This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2894713, IEEE Journal of Biomedical and Health Informatics

Deep Learning: Current and Emerging Applications in Medicine and Technology

A. Akay, *Member, IEEE*, H. Hess, Senior *Member, IEEE*

***Abstract*—Machine learning is enabling researchers to analyze and understand increasingly complex physical and biological phenomena in traditional fields such as biology, medicine, and engineering and emerging fields like synthetic biology, automated chemical synthesis, and bio-manufacturing. These fields require new paradigms towards understanding increasingly complex data and converting such data into medical products and services for patients. The move towards deep learning and complex modeling is an attempt to bridge the gap between acquiring massive quantities of complex data, and converting such data into practical insights. Here, we provide an overview of the field of machine learning, its current applications and needs in traditional and emerging fields, and discuss an illustrative attempt at using deep learning to understand swarm behavior of molecular shuttles.**

Index Terms — Deep learning, machine learning, neural networks, informatics

I. INTRODUCTION

The exponential growth of big and increasingly bigger data sets both in traditional and emerging fields has raised important questions: 1) How can increasingly complex phenomena in these data be correctly interpreted and analyzed? 2) How can the aforementioned data sets be converted into a deeper understanding of complex phenomena? and 3) How can the achieved understanding be converted into practical applications in medicine, ranging from rapid and precise diagnosis, over intelligent treatment, to targeted prevention? As we face the interacting pressures of rising chronic diseases, ageing populations, and dwindling resources, a paradigm shift towards intelligently extracting, analyzing, interpreting, and understanding increasingly complex data is required. The rapidly evolving field of machine learning is key to this paradigm shift.

Machine Learning (ML) has expanded both in scope and complexity as computing power increased. Graphics Processing Units (GPUs) and bigger storage spaces became more widely available, resulting in powerful parallel data processing. ML techniques have also evolved, and are now used to understand large data sets in medical image processing, speech recognition, computer vision, language processing, bioinformatics, and drug design [1]. Previous studies have investigated ML in bioinformatics [2], biology and medicine [3], computational biology [4,5], biomedicine [6], and super-resolution imaging [7].

The application of ML (and later, deep learning) has expanded into emerging fields including, but not limited to, bio-manufacturing [8], automated organic synthesis [9], material and molecular modeling [10], medical robotics [11], automated drug discovery [12], biological networks [13], and automated diagnosis [14]. Traditional fields are evolving in response to the development of ML techniques: for example, in drug design, as pressure is mounting on pharmaceutical companies to rapidly test promising molecules, engage in clinical trials, and place drugs on the market in a timely matter, ML can potentially aid in reducing the timeframe for drug discovery and trials. In the diagnosis field, rapid and precise information about the nature of a patient's health problem can be obtained with the help of machine learning algorithms.

Of particular interest to the authors is the design and study of molecular nano- and micro-robots [15-19], and their use in medical applications, including minimally invasive surgery [20], tissue engineering [21], and targeted drug delivery [22]. Despite rapid advances, various challenges remain with respect to design [23], materials selection [24], and the control of location and orientation [25].

The goal of this position paper is to provide background information about ML for the non-expert and discuss recent advances in ML techniques, present current applications in traditional and emerging fields, assess the current needs that ML can respond to, and to provide an example of how machine learning techniques can be used in molecular robotics, specifically in studying and understanding swarm behavior of molecular shuttles. Finally, we aim to address possible future trends, highlight both current challenges in implementing this evolving field and opportunities for advancing it.

II. TECHNOLOGY BACKGROUND

A. Artificial Intelligence, Machine Learning, and Deep Learning

The concept of artificial intelligence (AI) started at a computer science conference at Dartmouth College in 1956 [26]. The vision was to develop machines that can mimic human intelligence and the ability to learn. The limitation was that it could only perform simple, specific tasks (dubbed 'Narrow AI') [27]. To advance further, algorithms needed to be developed that could analyze data structures, learn from said data structures, and determine what steps to take based on the nature of the data structures. This engendered the concept of ML, whereby the algorithm sifts through massive amounts of data, 'learns' from it, and responds to it. Despite these advances, ML algorithms were still limited by relying extensively on data representations, known as 'features,' that required human expertise (known as 'domain expertise') to

Altug Akay, PhD is with the Department of Biomedical Engineering, Columbia University, 116th Street & Broadway, New York, NY, 10027, USA (ara2180@columbia.edu).

Henry Hess, PhD, is with the Department of Biomedical Engineering, Columbia University, 116th Street & Broadway, New York, NY, 10027, USA (hhess@columbia.edu).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2894713, IEEE Journal of Biomedical and Health Informatics

direct which features the algorithm should look for [28]. These algorithms have shown limited success to achieve optimal AI performance and to process highly complex data. Andrew Ng proposed to make artificial neural network (ANN) structures complex by increasing the number of hidden layers and neurons in the layers, and the use of massive data [29]. The demands placed on the AI community (more robust algorithms that required very little domain expertise), combined with the exponential growth in data sets and greater availability of graphics processing units (GPUs), resulted in the development of deep learning (DL) algorithms. Deep learning is a branch of ML that derives its origins from the concept of ANNs.

An ANN is a computational platform composed of inter-connected nodes ('artificial neurons') that resemble, and mimic, the brain's neuronal functions. The connections between the nodes ('edges') strengthen or weaken as the learning process progresses. Figure 1 shows a traditional ANN containing an input layer (that senses and detects signals within the environment), a hidden layer (that processes the signals sent by the input layer), and an output layer (the response to a signal or stimulus). Note that each neuron in both the hidden and output layers assigns a weight coefficient to its input. Therefore, the final output of neural network is determined by the contributions of each weight coefficients [30]. ANNs 'learn' the same way humans learn: by interacting with, and responding to, various stimuli within a local environment.

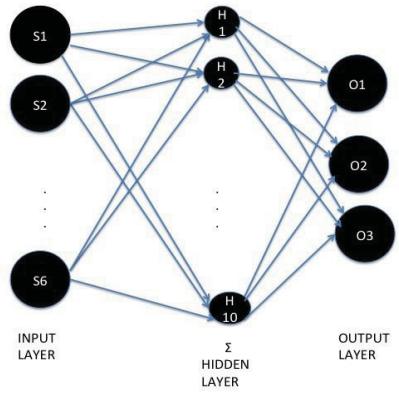

Figure 1: An ANN using a single hidden layer. S1-S6 are the nodes in the input layer, and O1-O3 are the nodes in the output layer.

Learning algorithms use statistics and analytics to enable computers to improve a given task's performance ('learn') using data without being given specific instructions [31]. They use supervised and unsupervised learning tasks to match, and identify, inputs. Supervised learning uses labelled (categorized) data that are fed into the input layer, which produces an output (the desired category with the highest score). Supervised ANNs are useful when the outcome of the feature vector is known. As in the case of regression, a feature vector is input into the net and through backward and forward propagation the optimal weights are found to train the network over the training examples. The network is then tested on the testing set. The network can then be used for prediction and accuracy measured.

