subtypes. More recently, deep learning has shown promise working with both sequences (Convolutional Neural Networks) [57] and the incorporation of past and current state (Recurrent Neural Networks, Long Short Term Memory Networks)[58].

Data sharing and privacy

Early successes using deep learning involved very large training datasets (ImageNet 1.4 million images)[59], but a responsibility to protect patient privacy limits the ability openly share large patient datasets. Limited dataset sizes may restrict the number of parameters that can be trained in a model, but the lack of sharing may also hamper reproducibility and confidence in results. Even without sharing data, algorithms trained on confidential patient data may present security risks or accidentally allow for the exposure of individual level patient data. Tramer et al. [60] showed the ability to steal trained models via public APIs and Dwork and Roth [61] demonstrate the ability to expose individual level information from accurate answers in a machine learning model.

Training algorithms in a differentially private manner provides a limited guarantee that the algorithms output will be equally likely to occur regardless of the participation of any one individual. The limit is determined by a single parameter which provides a quantification of privacy. Simmons et al. [62] present the ability to perform GWASs in a differentially private manner and Abadi et al. [63] show the ability to train deep learning classifiers under the differential privacy framework. Finally, Continuous Analysis [64] allows for the ability to automatically track and share intermediate results for the purposes of reproducibility without sharing the original data.

Biomedical data is often "Wide"

Biomedical studies typically deal with relatively small sample sizes but each sample may have millions of measurements (genotypes and other omics data, lab tests etc).

Classical machine learning recommendations were to have 10x samples per number of parameters in the model.

Number of parameters in an MLP. Convolutions and similar strategies help but do not solve

Bengio diet networks paper

Has deep learning already induced a strategic inflection point for one or more aspects?

I have looked through the papers that we have. I don't see a case in our collection where I felt that we'd be justified to say that deep learning has transformed how we categorize individuals with respect to health and disease. There are definitely interesting applications, but I don't see anything that we couldn't do similarly with some other method.

Will deep learning induce a strategic inflection point for categorization?

This section attempts to get at whether or not we think that deep learning will be transformational. Since we have some room to provide our perspective, I'd suggest that we take a relatively tough look at this once we review where we are in the parts above.

What unique potential does deep learning bring to this?

Are there areas that we expect deep learning to transform how we categorize disease that we haven't seen yet? Let's get fun with speculation/dreaming on this one.

Where would you point your deep learning efforts if you had the time?

This can be fun. We might eventually merge this with the section immediately above on deep learning's unique potential here.

How is deep learning used to study basic biological processes in a manner that may provide future insights into human disease?

The (awkward) placeholder section title is intended to help define the scope. We do not want this section to become a miscellaneous collection of everything that does not fit in Categorize and Treat.

One proposal is that we organize this roughly by what is being predicted, which will generally correspond to the types of data being used. For each sub-section we can quickly introduce the prediction problem and cite some examples of the relevance to disease. Hypothetically, if we had an algorithm that produced perfect predictions on the task, what would we learn and how could those predictions be used?

Existing reviews could be mentioned briefly.

It may not fit here, but there could be a general discussion of why different neural network architectures are particularly well-suited for different types of input data. For example, CNNs and RNNs for 1-dimensional data are used in several categories below.

A few suggestions for sub-sections follow. Some of these could be left out because our goal is not an exhaustive enumeration of methods. Some are important areas of biology, but there may not be much deep learning- specific content to present. Others may be important areas where we lack expertise, in which case we may acknowledge the application area but not dive into merits or weaknesses of individual methods.

Gene expression

Predicting gene expression levels and unsupervised approaches for learning from gene expression. Those could be divided into separate sub-sections.

Splicing

A separate section from general gene expression section above.

Transcription factors and RNA-binding proteins

Existing reviews have covered some of these papers rather well and we do not want to repeat what has already been well-stated elsewhere. This could be split into two sub-sections or kept very brief.

Promoters, enhancers, and related epigenomic tasks

We may want to be selective about what we discuss and not list every application in this area.

Micro-RNA binding

miRNAs are important biologically, but have neural networks produced anything particularly notable in this area?

Protein secondary and tertiary structure

Jinbo Xu is writing this

Signaling

There is not much content here. Can[65] be covered elsewhere?

Morphological phenotypes

A field poised for dramatic revolution by deep learning is bioimage analysis. Thus far, the primary use of deep learning for biological images has been for segmentation - that is, for the identification of biologically relevant structures in images such as nuclei, infected cells, or vasculature, in fluorescence or even brightfield channels [66]. Once so-called regions of interest have been identified, it is often straightforward to measure biological properties of interest, such as fluorescence intensities, textures, and sizes. Given the dramatic successes of deep learning in biological imaging, we simply refer to articles that review recent advancements [8]. We believe deep learning will become a commonplace tool for biological image segmentation once user-friendly tools exist.

We anticipate an additional kind of paradigm shift in bioimaging that will be brought about by deep learning: what if images of biological samples, from simple cell cultures to three-dimensional organoids and tissue samples, could be mined for much more extensive biologically meaningful information than is currently standard? For example, a recent study demonstrated the ability to predict lineage fate in hematopoietic cells up to three generations in advance of differentiation [68]. In biomedical research, by far the most common paradigm is for biologists to decide in advance what feature to measure in images from their assay system. But images of cells contain a wide variety of quantitative information, and deep learning may just be the tool to extract it. Although classical methods of segmentation and feature extraction can produce hundreds of metrics per cell in an image, deep learning is unconstrained by human intuition and can in theory extract more subtle features. Already, there is evidence deep learning can surpass the efficacy of classical methods [69], even using generic deep convolutional networks trained on natural images[70], known as transfer learning.