Although supervised ANNs have been successfully used in many biomedical applications, including DNA motif discovery [32], medical diagnosis [33], cancer identification and gene classification using DNA microarray gene expression patterns [34], and drug discovery [35], an ANN may exhibit inconsistent prediction performance and long training times, especially for architectures with many layers requiring large data sets. Architectures need to be carefully selected and fine-tuned to achieve the best performance. The selection of the number of layers, the number of nodes for each layer, and the activation functions for nodes in each layer are critical parameters that need to be determined for each application.

Unsupervised learning uses the same method as ANNs and can be considered as a less complex version of a DL network. Unsupervised learning is related to logistic regression through its (usually) sigmoid output layer. The self-organizing map (SOM) and adaptive resonance theory (ART) are widely used in unsupervised learning algorithms. The SOM is a topographic organization in which nearby locations in the map represent inputs with similar properties.

The ART model allows the number of clusters to vary with problem size and lets the user control the degree of similarity between members of the same clusters by means of a user-defined constant called the vigilance parameter. ART networks are also used for many pattern recognition tasks, such as automatic target recognition and seismic signal processing. Unsupervised ANNs have been used in medicine and biology, including medical data clustering [36], classical biomedical markers analysis [37], and tumor characterization [38]. The unsupervised ANN needs to label and group the raw data without any prior knowledge of patterns in the input data. This may lead to less accurate results. Finally, the output may not be ascertained since the number of classes is unknown and needs to be determined by the ANN-generated results.

A deep neural network (DNN) uses the same architecture as an ANN, except with more hidden layers between the input and output layers. The output values are sequentially computed within the network's hidden layers, with each input vector comprising the previous layer's output vector to produce a weighted sum. The result is a higher, abstract data representation that is attentive to small details while ignoring irrelevant information. DNNs have been widely been used in protein structure prediction research [39-42], biomedical imaging, with an emphasis on anomaly classification [43-45], segmentation [46], recognition [47,48], and brain decoding [49,50]. In biomedical signal processing, DNNs have been using EEG signals for motor skill [51], emotional classification [52], and anomaly detection using ECG signals [53,54].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2894713, IEEE Journal of Biomedical and Health Informatics

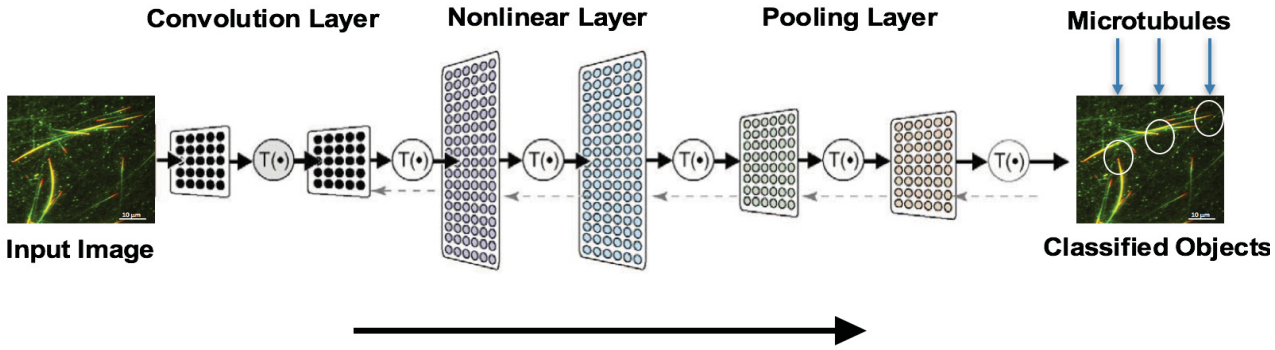

Fig. 2: A Convolutional Neural Network applied to the task of identifying objects in an image.

Convolutional Neural Networks (CNNs) and Recursive Neural Networks (RNNs) are the most common DNNs. A CNN, shown in Figure 2, is composed of convolution, non-linear, and pooling layers [55]. The convolution layers obtain local weighted sums ('feature maps') at every layer by computing filters that are repeatedly applied across the entire data set to improve training efficiency. The non-linear layers then increase the feature maps' non-linear properties. Finally, the pooling layers performs sub-sampling of non-overlapping regions in feature maps to enable the CNN to aggregate local features to identify complex features [6]. CNNs process multiple data types, particularly two-dimensional images. An RNN processes inputs one element at a time, maintaining a vector within its hidden units that contain information from all of the preceding element's sequence. The final output depends on the previous inputs. RNNs are ideal for sequential information processing, particularly natural language processing [56] and audio recordings [57].

CNNs have been used for gene sequence data [58], transcription-binding site predictions [59,60], anomaly detection and classification using CT image datasets [61], malignant tumor growth prediction [62], chemotherapy response prediction [63], a finger-joint detection platform to examine bone age and growth disorders [64,65], and brain decoding, in which extracted features were converted into 2-D pixel colors, to predict seizures [66]. RNNs have been used for protein structure prediction [67], gene expression regulation [68], protein classification [69], and seizure prediction [70].

Novel architectures have recently emerged to tackle increasingly complex biomedical data derived from emerging interdisciplinary fields. Capsule networks [71,72] use the CNN method to visualize, and direct, different stimuli to specialist capsules (modules). Rather than adding hidden layers, a capsule network adds layers within each hidden layer. Such a network can lead to a better understanding of disease mechanisms.

Multi-task learning exploits the ability to simultaneously learn multiple (but similar) tasks within a model: it has been applied to predict drug toxicity and sensitivity in cancer cell lines [73-75]. Other exciting research avenues can include looking for ways to minimize deleterious drug side effects (termed 'off-target effects').

A transfer learning neural network architecture is based on the logic that, since biological systems share similar characteristics, data from one system can help to understand the other [76]. It allows a model to be re-used from one task to another different (but similar) task.

Deep Spatio-Temporal Neural Networks (DST-NNs) learn multi-dimensional outputs through progressive refinement, which looks at local correlations through input feature components per layer: the spatial layer, where the original inputs are used in every layer, and the temporal features, which change to end up in the upper layers [77].

Multi-dimensional RNNs use RNN capabilities to treat non-sequential, multi-dimensional data as sequential data groups. In one study, a 2-D data set from four contexts under different data processing orders are reflected in four hidden units that are connected to a single output layer, with results that consider all possible contexts [78].

Convolutional auto-encoders (CAEs) use both auto-encoders (AE) and CNNs to combine good hierarchical representations of spatial information data while being regularized via unsupervised training. Reconstruction error is minimized using both an encoder (to extract feature vectors) and decoder (to recreate the data from the feature vectors) when training AEs. The principles of CNN can mimic an AE in the sense that encoders (deconvolution) and decoders (un-pooling) are integrated as a CAE and trained as an AE [79,80].

B. Molecular Robotics and Swarming

In this section, we give a short introduction to an application area of particular interest to our work, molecular robotics. One specific problem is how molecular robots can be induced to exhibit swarming behavior. Swarming is a phenomenon wherein large numbers of individuals organize into a coordinated motion. The ability of fish schools, insect swarms, or starling murmurations to shift shape as one, and coordinate their motion in space, has been extensively studied because of their implications for the evolution of social cognition, collective animal behavior, and artificial life [67],[81]-[82]. Understanding swarm mechanisms and operational principles can provide novel approaches for developing swarm formation control, autonomous agent distributed/cooperative control, and robot coordination [83]-[85].

Interest has grown in using foraging and swarming-based biomimicry for engineering applications such as distributed optimization, collective robotics, satellite clusters, mobile sensor networks, and autonomous aerial/underwater vehicles [86],[87], which has further motivated research into swarming

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2894713, IEEE Journal of Biomedical and Health Informatics

behavior and its applications. Intelligent, agent-based simulations for swarm analysis has greatly improved our ability to perform tasks such as image processing, pattern and speech recognition, and language interpretation using advanced computational techniques such as DNN [2] and intelligent agent swarms [86],[87]. Simulating agents in models allows easy parameter modification, voluminous data generation, full experiment reproducibility, and easier identification of complex phenomena's underlying dynamics [88]. Exploiting swarm behavior is a promising approach to controlling molecular robots, due to their typically large numbers combined with limited communication and processing abilities.

Molecular robotics takes robots, which are artifacts exhibiting intelligent behaviors by sensing-processing-actuating cycles [90], to the molecular level. The concept stresses the need to borrow mechanisms from living systems and applying them to robots [89]. Among the mechanisms include designing systems that encourage and engender desirable behaviors depending on the specific task at hand (foraging, swarming, autonomous decision-making, etc.) [90]. Powering molecular robots continues to be a challenge, and biomolecular motors have the key advantage of unmatched energy conversion efficiency [91].

Biomolecular motors, such as the motor proteins myosin and kinesin, convert the chemical energy stored in ATP into mechanical work as they move along their cytoskeletal filaments (actin filaments and microtubules, respectively) [92]. Biomolecular motors can be used as highly functional off-the-shelf components to construct molecular robots. An example for such a hybrid robot combining biological and synthetic molecular components is a kinesin-powered molecular shuttle [93],[94] (Fig. 3a): surface-adhered kinesin motor proteins propel microtubules which are functionalized with linkers to endow them with the ability to capture and carry cargo [95]. Depending on the ATP concentration and temperature, the shuttles move at speeds up to 1 μm/s [96]. Millions of shuttles can operate in parallel, and their interactions can be programmed [97]-[99]. The operation can be visualized using fluorescence microscopy if the microtubules are fluorescently labeled (Fig. 3b), and automated methods relying on image segmentation by thresholding to identify and track the microtubules have been developed [100]. Interactions between shuttles, or gliding actin filaments and microtubules in general, engender complex emergent patterns [101]-[103].

Machine learning can contribute to molecular robotics at two distinct levels: it can be applied to analyze microscopy images of molecular robots, where it can facilitate and improve the identification of the robots in crowded images with a high background. This is a classic image processing task that has been addressed with ML solutions in various fields. The second level is to "learn" the swarm behavior, thereby enabling its control, similar to the way a herding dog learns and controls the behavior of a flock of sheep. It is also desirable to infer the rules engendering the behavior [104]. An illustration of a possible use of DL to model molecular shuttle movement and organization will be illustrated. However in the next section, we first highlight how ML has been used in different applications in traditional and emerging fields.

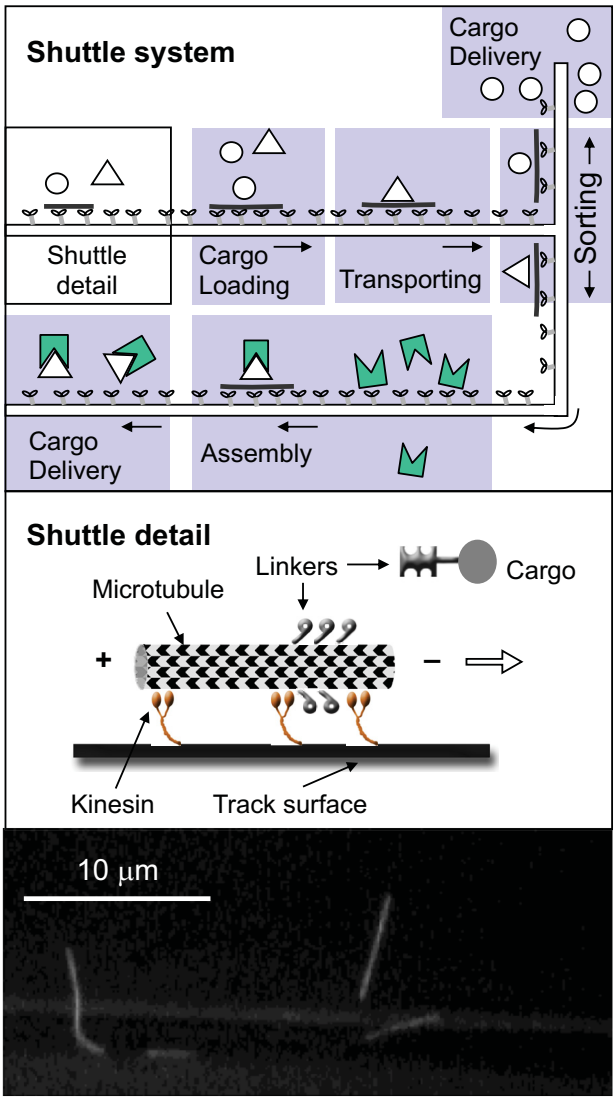

Fig. 3: Top and Middle: Molecular shuttles as conceived by Vogel and Howard utilize kinesins adhered in tracks and controllably activated with defined concentrations of ATP to propel cargo-carrying microtubules. Bottom: Gliding, fluorescently labeled microtubules can be imaged by fluorescence microscopy and appear as bright lines with variable shapes as they interact with guiding structures. Adapted with permission from [209]. Copyright 2003 American Chemical Society.

III. Current Applications

Machine learning and DL have been extensively used across traditional and emerging fields. Below is an overview of fields that have taken advantage of current ML and DL capabilities.

A. Diagnosis of Diabetic Retinopathy

A research group at Google automated the diagnosis of the severity of referable diabetic retinopathy using a DL algorithm, thus potentially aiding clinicians in rapidly diagnosing and treating this disease [14]. The algorithm obviated the need for domain expertise, resulting in significantly reducing the time it took to learn what to look for, and, by extension, closely match clinician's classification of the severity of the disease.

B. Microrobot Localization

Sitti et al., used an advanced CNN architecture to enable real-time localization of a minimally-invasive and therapeutic endoscopy robot for the gastrointestinal (GI) tract [11]. This approach enabled the robot to establish its position and orientation within a frame of reference. They have merged the CNN architecture with sophisticated sensor platforms to enable precise, real-time robot localization within the GI tract. This can open new research avenues towards fusing molecular robot technologies with CNN (and other) architectures that can aid in precise diagnosis and interventions, and for molecular shuttle localization *in vivo*.