The impact of further improvements on biomedicine could be enormous. Comparing cell population morphologies using conventional methods of segmentation and feature extraction has already proven useful for functionally annotating genes and alleles, identifying the cellular target of small molecules, and identifying disease-specific phenotypes suitable for drug screening [71]. Deep learning would bring to these new kinds of experiments - known as image-based profiling or morphological profiling - a higher degree of accuracy, stemming from the freedom from human-tuned feature extraction strategies.

TODO: Make sure that at the end we clearly emphasize our excitement around unsupervised uses.

Single-cell

There are not many neural network papers in this area (yet), unless we count imaging applications. But there is still plenty to discuss. The existing methods [74] use interesting network architectures to approach single-cell data. [76] could fit here.

Metagenomics

TODO: Add reference tags to this section Metagenomics (which refers to the study of genetic material, 16S rRNA and/or whole-genome shotgun DNA, from microbial communities) has revolutionized the study of micro-scale ecosystems within us and around us. There is increasing literature of applying machine learning in general to metagenomic analysis. In the late 2000's, a plethora of machine learning methods were applied to classifying DNA sequencing reads to the thousands of species within a sample. An important problem is genome assembly from these mixed-organism samples. And to do that, the organisms should be "binned" before assembling. Binning methods began with many k-mer techniques [refs] and then delved into other clustering algorithms, such as self-organizing maps (SOM). Then came the taxonomic classification problem, with researchers naturally using BLAST [blast], followed by other machine learning techniques such as SVMs [McHardy], naive Bayesian classifiers [nbc], etc. to classify each read. Then, researchers began to use techniques that could be used to estimate relative abundances of an entire sample, instead of the precise but painstakingly slow read-by-read classification. Relative abundance estimators (a.k.a diversity profilers) are MetaPhlan[ref], (WGS)Quikr[ref], and some configurations of tools like OneCodex[ref] and LMAT[ref]. While one cannot identify which reads were mapped back to an organism using relative abundance estimators, they can be useful for faster comparative and other downstream analyses. Newer methods hope to classify reads and estimate relative abundances at faster rates [Vervier] and as of this writing, there are more than 70 metagenomic taxonomic classifiers in existence. Besides binning and classification of species, there is functional identification and annotation of sequence reads [Yok,Soueidan]. However, the focus on taxonomic/functional annotation is just the first step. Once organisms are identified, there is the interest in understanding the interrelationship between these organisms and host/environment phenotypes [Guetterman]. One of the first attempts was a survey of supervised classification methods for microbes->phenotype classification [Knights], followed by similar studies that are more massive in scale [Stratnikov, Segata]. There have been techniques that bypass the taxonomic classification step altogether [Ding et al] , (sequence composition to phenotype classification). Also, researchers have looked into how feature selection can improve classification [Liu, Segata], and techniques have been proposed that are classifier-independent [Ditzler,Ditzler].

So, how have neural networks (NNs) been of use? Most neural networks are being used for short sequence->taxa/function classification, where there is a lot of data for training (and thus suitable for NNs). Neural networks have been applied successfully to gene annotation (e.g. Orphelia [Hoff] and FragGeneScan [77]), which usually has plenty of training examples. Representations (similar to Word2Vec [ref] in natural language processing) for protein family classification has been introduced and classified with a skip-gram neural network [Asgari]. Recurrent neural networks show good performance for homology and protein family identification [Hochreiter, Sonderby]. Interestingly, Hochreiter, who invented Long Short Term Memory, delved into homology/protein family classification in 2007, and therefore, deep learning is deeply rooted in functional classification methods.

One of the first techniques of "de novo" genome binning used self-organizing maps, a type of NN [Abe]. Essinger et al. use ART, a neural network algorithm called Adaptive Resonance Theory, to cluster similar genomic fragments and showed that it has better performance than K-means. However, other methods based on interpolated Markov models [Salzberg] have performed better than these early genome binners. Also, neural networks can be slow, and therefore, have had limited use for reference-based taxonomic classification, with TAC-ELM [tac-elm] being the only NN-based algorithm to taxonomically classify massive amounts of metagenomic data. Also, neural networks can fail to perform if there are not enough training examples, which is the case with taxonomic classification (since only ~10% of estimated species have been sequenced). An initial study shows that deep neural networks have been successfully applied to taxonomic classification of 16S rRNA genes, with convolutional networks provide about 10% accuracy genus-level improvement over RNNs and even random forests [Mrzelj]. However, this study performed 10-fold cross-validation on 3000 sequences in total.

Due to the traditionally small numbers of metagenomic samples in studies, neural network uses for classifying phenotype from microbial composition are just beginning. A standard MLP was able to classify wound severity from microbial species present in the wound [78]. Recently, multi-layer, recurrent networks (and convolutional networks) have been applied to microbiome genotype-phenotype, with Ditzler et al. being the first to associate soil samples with pH

level using multi-layer perceptrons, deep-belief networks, and recursive neural networks (RNNs) [Ditzler]. Besides classifying the samples appropriately, Ditzler shows that internal phylogenetic tree nodes inferred by the networks are appropriate features representing low/high pH, which can provide additional useful information and new features for future metagenomic sample comparison. Also, an initial study has shown promise of these networks for diagnosing disease [Faruqi].

There are still a lot of challenges with applying deep neural networks to metagenomics problems. They are not ideal for microbial/functional composition->phenotype classification because most studies contain tens of samples (~20->40) and hundreds/thousands of features (aka species). Such underdetermined/ill-conditioned problems are still a challenge for deep neural networks that require many more training examples than features to sufficiently converge the weights on the hidden layers. Also, due to convergence issues (slowness and instability due to large neural networks modeling very large datasets [79]), taxonomic classification of reads from whole genome sequencing seems out of reach at the moment for deep neural networks -- due to only thousands of full-sequenced genomes as compared to hundreds of thousands of 16S rRNA sequences available for training.