C. Computational Biology

Computational biology uses computers to model, and understand, the essential functions of life, with particular emphasis on representing and simulating biological systems [105]. The quantitative increase of biological data, both in size and complexity, has initially engendered the use of traditional ML techniques that required extensive domain expertise and manual interventions to adjust the neural network for precisely measuring specific phenomena. Deep learning requires minimal domain expertise and can 'learn' which features to look for from available data sets.

In regulatory genomics, where specific regulatory mechanisms and non-coding transcription factors within the genome are studied, the main challenges are mapping the effects of mutations within a population, especially of rare mutations that can result in limited sample sizes, and predicting DNA sequence because of its multi-layered abstraction among its variations, interactions with different regulatory mechanisms, and how it effects one region (or several regions) of interest. DNNs have been used in predicting splicing [106],[107], and DNA-binding protein specificities [59]. Alipanahi et al. have expanded the use of CNNs to predict the effects of mutations, including the visualization of all possible mutations within a sequence, which, in turn, can lead to identifying single nucleotide variants (SNVs) that contribute to various mutation(s) [59]. Researchers have also started merging different DL architectures to predict multiple trait development across genomes and how they affect each other: Zhou and Troyanskaya used multiple architectures to capture multiple genomic sequences [108]: Dahl, et al. used multi-task neural networks to predict multiple chromatin states within regions of interest, in parallel [109].

In biological image analysis, CNNs are commonly used for classification, feature detection and extraction, and pattern recognition. Adding more hidden layers in the CNN results in detecting more abstract features [62]. Image analysis architectures and algorithms can potentially be expanded to include multiple, stacked images, either from multi-scale, or multi-modal imaging techniques, and trained to look for increasingly complex features. This could greatly aid in developing more precise diagnoses in clinical settings, especially from patients suffering from multiple ailments.

The field has recently expanded to include analyzing entire cells, and cell populations, to understand their interactions with each other, and their environment. Such efforts have resulted in architectures that have merged classification with segmentation tasks to model entire full-resolution fungal microscopy images [110] and quantify bacteria colonies on agar plates [111]. These technologies can greatly expand our understanding of cell population behavior, both in healthy and diseased states. These data can also serve as a reference for possible rapid clinical diagnosis of patients suffering from similar ailments.

D. Bioinformatics

Bioinformatics uses computational tools to discern biological data. With the increasing size and complexity of data in the '-omics' field, pressure is mounting to develop sophisticated tools that can quickly and efficiently convert data into healthcare products and services. Deep learning architectures have contributed immensely to this field. It is expected that, as the field continues to evolve and more complex data is gathered, more sophisticated architectures and algorithms will be needed to analyze and interpret various inter-connected phenomena.

Large amounts of raw sequential data are used (DNA and RNA sequences, for example). RNNs are used because of their ability to process sequential information. Among avenues of interest are protein structure prediction and classification that can aid in deciphering under what circumstances proteins change structurally [39-42], gene expression regulation to decipher the circumstances where genes turn 'on' or 'off'[112]-[122], and anomaly classification and detection [123]-[125] to ascertain when, for example, the circumstances lead to the development of malignant tumors.

In biomedical imaging, data sources come from MRI [126],[127], radio-graphs [128],[129], PET [130], and histopathology [131] images. CNNs are used since they mimic the human visual cortex that looks at the general features in the environment (mimicking photoreceptors in the retina), and are designed to study variable and intricate features within an image, or set of images. Among avenues of interest are anomaly detection [61],[132]-[137], segmentation [46],[138]-[143], recognition [144]-[147], and brain behavior 'decoding' [50].

In biomedical signal processing, researchers primarily use sequential recorded electrical activity from the human body [115]. Sources include EEG [148], ECoG [149], ECG [150], EMG [151], and EOG [152]. RNNs are used because biomedical signals are sequential data. Avenues of research are, for example, brain decoding [153]-[162] and anomaly classification for diagnosis of neuro-degenerative disorders and heart ailments [163]-[168].

E. Network Biology

Network biology studies biomolecule interactions that contribute to the structure and function of living cells, the reconstruction (and analysis) of large-scale endogenous biological networks, and the design (and construction) of small-scale synthetic gene networks [13]. Machine learning algorithms can exploit, and merge, disparate biological datasets to develop increasingly complex, and stacked, models to study interactions of various phenomena from within cells to between cells in tissues and organs in both healthy and diseased states.

Disease biology can be understood using network biology principles (discovering networks and sub-networks of the interactions required for the commencement of a disease phase)

[169-175], and using ML to identify and comprehend disease mechanisms from the surface to deeper levels. There has been progress [176-179], but work remains, specifically at understanding how disrupting biological process networks can engender diseases. Capsule networks [71,72] use similar approaches as CNNs to visualize different, and direct, stimuli to specialist capsules (modules). A capsule network adds layers within each hidden layer, rather than adding more layers.

Drug discovery, in the context of network biology and ML, requires characterizing the actions of compounds, predicting both on-target and off-target effects of said compounds, and the merging of multiple drugs to attack complex diseases [180]. Large multi-omics datasets, when merged with genotype archives and databases such as PubChem [181], DrugBank [182], and ZINC [183], can potentially be integrated to develop increasingly predictive complex disease outcomes, and their interactions within the human body can be studied and modeled. Multi-task neural networks exploit the ability to simultaneously learn multiple, yet similar, tasks within a model: this method has been applied to predict drug toxicity, and drug sensitivity, in cancer cell lines [73]-[75]. Other exciting avenues of research can include looking for ways to minimize the deleterious side effects of drugs ('off-target effects').

The human microbiome is composed of micro-organisms within the human body. These micro-organisms play essential roles in the body's health and development. Studying micro-organism interactions, and the evolution of their exchange networks, can lead to new insights towards the interaction between these micro-organisms and their hosts. A transfer learning neural network is the method of choice for the limited data surrounding these 'meta-metabolic networks' [76].

Synthetic biology seeks to develop artificial biological components for research, medical, and industrial applications. One exciting avenue of research is synthetic gene circuit development [184]-[188]. The main challenges are understanding design principles and how multiple components (at multiple levels) interact with each other: it demands a neural network that combines the CNN's ability to detect abstract and minute details to understand the organization of regulatory units in synthetic networks, with the RNN's ability to sequentially predict the outcomes of different components to ensure the system's stability as its complexity increases.

F. Drug Discovery

Virtual screening (VS) uses computers to peruse molecular libraries to identify molecules that can successfully bind to receptor sites to be used for drug development [31]. The need for automatic search methods that successfully identify promising molecules has resulted in the development of ML applications used for screening and matching molecules.

Support Vector Machines (SVMs) classify and rank properties, in the form of binary outcomes [32-39]. SVMs have been applied towards ranking compounds to predict activity levels of their binding sites [40], along with novel hybrid techniques for classification purposes [41,42].

Decision Trees (DTs) outline decisions and their consequences in the form of a tree. It takes the form of an upside-down tree that branches out into different branches (decisions) until the leaves (decisions, or nodes) have been developed. DTs have been used to develop binary property classifications ranging from drug permeation to metabolic stability (or lack thereof) [43-51].

Naive Bayesian classifier uses Bayes' theorem that describes the probability of an event that has originated from two (or more) causes [52]. The classifier provides a framework where an established opinion, or rule, changes when new knowledge (data) enters into the equation, and has been used with ML techniques for similarity-based clustering [53-55].

G. Automating Small Molecule Synthesis

The move from manual small-molecule synthesis to automated small-molecule synthesis is an attempt to create machines that can mass-produce multiple small molecules of different types. Challenges remain towards automating small molecule synthesis for mass-production, with two camps attempting to solve the problem: one group wants to use optimized machines to customize the synthesis of one specific small molecule, while another group wants to use general machines to synthesize many different small molecule types [189]. The strategies of the latter group include automating the retrosynthetic analysis process (a three-step process whereby one first converts all possible reactions to deconstruct a target area into a set of starting components, decides the best route for deconstructing the target, then manually validates the synthesis in a laboratory [56-60]) using ML techniques [61,62]. Machine learning has also merged with online reaction monitoring and self-optimization procedures to automate synthesis development for both customized and generalized paradigms [190]. Robot-mediated experimental discovery and testing has also used ML for both laboratory automation [191] and autonomous prediction of optimal molecular structures [192].

H. Bio-manufacturing

Bio-manufacturing uses biological products to produce bio-materials for consumer and industrial applications. Genome-scale modeling (GSM) has been used to predict microbial factory performance and identify potentially useful gene targets [9]. Machine learning, along with data mining, and genomic modeling, can aid in deciphering complex intra- and inter-cellular phenomena to develop prediction scenarios to improve medical, food, and industrial yields [193]-[195]. Developing optimal strains for industrial applications requires understanding cellular and genetic metabolism and regulation, respectively. Challenges remain, including accounting for enzymatic and product consumption of critical building blocks [196], ensuring that stresses imposed by bio-production processes remain below minimally acceptable levels [197], accounting for both the unpredictability in the bio-reactor environments (random mutations, differing cell variations) and synthetic component behavior [198].

Deep learning, especially advanced DL, can investigate both 'noisy' biological data and merge incomplete input/output variables in data sets [106,199]. Despite initial problems (a plethora of non-standardized data with different measured variables), techniques such as transfer learning [55], and, more importantly, developing database standards in metabolic engineering and systems biology fields [200] will greatly aid in advancing this novel and exciting field.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2894713, IEEE Journal of Biomedical and Health Informatics

IV. Current needs

As traditional and emerging fields continue to evolve, it is of interest to identify human needs that necessitate a paradigm shift towards intelligent approaches.

A. Drug Design

Patient-centric drug design aims to use patient meta-data to develop personalized health solutions ranging from dietary and exercise guidelines to timely drug delivery. Drug design must be transformed from a top-down approach (mass-production) to a bottom-up approach (patient-centric using meta-data).

Deep learning architectures can use predictive analysis and patient database meta-searches to ascertain optimal drug design, thereby preventing dangerous side-effects during the drug regimen. Once designed, the drug must be intelligently delivered. A futuristic approach is to merge molecular robotics with DL architectures to develop optimal routes for drug delivery to targeted areas. One strategy could be to send molecular robots for 'training runs' to their specific targets, so that once they are inside the patient with the drugs, they "know" exactly where to deliver them.

B. Multi-platform Data

Unlike homogenous data that are restricted to one system and can be opened by only one infrastructure and architecture, Multiplatform data (MDA) can be read, and used, across multiple platforms. This goes beyond merely opening separated data warehouse silos: an integrated architecture can satisfy many requirements ranging from meta-data analysis to sophisticated modeling. The merging of previously incompatible data sets is an opportunity that ML algorithms can greatly benefit from.

The need to integrate raw data from different formats, without reducing data quality, requires platforms that can be used beyond merely stacking the raw data, but also integrating the data into complex models to capture the depth of phenomena. These models can be used for making intelligent diagnoses, designing intelligent drugs (and delivery platforms), and developing novel therapeutic strategies for complex scenarios, for example, patients with multiple aliments who require multiple drugs.

C. Intelligent Pre-Diagnosis

As life expectancies continue to rise, concerns are mounting that multiple ailments due to advanced old age require multiple drugs that are necessary for well-being but can have harmful side effects due to drug interactions. Predictive models of drug interactions would enable better management of drug combinations and development of safer alternatives. Machine learning algorithms can play a major role in addressing this urgent clinical need.

V. Deep Learning for Molecular Robotics: An example

Experiments with kinesin-powered molecular shuttles described in section II.b and Figure 3 deliver sequential digital images in a fluorescence microscope's field of view [201]. These images can be used to train a neural network that, upon successfully "learning" the shuttle dynamics, may acquire the ability to generate a new sequence of images from a starting image with dynamics that are nearly indistinguishable to the actual shuttle dynamics.

This task can be approached using a RNN, since it can process an input sequence one element at a time while maintaining a 'state vector' within their hidden units that contains information about the sequence element's entire history [202]-[204]. Training RNNs requires a backpropagation algorithm, which is problematic since the gradients either grow or shrink at each time step [205],[206]. Further, the layers share the same weights, making it difficult to learn to store information for long periods of time [207],[208].

These problems can be addressed with a long short-term memory (LSTM)-based convolutional RNNs (CRNNs). An LSTM algorithm uses memory cells to remember preceding inputs. It is an accumulator that connects to itself at the next time step, copying its own real-valued state and accumulating the external signal. This self-connection is gated by another unit that learns when to clear its memory content [50].

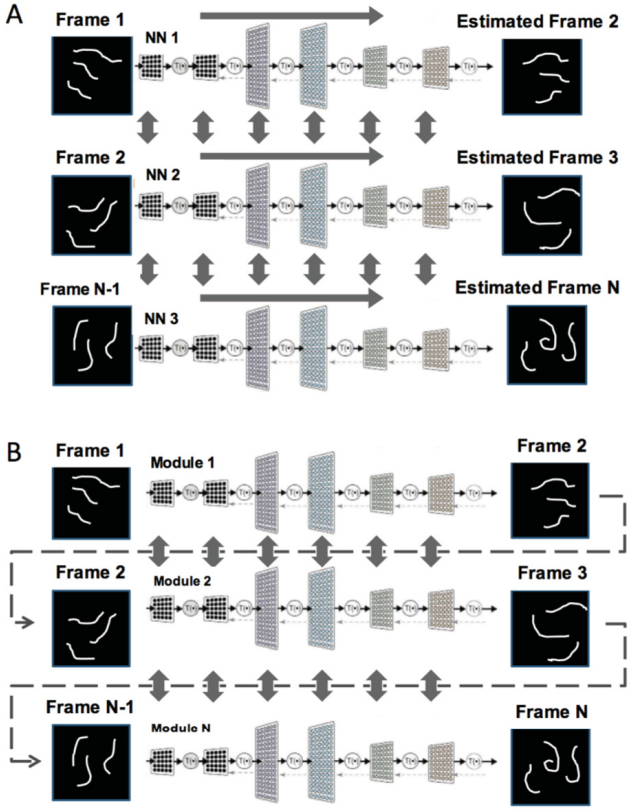

Fig. 4: The CRNN Architecture. The Training phase (A) results in the creation of sequential estimated frames. In the Testing phase (B), the first frame of an unused video is the starting point for the sequential estimated frames. 'NN' = Neural Network.