However, because recurrent neural networks are showing success for base-calling (and thus removing the large error in the measurement of a pore's current signal) for the relatively new Oxford Nanopore sequencer [Boza], there is hope that the process of denoising->organism/function classification can be combined into one step in using powerful LSTM's. LSTM's are working miracles in raw speech signal->meaning translation [ref], and combining steps in metagenomics are not out of the question. For example, metagenomic assembly usually requires binning then assembly, but could deep neural nets accomplish both tasks in one network? Does functional/taxonomic classification need to be separate processes? The largest potential in deep learning is to learn "everything" in one complex network, with a plethora of labeled (reference) data and unlabeled (microbiome experiments) examples.

Sequencing and variant calling

We have one nanopore paper in the issues and very recent work on variant calling that looks worthy of inclusion.

The impact of deep learning in treating disease and developing new treatments

There will be some overlap with the Categorize section, and we may have to determine which methods categorize individuals and which more directly match patients with treatments. The sub-section titles are merely placeholders.

Categorizing patients for clinical decision making

How can deep learning match patients with clinical trials, therapies, or other interventions? As an example,[80] predicts individuals who are most likely to decline during a clinical trial and benefit from the treatment.

Effects of drugs on transcriptomic responses

We discussed a few papers that operate on Library of Network-Based Cellular Signatures (LINCS) gene expression data. We could briefly introduce the goals of that resource and comment on the deep learning applications. In the Issues, we had reservations about whether the improvements in expression prediction are good enough to make a practical difference in the domain and feature selection and construction.

Ligand-Based Prediction of Bioactivity

TODO: expand outline

- Short introduction to problem, related reviews, use vHTS definition from[81] (vHTS doesn't fit neatly into classic classification, regression, or ranking)
- Introduce ligand-based approaches, hype and excitement surrounding performance of a "high-parameter" network on the Merck Kaggle challenge, cover other neural networks trained on fingerprints or descriptors as features that followed, Tox21 Data Challenge
- Multitask networks related to the above point
- Realistic view of where things stand today, high-parameter networks struggle with overfitting, cross validation needs to be done carefully because of temporal structure [82], low parameter networks based on chemical similarity (IRV) work very well, especially well-suited for the domain in which training data can be limited and contains few positive instances, may touch on BACE example here and other discussions of training data limitations (e.g. [83])
- "Creative experimentation" phase of the field, new ideas for representation learning and novel approaches including graph convolutions, autoencoders, one shot learning, and generative models
- These "creative" approaches are definitely interesting but aren't necessarily outperforming existing methods, improvements on the software and reusability side could be important to help establish more rigorous benchmarking, DeepChem as example of this

- Future outlook, what would need to happen for the "creative" approaches to overtake the current state of the art, can representation learning be improved by incorporating more information about chemical properties or even more "tasks" during training, how much will future growth depend on data versus algorithms
- Future outlook part 2, how the above approaches relate to traditional methods like docking (note neural networks that include docking scores as features), deep learning efforts in this direction that use structure (e.g. [24]), "zero-shot learning", analogies to other domains where deep learning can capture the behavior of complex physics (e.g. quantum physics example), maybe briefly mention other compound-protein interaction-based networks although that doesn't seem to fit here and is somewhat out of scope
- Future output part 3 (most speculative), what would successful generative networks mean for the HTS field?

Modeling Metabolism and Chemical Reactivity

Add a review here of metabolism and chemical reactivity.

Discussion

This section provides meta-commentary that spans the Categorize, Study, and Treat subject areas. The candidate sub-sections below are initial ideas that can be further pruned.

Evaluation

What are the challenges in evaluating deep learning models that are specific to this domain? This can include a discussion of ROC versus precision-recall curves for the imbalanced classes often encountered in biomedical datasets. It could also mention alternative metrics that are used in specific sub-areas such as enrichment factors in virtual screening. A lack of true gold standard data for some problems complicates both training and evaluation. Confidence-weighted labels are valuable when available.

Interpretation

Most of our examples pertain to the Study papers. Does this discussion belong there or can we generalize it and keep it here? Specific points would include the dangers of over-interpreting hidden units, pros/cons of specific techniques (see issues), and recommendations. Some other reviews have addressed this in part as well.

Data limitations

Related to evaluation, are there data quality issues in genomic, clinical, and other data that make this domain particularly challenging? Are these worse than what is faced in other non-biomedical domains?

Many applications have used relatively small training datasets. We might discuss workarounds (e.g. semi-synthetic data, splitting instances, etc.) and how this could impact future progress. Might this be why some studies have resorted to feature engineering instead of learning representations from low-level features? Is there still work to be done in finding the right low-level features in some problems?

Hardware limitations and scaling

Several papers have stated that memory or other hardware limitations artificially restricted the number of training instances, model inputs/outputs, hidden layers, etc. Is this a general problem worth discussing or will it be solved naturally as hardware improves and/or groups move to distributed deep learning frameworks? Does hardware limit what types of problems are accessible to the average computational group, and if so, will that limit future progress? For instance, some hyperparameter search strategies are not feasible for a lab with only a couple GPUs.

Some of this is also outlined in the Categorize section. We can decide where it best fits.

Efficiently scaling deep learning is challenging, and there is a high computational cost (e.g., time, memory, energy) associated with training neural networks and using them for classification. As such, neural networks have only recently found widespread use [5].

Many have sought to curb the costs of deep learning, with methods ranging from the very applied (e.g., reduced numerical precision [85]) to the exotic and theoretic (e.g., training small networks to mimic large networks and ensembles [89]). The largest gains in efficiency have come from computation with graphics processing units (GPUs) [91], which excel at the matrix and vector operations so central to deep learning. The massively parallel nature of GPUs allows additional optimizations, such as accelerated mini-batch gradient descent [92]. However, GPUs also have a limited quantity of memory, making it difficult to implement networks of significant size and complexity on a single GPU or machine [91]. This restriction has sometimes forced computational biologists to use workarounds or limit the size of an analysis. For example, Chen et al. [99] aimed to infer the expression level of all genes with a single neural network, but due to memory restrictions they randomly partitioned genes into two halves and analyzed each separately. In other cases, researchers limited the size of their neural network [17]. Some have also chosen to use slower CPU

implementations rather than sacrifice network size or performance [101].

Steady improvements in GPU hardware may alleviate this issue somewhat, but it is not clear whether they can occur quickly enough to keep up with the growing amount of available biological data or increasing network sizes. Much has been done to minimize the memory requirements of neural networks [102], but there is also growing interest in specialized hardware, such as field-programmable gate arrays (FPGAs) [95] and application-specific integrated circuits (ASICs). Specialized hardware promises improvements in deep learning at reduced time, energy, and memory [95]. Logically, there is less software for highly specialized hardware [104], and it could be a difficult investment for those not solely interested in deep learning. However, it is likely that such options will find increased support as they become a more popular platform for deep learning and general computation.

Distributed computing is a general solution to intense computational requirements, and has enabled many large-scale deep learning efforts. Early approaches to distributed computation [105] were not suitable for deep learning [107], but significant progress has been made. There now exist a number of algorithms [107], tools [109], and high-level libraries [112] for deep learning in a distributed environment, and it is possible to train very complex networks with limited infrastructure [114]. Besides handling very large networks, distributed or parallelized approaches offer other advantages, such as improved ensembling [115] or accelerated hyperparameter optimization [116].

Cloud computing, which has already seen adoption in genomics [118], could facilitate easier sharing of the large datasets common to biology [119], and may be key to scaling deep learning. Cloud computing affords researchers significant flexibility, and enables the use of specialized hardware (e.g., FPGAs, ASICs, GPUs) without significant investment. With such flexibility, it could be easier to address the different challenges associated with the multitudinous layers and architectures available [121]. Though many are reluctant to store sensitive data (e.g., patient electronic health records) in the cloud, secure/regulation-compliant cloud services do exist [122].

TODO: Write the transition once more of the Discussion section has been fleshed out.

Code, data, and model sharing

Reproducibility is important for science to progress. In the context of deep learning applied to advance human healthcare, does reproducibility have different requirements or alternative connotations? With vast hyperparameter spaces, massively heterogeneous and noisy biological data sets, and black box interpretability problems, how can we best ensure reproducible models? What might a clinician, or policy maker, need to see in a deep model in order to influence healthcare decisions? Or, is deep learning a hypothesis generation machine that requires manual validation?

Transfer learning/transferability of features

- <https://github.com/greenelab/deep-review/issues/139#issuecomment-268901804>

Conclusions

Final thoughts and future outlook here. The Discussion will give an overview and the Conclusion will provide a short, punchy take home message.

Points to mention based on discussion thus far that may make the bar for conclusions:

- Limitations of data & workarounds [availability impacts on amount, etc]
- Transferability of features
- Strong enthusiasm for unsupervised approaches.
- Right to an explanation [possibly, depends if raised in multiple areas]

Author contributions

TODO: not sure if it should go here, but somewhere we should talk about how we wrote this thing, since it is still somewhat unconventional to have a
We recognized that writing a review on a rapidly developing area in a manner that allowed us to provide a forward-looking perspective on diverse approaches and biological problems would require expertise from across computational biology and medicine. We created an open repository on the GitHub version control system and engaged with numerous authors from papers within and outside of the area. Paper review was conducted in the open by # individuals, and the manuscript was drafted in a series of commits from # authors. Individuals who met the ICJME standards of authorship are included as authors. These were individuals who contributed to the review of the literature; drafted the manuscript or provided substantial critical revisions; approved the final manuscript draft; and agreed to be accountable in all aspects of the work. Individuals who did not contribute in one or more of these ways, but who did participate, are acknowledged at the end of the manuscript.

1. Stephens ZD *et al.* 2015 Big Data: Astronomical or Genomical? *PLOS Biology* **13**, e1002195. (doi:[10.1371/journal.pbio.1002195](https://doi.org/10.1371/journal.pbio.1002195))