During the training session, a sequential image batch is fed into the first CRNN, whose output are the next estimated images. In the first iteration (shown in Figure 4A), each original frame (Frame 1, Frame 2…Frame N-1) is fed into sequential CRNNs concurrently and trained in parallel, resulting in sequential estimated frames. The CRNNs were trained to

extract specific features (in this case, fluorescently labeled microtubules) to estimate their trajectories using the LSTM algorithm. During the training phase, the number of iterations depends on how well the network learns to adjust its weights in the internal layer to predict the correct labeling of the inputs (the sequential image batch).

Upon the completion of the training phase, the first frame of a second image sequence can be used as the starting point from which all following frames are estimated by the trained neural network. The estimated image sequence can then be compared to the actual image sequence and validated to determine the neural network's performance with respect to predicting the behavior of a swarm of molecular shuttles.

VI. Trends

A. Multi-modal Deep Learning Architectures

The merging of data sets from multiple sources can lead to discovering previously unknown bio-markers, matching genotype and phenotype relations, and developing strong associations among various, inter-related biological phenomena. Other data sources can include images, biomedical signals, entire research databases, and electronic health records. A multi-modal DL architecture can classify information from multiple sources and by using these data streams compensate for missing information [210]. Promising research on multi-modal deep learning [145],[162],[211] can open the door to advances in traditional fields and emerging fields such as complex diseases, drug-protein and drug-drug interactions in variable environments.

Furthermore, multi-modal architectures can develop an overview of an entire system to elucidate complex interactions, develop single (or multiple) scenarios using current knowledge and permit optimal scenario development that can be tested in the laboratory.

Molecular robotics – our field of research – can also be expected to benefit from multi-modal DL architectures. In one scenario, tracking a single bio-sensor or drug delivery transport in the human body can be challenging without intrusive equipment and expensive procedures. Multi-modal DL can use 'incomplete' information of the molecular robot's surroundings to provide a better idea of its location.

B. Multi-scale Modeling

Multi-scale modeling merges models at different scales to describe a system [212]. The realization that intricate and inter-related complex phenomena play important roles at multiple levels requires a new modeling platform that balances macroscale accuracy and microscale model efficiency.

The two multi-scale modeling paradigms are sequential (which uses microscale models to form a macroscale model) and concurrent (which uses microscale models as needed as a macroscale model continues to be developed).

Interdisciplinary research and development in molecular robotics for intelligent diagnosis and personalized drug delivery require considerations in milli/micro-robot design, their behavior in *in vitro* and *in vivo* environments under variable conditions, and the internal phenomena within the human body under variable conditions, to name a few.

In drug discovery, multi-scale modeling can be merged with VS to find promising molecules and binding sites, and circumvent the need for expensive clinical trials, particularly if the emphasis is towards medical conditions that affect small populations. Multi-scale anatomy and physiology models in diseased conditions, when merged with ML and VS algorithms, can be used to predict single, or multiple, drug molecule effects.

C. Molecular Robot Localization

As emerging fields develop, research is needed to ascertain the technology's potential in intelligent drug delivery and release in affected areas, and real-time localization. This is critical for individuals who suffer from multiple diseases and require multiple drug treatments, particularly if molecular robot 'teams' with different drug cargoes must be injected *in vivo* and have to be tracked to ensure proper drug delivery and release in targeted regions.

Sitti et al have merged a CNN architecture with sophisticated sensors to provide real-time robot localization [11]. More research should focus on enhancing localization capabilities of multiple molecular shuttles carrying cargo (drugs, nano-manufacturing components for 'on-site' assembly, etc.). Traditional and emerging DL architectures can, when merged with multi-modal and/or multi-scale modeling, provide opportunities for training molecular shuttles to autonomously determine their position *in vivo*. The number of molecular shuttles can increase, potentially providing the basis for an intelligent swarm to go to multiple designated sites and release their drug cargo.

D. Clinical Trials

The introduction of traditional and emerging ML algorithms can potentially reduce the time and expense across the entire drug discovery and delivery pipeline, leading to timely discoveries that can be rapidly tested in clinical trials and delivered to the consumer marketplace.

To begin with, VS, multi-modal, and multi-scale technologies can be used for initial molecule discovery that can be merged with patient data from various sources to optimally match molecules for specific treatment. The merged meta-data can also account for possible side-effects of molecule combinations in patients.

Other ML aspects that can aid in clinical trials can be to use collected meta-data to predict drug molecule outcomes. If enough patient meta-data is acquired, one possible use can be to mimic short-term and long-term effects of simulated drug intake in virtual patients. An optimal drug combination, once discovered, can be developed and tested on real patients to validate the virtual patient models in their outcomes (the virtual patient feeling better, or worse).

These patient models can contribute to a knowledge of the effects of molecules in different patient profiles, potentially obviating the need to repeat the same trials in real-time, saving time and resources.

As lifespans continue to grow, patients will be suffering from multiple ailments that require multiple drugs, resulting in deleterious side effects. As personalized medicine continues to evolve, another potential research avenue will be to use patient meta-data, including ML architectures to account for optimal,

and possibly multiple, drug molecule interactions and delivery platforms to develop a personalized, patient-centric solution. This could open the way towards research into rare diseases that were, until fairly recently, not considered a priority by major pharmaceutical companies. Implementing ML could provide a paradigm shift towards treating rare diseases.

VII. CHALLENGES AND OPORTUNITIES

Implementing DL techniques and platforms in traditional and emerging fields will require the ML research and development community to circumvent challenges.

A. Black Box to White Box

Although DL platforms can detect increasingly abstract areas of interest in various data platforms, how the data are analyzed and obtained remains unknown, engendering criticisms of the 'black box' nature of DL platforms, where we do not know how the results are obtained. As data becomes increasingly complex, the shift from a 'black box' (where we know little about how the results are obtained) to a 'white box' (where we know how results are obtained) will be required to meaningfully interpret results. One solution is a de-convolutional neural network (that visualizes specific input representations in CNNs) [213].

The key problem is that when hundreds, if not thousands, of neural units with multiple input layers are included, it is difficult to understand how the algorithms interpret the data, and how they come to the conclusions that they do, unlike past studies relying on simple neural networks with few neural units and domain expertise. One solution is to develop protocols within the algorithm that would 'explain' how it came to the conclusions that it did. However, as DL algorithms continue to evolve to meet the demands of studying and understanding increasingly complex data, simple messages and rationales will be insufficient.

This urgency is not only limited to research labs and industries, but also in the everyday lives of citizens who use these algorithms. The main problem is that we still have a limited understanding of how the brain actually works: we are replicating our limited understanding of biological brains with a 'rough copy' of ANNs.

One solution is use neuroscience to better understand the biological brain's capability and learning process, and using these insights to develop more sophisticated DL algorithms. This implies studying, and understanding, how the brain learns and changes from learning.