2. In press.
3. LeCun Y, Bengio Y, Hinton G. 2015 Deep learning. *Nature* **521**, 436–444. (doi:[10.1038/nature14539](https://doi.org/10.1038/nature14539))
4. Block HD, Knight BW, Rosenblatt F. 1962 Analysis of a Four-Layer Series-Coupled Perceptron. II. *Reviews of Modern Physics* **34**, 135–142. (doi:[10.1103/revmodphys.34.135](https://doi.org/10.1103/revmodphys.34.135))
5. Schmidhuber J. 2015 Deep learning in neural networks: An overview. *Neural Networks* **61**, 85–117. (doi:[10.1016/j.neunet.2014.09.003](https://doi.org/10.1016/j.neunet.2014.09.003))
6. Park Y, Kellis M. 2015 Deep learning for regulatory genomics. *Nature Biotechnology* **33**, 825–826. (doi:[10.1038/nbt.3313](https://doi.org/10.1038/nbt.3313))
7. Gawehn E, Hiss JA, Schneider G. 2015 Deep Learning in Drug Discovery. *Molecular Informatics* **35**, 3–14. (doi:[10.1002/minf.201501008](https://doi.org/10.1002/minf.201501008))
8. Kraus OZ, Frey BJ. 2016 Computer vision for high content screening. *Critical Reviews in Biochemistry and Molecular Biology* **51**, 102–109. (doi:[10.3109/10409238.2015.1135868](https://doi.org/10.3109/10409238.2015.1135868))
9. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. 2016 Applications of Deep Learning in Biomedicine. *Molecular Pharmaceutics* **13**, 1445–1454. (doi:[10.1021/acs.molpharmaceut.5b00982](https://doi.org/10.1021/acs.molpharmaceut.5b00982))
10. Angermueller C, Pärnamaa T, Parts L, Stegle O. 2016 Deep learning for computational biology. *Molecular Systems Biology* **12**, 878. (doi:[10.15252/msb.20156651](https://doi.org/10.15252/msb.20156651))
11. Min S, Lee B, Yoon S. 2016 Deep learning in bioinformatics. *Briefings in Bioinformatics*, bbw068. (doi:[10.1093/bib/bbw068](https://doi.org/10.1093/bib/bbw068))
12. Parker JS *et al.* 2009 Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Journal of Clinical Oncology* **27**, 1160–1167. (doi:[10.1200/jco.2008.18.1370](https://doi.org/10.1200/jco.2008.18.1370))
13. Mayer IA, Abramson VG, Lehmann BD, Pietenpol JA. 2014 New Strategies for Triple-Negative Breast Cancer—Deciphering the Heterogeneity. *Clinical Cancer Research* **20**, 782–790. (doi:[10.1158/1078-0432.ccr-13-0583](https://doi.org/10.1158/1078-0432.ccr-13-0583))
14. TAN J, UNG M, CHENG C, GREENE CS. 2014 UNSUPERVISED FEATURE CONSTRUCTION AND KNOWLEDGE EXTRACTION FROM GENOME-WIDE ASSAYS OF BREAST CANCER WITH DENOISING AUTOENCODERS. In *Biocomputing 2015*, World Scientific Pub Co Pte Lt. (doi:[10.1142/9789814644730_0014](https://doi.org/10.1142/9789814644730_0014))
15. Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. 2013 Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013* pp. 411–418. Springer Nature. (doi:[10.1007/978-3-642-40763-5_51](https://doi.org/10.1007/978-3-642-40763-5_51))
16. Zurada J. In press. End effector target position learning using feedforward with error back-propagation and recurrent neural networks. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)* Institute of Electrical and Electronics Engineers (IEEE). (doi:[10.1109/icnn.1994.374637](https://doi.org/10.1109/icnn.1994.374637))
17. Wang S, Sun S, Li Z, Zhang R, Xu J. 2016 Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. (doi:[10.1101/073239](https://doi.org/10.1101/073239))
18. Spencer M, Eickholt J, Cheng J. 2015 A Deep Learning Network Approach to Protein Secondary Structure Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **12**, 103–112. (doi:[10.1109/tcbb.2014.2343960](https://doi.org/10.1109/tcbb.2014.2343960))
19. Wang S, Peng J, Ma J, Xu J. 2016 Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. *Scientific Reports* **6**, 18962. (doi:[10.1038/srep18962](https://doi.org/10.1038/srep18962))
20. Liu F, Li H, Ren C, Bo X, Shu W. 2016 PEDLA: predicting enhancers with a deep learning-based algorithmic framework. (doi:[10.1101/036129](https://doi.org/10.1101/036129))
21. Li Y, Chen C-Y, Wasserman WW. 2015 Deep Feature Selection: Theory and Application to Identify Enhancers and Promoters. In *Lecture Notes in Computer Science*, pp. 205–217. Springer Nature. (doi:[10.1007/978-3-319-16706-0_20](https://doi.org/10.1007/978-3-319-16706-0_20))
22. Klefogiannis D, Kalnis P, Bajic VB. 2014 DEEP: a general computational framework for predicting enhancers. *Nucleic Acids Research* **43**, e6–e6. (doi:[10.1093/nar/gku1058](https://doi.org/10.1093/nar/gku1058))
23. Quang D, Chen Y, Xie X. 2014 DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763. (doi:[10.1093/bioinformatics/btu703](https://doi.org/10.1093/bioinformatics/btu703))
24. Wallach I, Dzamba M, Heifets A. 2015 AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery.

25. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. 2016 Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. *Molecular Pharmaceutics* **13**, 2524–2530. (doi:[10.1021/acs.molpharmaceut.6b00248](https://doi.org/10.1021/acs.molpharmaceut.6b00248))
26. Wang Y, Zeng J. 2013 Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics* **29**, i126–i134. (doi:[10.1093/bioinformatics/btt234](https://doi.org/10.1093/bioinformatics/btt234))
27. Dhungel N, Carneiro G, Bradley AP. 2016 The Automated Learning of Deep Features for Breast Mass Classification from Mammograms. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* pp. 106–114. Springer Nature. (doi:[10.1007/978-3-319-46723-8_13](https://doi.org/10.1007/978-3-319-46723-8_13))
28. Dhungel N, Carneiro G, Bradley AP. 2015 Deep Learning and Structured Prediction for the Segmentation of Mass in Mammograms. In *Lecture Notes in Computer Science*, pp. 605–612. Springer Nature. (doi:[10.1007/978-3-319-24553-9_74](https://doi.org/10.1007/978-3-319-24553-9_74))
29. Zhu W, Lou Q, Vang YS, Xie X. 2016 Deep Multi-instance Networks with Sparse Label Assignment for Whole Mammogram Classification. (doi:[10.1101/095794](https://doi.org/10.1101/095794))
30. Zhu W, Xie X. 2016 Adversarial Deep Structural Networks for Mammographic Mass Segmentation. (doi:[10.1101/095786](https://doi.org/10.1101/095786))
31. Beaulieu-Jones BK, Greene CS. 2016 Semi-Supervised Learning of the Electronic Health Record for Phenotype Stratification. (doi:[10.1101/039800](https://doi.org/10.1101/039800))
32. Wang D, Khosla A, Gargya R, Irshad H, Beck AH. 2016 Deep learning for identifying metastatic breast cancer.
33. Lee CS, Baughman DM, Lee AY. 2016 Deep learning is effective for the classification of OCT images of normal versus Age-related Macular Degeneration. (doi:[10.1101/094276](https://doi.org/10.1101/094276))
34. In press.
35. Ohno-Machado L. 2011 Realizing the full potential of electronic health records: the role of natural language processing. *Journal of the American Medical Informatics Association* **18**, 539–539. (doi:[10.1136/amiajnl-2011-000501](https://doi.org/10.1136/amiajnl-2011-000501))
36. de Bruijn B, Cherry C, Kiritchenko S, Martin J, Zhu X. 2011 Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010. *Journal of the American Medical Informatics Association* **18**, 557–562. (doi:[10.1136/amiajnl-2011-000150](https://doi.org/10.1136/amiajnl-2011-000150))
37. Chalapathy R, Borzeshi EZ, Piccardi M. 2016 Bidirectional lstm-crf for clinical concept extraction.
38. Yoon H-J, Ramanathan A, Tourassi G. 2016 Multi-task Deep Neural Networks for Automated Extraction of Primary Site and Laterality Information from Cancer Pathology Reports. In *Advances in Big Data*, pp. 195–204. Springer Nature. (doi:[10.1007/978-3-319-47898-2_21](https://doi.org/10.1007/978-3-319-47898-2_21))
39. Lasko TA, Denny JC, Levy MA. 2013 Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data. *PLoS ONE* **8**, e66341. (doi:[10.1371/journal.pone.0066341](https://doi.org/10.1371/journal.pone.0066341))
40. Beaulieu-Jones BK, Greene CS. 2016 Semi-supervised learning of the electronic health record for phenotype stratification. *Journal of Biomedical Informatics* **64**, 168–178. (doi:[10.1016/j.jbi.2016.10.007](https://doi.org/10.1016/j.jbi.2016.10.007))
41. Miotto R, Li L, Kidd BA, Dudley JT. 2016 Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific Reports* **6**, 26094. (doi:[10.1038/srep26094](https://doi.org/10.1038/srep26094))
42. Ranganath R, Perotte A, Elhadad N, Blei D. 2016 Deep survival analysis.
43. Xiang A, Lapuerta P, Ryutov A, Buckley J, Azen S. 2000 Comparison of the performance of neural network methods and Cox regression for censored survival data. *Computational Statistics & Data Analysis* **34**, 243–257. (doi:[10.1016/s0167-9473\(99\)00098-5](https://doi.org/10.1016/s0167-9473(99)00098-5))
44. Katzman J, Shaham U, Bates J, Cloninger A, Jiang T, Kluger Y. 2016 Deep survival: A deep cox proportional hazards network.
45. Ranganath R, Tang L, Charlin L, Blei DM. 2014 Deep exponential families.
46. Bowman S. 2013 Impact of Electronic Health Record Systems on Information Integrity: Quality and Safety Implications. *Perspect Health Inf Manag* **10**, 1c.
47. Botsis T, Hartvigsen G, Chen F, Weng C. 2010 Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. *Summit on Translat Bioinforma* **2010**, 1–5.
48. Just BH, Marc D, Munns M, Sandefer R. 2016 Why Patient Matching Is a Challenge: Research on Master Patient

Index (MPI) Data Discrepancies in Key Identifying Fields. *Perspect Health Inf Manag* **13**, 1e.

49. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, Lai AM. 2014 A review of approaches to identifying patient phenotype cohorts using electronic health records. *Journal of the American Medical Informatics Association* **21**, 221–230. (doi:[10.1136/amiajnl-2013-001935](https://doi.org/10.1136/amiajnl-2013-001935))
50. WILEY LK, VANHOUTEN JP, SAMUELS DC, ALDRICH MC, RODEN DM, PETERSON JF, DENNY JC. 2016 STRATEGIES FOR EQUITABLE PHARMACOGENOMIC-GUIDED WARFARIN DOSING AMONG EUROPEAN AND AFRICAN AMERICAN INDIVIDUALS IN A CLINICAL POPULATION. In *Biocomputing 2017*, World Scientific Pub Co Pte Lt. (doi:[10.1142/9789813207813_0050](https://doi.org/10.1142/9789813207813_0050))
51. In press.
52. Jaggi R, Surender R. 2004 Regulation of junior doctors' work hours: an analysis of British and American doctors' experiences and attitudes. *Social Science & Medicine* **58**, 2181–2191. (doi:[10.1016/j.socscimed.2003.08.016](https://doi.org/10.1016/j.socscimed.2003.08.016))
53. Liapis CD. 2003 Effects of limited work hours on surgical training. *Journal of the American College of Surgeons* **196**, 662–663. (doi:[10.1016/s1072-7515\(03\)00097-8](https://doi.org/10.1016/s1072-7515(03)00097-8))
54. Gravenstein JS, Cooper JB, Orkin FK. 1990 Work and Rest Cycles in Anesthesia Practice *Anesthesiology* **72**, 737–742. (doi:[10.1097/0000542-199004000-00024](https://doi.org/10.1097/0000542-199004000-00024))
55. Firth-Cozens J, Greenhalgh J. 1997 Doctors' perceptions of the links between stress and lowered clinical care. *Social Science & Medicine* **44**, 1017–1022. (doi:[10.1016/s0277-9536\(96\)00227-4](https://doi.org/10.1016/s0277-9536(96)00227-4))
56. Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, Jensen PB, Jensen LJ, Brunak S. 2014 Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nature Communications* **5**. (doi:[10.1038/ncomms5022](https://doi.org/10.1038/ncomms5022))
57. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. 2016 Deepr: A convolutional net for medical records.
58. Pham T, Tran T, Phung D, Venkatesh S. 2016 DeepCare: A deep dynamic memory model for predictive medicine.
59. Russakovsky O *et al.* 2014 ImageNet large scale visual recognition challenge.
60. Tramèr F, Zhang F, Juels A, Reiter MK, Ristenpart T. 2016 Stealing machine learning models via prediction apis.
61. Dwork C, Roth A. 2013 The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science* **9**, 211–407. (doi:[10.1561/04000000042](https://doi.org/10.1561/04000000042))
62. Simmons S, Sahinalp C, Berger B. 2016 Enabling Privacy-Preserving GWASs in Heterogeneous Human Populations. *Cell Systems* **3**, 54–61. (doi:[10.1016/j.cels.2016.04.013](https://doi.org/10.1016/j.cels.2016.04.013))
63. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, Zhang L. 2016 Deep learning with differential privacy. (doi:[10.1145/2976749.2978318](https://doi.org/10.1145/2976749.2978318))
64. Beaulieu-Jones BK, Greene CS. 2016 Reproducible Computational Workflows with Continuous Analysis. (doi:[10.1101/056473](https://doi.org/10.1101/056473))
65. Chen L, Cai C, Chen V, Lu X. 2015 Trans-species learning of cellular signaling systems with bimodal deep belief networks. *Bioinformatics* **31**, 3008–3015. (doi:[10.1093/bioinformatics/btv315](https://doi.org/10.1093/bioinformatics/btv315))
66. Van Valen DA *et al.* 2016 Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments. *PLOS Computational Biology* **12**, e1005177. (doi:[10.1371/journal.pcbi.1005177](https://doi.org/10.1371/journal.pcbi.1005177))
67. Ronneberger O, Fischer P, Brox T. 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science*, pp. 234–241. Springer Nature. (doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28))
68. Buggenthin F *et al.* 2017 Prospective identification of hematopoietic lineage choice by deep learning. *Nature Methods* (doi:[10.1038/nmeth.4182](https://doi.org/10.1038/nmeth.4182))
69. Eulenberg P, Koehler N, Blasi T, Filby A, Carpenter AE, Rees P, Theis FJ, Wolf FA. 2016 Deep Learning for Imaging Flow Cytometry: Cell Cycle Analysis of Jurkat Cells. (doi:[10.1101/081364](https://doi.org/10.1101/081364))
70. Pawlowski N, Caicedo JC, Singh S, Carpenter AE, Storkey A. 2016 Automating Morphological Profiling with Generic Deep Convolutional Networks. (doi:[10.1101/085118](https://doi.org/10.1101/085118))
71. Caicedo JC, Singh S, Carpenter AE. 2016 Applications in image-based profiling of perturbations. *Current Opinion in Biotechnology* **39**, 134–142. (doi:[10.1016/j.copbio.2016.04.003](https://doi.org/10.1016/j.copbio.2016.04.003))
72. Bougen-Zhukov N, Loh SY, Lee HK, Loo L-H. 2016 Large-scale image-based screening and profiling of cellular

- phenotypes. *Cytometry Part A* **91**, 115–125. (doi:[10.1002/cyto.a.22909](https://doi.org/10.1002/cyto.a.22909))
73. Grys BT, Lo DS, Sahin N, Kraus OZ, Morris Q, Boone C, Andrews BJ. 2016 Machine learning and computer vision approaches for phenotypic profiling. *The Journal of Cell Biology* **216**, 65–71. (doi:[10.1083/jcb.201610026](https://doi.org/10.1083/jcb.201610026))
74. Arvaniti E, Claassen M. 2016 Sensitive detection of rare disease-associated cell subsets via representation learning. (doi:[10.1101/046508](https://doi.org/10.1101/046508))
75. Angermueller C, Lee H, Reik W, Stegle O. 2016 Accurate prediction of single-cell DNA methylation states using deep learning. (doi:[10.1101/055715](https://doi.org/10.1101/055715))
76. Shaham U, Stanton KP, Zhao J, Li H, Raddassi K, Montgomery R, Kluger Y. 2016 Removal of batch effects using distribution-matching residual networks.
77. Rho M, Tang H, Ye Y. 2010 FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research* **38**, e191–e191. (doi:[10.1093/nar/gkq747](https://doi.org/10.1093/nar/gkq747))
78. Chudobova D *et al.* 2015 Influence of microbiome species in hard-to-heal wounds on disease severity and treatment duration. *The Brazilian Journal of Infectious Diseases* **19**, 604–613. (doi:[10.1016/j.bjid.2015.08.013](https://doi.org/10.1016/j.bjid.2015.08.013))
79. Bengio Y, Boulanger-Lewandowski N, Pascanu R. 2012 Advances in optimizing recurrent networks.
80. Ithapu VK, Singh V, Okonkwo OC, Chappell RJ, Dowling NM, Johnson SC. 2015 Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. *Alzheimer's & Dementia* **11**, 1489–1499. (doi:[10.1016/j.jalz.2015.01.010](https://doi.org/10.1016/j.jalz.2015.01.010))
81. Swamidass SJ, Azencott C-A, Lin T-W, Gramajo H, Tsai S-C, Baldi P. 2009 Influence Relevance Voting: An Accurate And Interpretable Virtual High Throughput Screening Method. *Journal of Chemical Information and Modeling* **49**, 756–766. (doi:[10.1021/ci8004379](https://doi.org/10.1021/ci8004379))
82. Kearnes S, Goldman B, Pande V. 2016 Modeling industrial admet data with multitask networks.
83. Altae-Tran H, Ramsundar B, Pappu AS, Pande V. 2016 Low data drug discovery with one-shot learning.
84. Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR. 2016 Protein-ligand scoring with convolutional neural networks.
85. Gupta S, Agrawal A, Gopalakrishnan K, Narayanan P. 2015 Deep learning with limited numerical precision.
86. Courbariaux M, Bengio Y, David J-P. 2014 Training deep neural networks with low precision multiplications.
87. Sa CD, Zhang C, Olukotun K, Ré C. 2015 Taming the wild: A unified analysis of hogwild!-style algorithms.
88. Hubara I, Courbariaux M, Soudry D, El-Yaniv R, Bengio Y. 2016 Quantized neural networks: Training neural networks with low precision weights and activations.
89. Ba LJ, Caruana R. 2013 Do deep nets really need to be deep?
90. Hinton G, Vinyals O, Dean J. 2015 Distilling the knowledge in a neural network.
91. Raina R, Madhavan A, Ng AY. 2009 Large-scale deep unsupervised learning using graphics processors. In *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09* Association for Computing Machinery (ACM). (doi:[10.1145/1553374.1553486](https://doi.org/10.1145/1553374.1553486))
92. In press.
93. Seide F, Fu H, Droppo J, Li G, Yu D. 2014 On parallelizability of stochastic gradient descent for speech DNNs. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* Institute of Electrical and Electronics Engineers (IEEE). (doi:[10.1109/icassp.2014.6853593](https://doi.org/10.1109/icassp.2014.6853593))
94. Hadjis S, Abuzaid F, Zhang C, Ré C. 2015 Caffe con troll: Shallow ideas to speed up deep learning.
95. Edwards C. 2015 Growing pains for deep learning. *Communications of the ACM* **58**, 14–16. (doi:[10.1145/2771283](https://doi.org/10.1145/2771283))
96. Su H, Chen H. 2015 Experiments on parallel training of deep neural network using model averaging.
97. Li M, Zhang T, Chen Y, Smola AJ. 2014 Efficient mini-batch training for stochastic optimization. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, Association for Computing Machinery (ACM). (doi:[10.1145/2623330.2623612](https://doi.org/10.1145/2623330.2623612))
98. In press.

99. Chen Y, Li Y, Narayan R, Subramanian A, Xie X. 2016 Gene expression inference with deep learning. *Bioinformatics* **32**, 1832–1839. (doi:[10.1093/bioinformatics/btw074](https://doi.org/10.1093/bioinformatics/btw074))
100. Gómez-Bombarelli R, Duvenaud D, Hernández-Lobato JM, Aguilera-Iparraguirre J, Hirzel TD, Adams RP, Aspuru-Guzik A. 2016 Automatic chemical design using a data-driven continuous representation of molecules.
101. Hamanaka M, Taneishi K, Iwata H, Ye J, Pei J, Hou J, Okuno Y. 2016 CGBVS-DNN: Prediction of Compound-protein Interactions Based on Deep Learning. *Molecular Informatics* **36**, 1600045. (doi:[10.1002/minf.201600045](https://doi.org/10.1002/minf.201600045))
102. Chetlur S, Woolley C, Vandermersch P, Cohen J, Tran J, Catanzaro B, Shelhamer E. 2014 CuDNN: Efficient primitives for deep learning.
103. Chen W, Wilson JT, Tyree S, Weinberger KQ, Chen Y. 2015 Compressing neural networks with the hashing trick.
104. Lacey G, Taylor GW, Areibi S. 2016 Deep learning on fpgas: Past, present, and future.
105. Dean J, Ghemawat S. 2008 MapReduce. *Communications of the ACM* **51**, 107. (doi:[10.1145/1327452.1327492](https://doi.org/10.1145/1327452.1327492))
106. Low Y, Bickson D, Gonzalez J, Guestrin C, Kyrola A, Hellerstein JM. 2012 Distributed GraphLab. *Proceedings of the VLDB Endowment* **5**, 716–727. (doi:[10.14778/2212351.2212354](https://doi.org/10.14778/2212351.2212354))
107. In press.
108. In press.
109. Moritz P, Nishihara R, Stoica I, Jordan MI. 2015 SparkNet: Training deep networks in spark.
110. Meng X *et al.* 2015 MLlib: Machine learning in apache spark.
111. In press.
112. In press.
113. In press.
114. In press.
115. Sun S, Chen W, Liu T-Y. 2016 Ensemble-compression: A new method for parallel training of deep neural networks.
116. In press.
117. In press.
118. Schatz MC, Langmead B, Salzberg SL. 2010 Cloud computing and the DNA data race. *Nature Biotechnology* **28**, 691–693. (doi:[10.1038/nbt0710-691](https://doi.org/10.1038/nbt0710-691))
119. Muir P *et al.* 2016 The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biology* **17**. (doi:[10.1186/s13059-016-0917-0](https://doi.org/10.1186/s13059-016-0917-0))
120. Stein LD. 2010 The case for cloud computing in genome informatics. *Genome Biology* **11**, 207. (doi:[10.1186/gb-2010-11-5-207](https://doi.org/10.1186/gb-2010-11-5-207))
121. Krizhevsky A. 2014 One weird trick for parallelizing convolutional neural networks.
122. Armbrust M *et al.* 2010 A view of cloud computing. *Communications of the ACM* **53**, 50. (doi:[10.1145/1721654.1721672](https://doi.org/10.1145/1721654.1721672))