B. Data Complexity

As traditional fields and emerging fields continue to evolve, data are becoming increasingly complex, as multiple phenomena are being recorded. Such complex data is making it increasingly difficult to measure and interpret using traditional ML approaches. Even with DL algorithms, proving correlation and causation from numerous phenomena is becoming increasingly difficult. The increasing demands placed on correctly interpreting increasingly intricate phenomena require numerous quality control mechanisms and numerous interpretation mechanisms in the workflows. Emerging DL algorithms can provide some means of overcoming data complexity.

Developing casual relationships among numerous, inter-related phenomena will require a mixed strategy of stacked, multi-modal architectures that study features in one system to ascertain whether phenomena affect each other (which could have been missed if such data had been studied alone) and DL architectures that can begin with identifying and classifying to what extent phenomena affect each other, and begin developing casual relationships that can be replicated in real-time.

Deep learning algorithms may also require multiple forms of human domain expertise, or have specified capsule networks to include different aspects of a phenomenon.

C. Small Data Sets and Overfitting

Collected data sets often remain frustratingly too small to be used for DL algorithms. Furthermore, it is generally assumed that data are of both sufficient quantity and generally balanced. Imbalanced data sets extend into other areas, such as clinical trials where less data exists with diseased groups vs. control groups. Privacy laws further limit data availability from diseased groups. This can severely skew the results of DL algorithms that require large amounts of balanced data to provide accurate measurements.

Data sets of insufficient size also result in overfitting: the training data error is low, but the testing data error is high. This results in the model failing to learn a proper generalization of the knowledge in the data. Although methods exist to rectify this problem (for example, the dropout method, where random nodes, and their edges, are 'dropped' from the network during training [214]), overfitting remains a huge problem with small data sets. One solution is data preprocessing (sampling and basic feature extraction) [215]-[217]: sampling balances an imbalanced data distribution, thereby lessening imbalanced data's potential impact. Other methods include cost-sensitive learning, that calculates the costs of misclassifying data [218], and algorithmic modification, where algorithms can be modified to suit small datasets [218].

D. Displacing the Workforce through Automation

The ability of machine learning algorithms to predict, and model, the outcome of experiments has well been documented, from simple chemical reactions to complex organic systems [2-11]. This enables the construction of experiments *in silico* and, potentially obviates the need for performing experiments, especially if supplies are limited. As databases continue to grow in quantity and modeling becomes more sophisticated, the potential for constructing, and performing, experiments in a 'virtual lab' will continue to grow.

However, concerns are mounting that as automation progresses, traditional fields are in danger of being displaced [12]. As ML algorithms continue to be developed and improved, concerns may mount that the entire drug discovery and testing workflow may be automated [2]. From the drug manufacturer's viewpoint, such investments may pay off in the long-term, as pressure is mounting to reduce the time and costs towards developing new drugs, combined with calls for more 'patient-centric' medicine. This includes drugs that account for the person's genomic profile, thereby negating dangerous drug side effects. For the workers in the drug discovery and clinical workflow, there is little comfort towards losing one's job to an

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2894713, IEEE Journal of Biomedical and Health Informatics

ML platform. These concerns should be addressed as pressure continues to mount to rapidly convert complex data into products and services.

E. Data Preparation

Preparing the data to ensure that they can be correctly interpreted by the ML algorithm is a tedious and frustrating, yet necessary, step: the consequences of not thoroughly executing each step would be data misinterpretation. This can result in added cost, and, even worse, in loss of human life.

Every aspect of data preparation must be accounted for: to begin with, the data set must be of considerable size, as too little data can result in misinterpretations. Data sets must also be properly partitioned into training and testing sets. The models must be properly selected, trained to know what parameters to look for, and tested on different data sets to prevent overfitting, and being able to test the data properly. The raw data must then be normalized to adjust variables from different data sets into a common scale, which would ensure that proper results comes out from the algorithm's output layer.

F. Selecting Correct Architectures and Hyper-parameters

Using the correct ML algorithm is critical and requires a thorough understanding of each algorithm's capabilities, advantages, and disadvantages to obtain optimal results and prevent data misinterpretation. To begin with, the algorithms can be divided into three classes: DNNs for internal correlations in high-dimensional data, CNNs for spatial information analysis, and RNNs for sequential data analysis [29]. Even when the correct algorithm is selected, the 'correct' hyper-parameters are important, as it can affect the results [2]. Automated hyper-parameter selection is slowly superseding the use of human ML experts for hyperparameter selection [3].

G. Training the Model

Training a model involves feeding data into a network to make it 'learn' to look for specific parameters within a data set. The data is sent into the first of many neural layers, whereby weights are assigned to specific features. An image may begin with the first layer looking at edges, followed by the second layer looking at specific features, followed by layers that continue to pass the image until the final layer releases the final output [2]. Initial parameters should be randomized to prevent fixed initialization and independently sampled from normal distributions with minimal variances [3],[4]. The batch size and training rates can affect both the training speed and model performance [5].

Overfitting continues to be a problem. The dropout method is one way of addressing this problem: another method, called 'early stopping,' stops the moment that validation performance begins to either saturate or deteriorate, leaving the parameters with the best performance. Another method is layer-wise pre-training, which pre-trains unsupervised layers using either autoencoders or restricted Boltzmann machines, with the entire network then being fine-tuned.

H. Opportunities

Although ML algorithms are powerful and have been used for multiple applications, challenges remain, despite efforts to improve their capabilities. This can open opportunities for researchers to advance the ML field to cope with increasingly complex data. Two opportunities are discussed from our attempt at organizing the swarm behavior of molecular robots, and the third opportunity on how to advance the field using research from neuroscience.

The first opportunity is to advance automated image analysis to identify molecular robots and their dynamic state (position, direction of movement, etc.). Current algorithms struggle to quickly and reliably identify kinesin-propelled microtubules serving as molecular shuttles when they both operate at high densities and dynamically change their shape. Artificial neural networks can make a significant contribution, although the best parameter optimization approach has to be identified [104], [219]-[221].

The second opportunity is to apply DL techniques to model and elucidate swarm behavior. Multimodal and multiscale information can be addressed with state-of-the-art ML techniques [222],[223]. However, the presence of thermal noise at the molecular and nanoscale can negatively affect the ML algorithm's ability to predict behavior.

The biggest opportunity can come from tackling the major criticism that nobody really knows how an ML algorithm draws conclusions ('black box'). As more input layers are added within an algorithm, confusion increases. Forging research partnerships with neuroscientists can aid in understanding and appreciating brain functions. Their insight could be the key towards developing the next generation of ML algorithms that can 'explain' their decision-making processes.

VIII. CONCLUSION

Machine learning is presenting limitless opportunities for traditional and emerging fields. As more data are collected and analyzed to understand complex phenomena and their roles in the development, maintenance, and regulation of systems from the nano-level to the macro-level, the field must continue to expand into new areas. Although tremendous progress has been made, more work remains, particularly when working with systems that lack the massive quantity of data required for DL algorithms. As data become increasingly complex, more measures must be taken to ensure that the right DL algorithms are used, and to understand how the algorithms obtained their results.

The molecular robotics field and, in a broader sense, microscopic agents designed and programmed by synthetic biologists [224], promise advances in drug delivery, biosensors, regenerative medicine and experimental therapies. The ability of data scientists to analyze, interpret, and model, unstructured data to convert it into a deeper understanding of complex phenomena, can assist in understanding the underlying dynamics in these applications. The use of DL algorithms towards understanding the swarming dynamics in molecular shuttles is an example of the potential applications that DL algorithms have.

We are looking forward to the large impact ML will have in in the decades to come.

REFERENCES

[1] F. Ghasemi, A. Mehridehnavi, A. Fassihi, and H. J. A. S. C. Perez-Sanchez, "Deep Neural Network in Biological Activity Prediction using Deep Belief Network," 2017.

[2] S. Min, B. Lee, and S. J. B. i. b. Yoon, "Deep learning in bioinformatics," *Briefings in Bioinformatics,* vol. 18, no. 5, pp. 851-869, 2017.

[3] T. Ching *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *J R Soc Interface,* vol. 15, no. 141, 2018.

[4] K. Lan, D. T. Wang, S. Fong, L. S. Liu, K. K. L. Wong, and N. Dey, "A Survey of Data Mining and Deep Learning in Bioinformatics," *J Med Syst,* vol. 42, no. 8, p. 139, 2018.

[5] C. Angermueller, T. Parnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Mol Syst Biol,* vol. 12, no. 7, p. 878, 2016.

[6] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov, "Applications of Deep Learning in Biomedicine," *Mol Pharm,* vol. 13, no. 5, pp. 1445-54, 2016.

[7] M. Ziatdinov *et al.*, "Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations," *ACS Nano,* vol. 11, no. 12, pp. 12742-12752, 2017.

[8] T. Oyetunde, F. S. Bao, J. W. Chen, H. G. Martin, and Y. J. Tang, "Leveraging knowledge engineering and machine learning for microbial bio-manufacturing," *Biotechnol Adv,* vol. 36, no. 4, pp. 1308-1315, 2018.

[9] B. Maryasin, P. Marquetand, and N. Maulide, "Machine Learning for Organic Synthesis: Are Robots Replacing Chemists?," *Angew Chem Int Ed Engl,* vol. 57, no. 24, pp. 6978-6980, 2018.

[10] A. P. Bartok *et al.*, "Machine learning unifies the modeling of materials and molecules," *Sci Adv,* vol. 3, no. 12, p. e1701816, Dec 2017.

[11] M. Turan, J. Shabbir, H. Araujo, E. Konukoglu, and M. Sitti, "A deep learning based fusion of RGB camera information and magnetic localization information for endoscopic capsule robots," *Int J Intell Robot Appl,* vol. 1, no. 4, pp. 442-450, 2017.

[12] A. Lavecchia, "Machine-learning approaches in drug discovery: methods and applications," *Drug Discov Today,* vol. 20, no. 3, pp. 318-31, 2015.

[13] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins, "Next-Generation Machine Learning for Biological Networks," *Cell,* vol. 173, no. 7, pp. 1581-1592, 2018.

[14] V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA,* vol. 316, no. 22, pp. 2402-2410, 2016.

[15] M. Hagiya, A. Konagaya, S. Kobayashi, H. Saito, and S. Murata, "Molecular robots with sensors and intelligence," *Acc Chem Res,* vol. 47, no. 6, pp. 1681-90, 2014.

[16] A. Agarwal and H. Hess, "Biomolecular motors at the intersection of nanotechnology and polymer science," *Progress in Polymer Science,* vol. 35, no. 1-2, pp. 252-277, 2010.

[17] K. Lund *et al.*, "Molecular robots guided by prescriptive landscapes," *Nature,* vol. 465, no. 7295, pp. 206-10, 2010.

[18] A. J. Wollman, C. Sanchez-Cano, H. M. Carstairs, R. A. Cross, and A. J. Turberfield, "Transport and self-organization across different length scales powered by motor proteins and programmed by DNA," *Nat Nanotechnol,* vol. 9, no. 1, pp. 44-7, 2014.

[19] W. Gao and J. Wang, "The environmental impact of micro/nanomachines: a review," *ACS Nano,* vol. 8, no. 4, pp. 3170-80, 2014.

[20] S. Yim, E. Gultepe, D. H. Gracias, and M. Sitti, "Biopsy using a magnetic capsule endoscope carrying, releasing, and retrieving untethered microgrippers," *IEEE Transactions on Biomedical Engineering,* vol. 61, no. 2, pp. 513-521, 2014.

[21] S. Tasoglu, E. Diller, S. Guven, M. Sitti, and U. Demirci, "Untethered micro-robotic coding of three-dimensional material composition," *Nat Commun,* vol. 5, p. 3124, 2014.

[22] R. W. Carlsen, M. R. Edwards, J. Zhuang, C. Pacoret, and M. Sitti, "Magnetic steering control of multi-cellular bio-hybrid microswimmers," *Lab Chip,* vol. 14, no. 19, pp. 3850-9, 2014.

[23] D. R. Frutiger, K. Vollmers, B. E. Kratochvil, and B. J. Nelson, "Small, fast, and under control: wireless resonant magnetic micro-agents," *The International Journal of Robotics Research,* vol. 29, no. 5, pp. 613-636, 2010.

[24] S. A. Bencherif *et al.*, "Injectable preformed scaffolds with shape-memory properties," *Proc Natl Acad Sci U S A,* vol. 109, no. 48, pp. 19590-5, 2012.

[25] S. J. J. o. M.-B. R. Martel, "Microrobotics in the vascular network: present status and next challenges," *Journal of Micro-Bio Robotics*, vol. 8, no. 1, pp. 41-52, 2013.

[26] Copland, M. (2018, September 06). The Difference Between AI, Machine Learning, and Deep Learning? | NVIDIA Blog. Retrieved from https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

[27] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. J. A. m. Shannon, "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955," vol. 27, no. 4, p. 12, 2006.

[28] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 609-616: ACM.

[29] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning, vol. 1," ed: MIT press Cambridge, 2016.

[30] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput,* vol. 18, no. 7, pp. 1527-54, 2006.

[31] A. L. J. I. J. o. r. Samuel and development, "Some studies in machine learning using the game of checkers," vol. 3, no. 3, pp. 210-229, 1959.

[32] M. Tompa *et al.*, "Assessing computational tools for the discovery of transcription factor binding sites," *Nat Biotechnol,* vol. 23, no. 1, pp. 137-44, 2005.

[33] F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," ed: Elsevier, 2013.

[34] Sarhan A.H., "Cancer classification based on microarray gene expression data using DCT and ANN," *Journal of Theoretical and Applied Information Technology,* vol. 6, no. 2, 2009.

[35] M. Puri, Y. Pathak, V. K. Sutariya, S. Tipparaju, and W. Moreno, *Artificial Neural Network for Drug Design, Delivery and Disposition*. Academic Press, 2015.

[36] A. K. Jain, M. N. Murty, and P. J. J. A. c. s. Flynn, "Data clustering: a review," vol. 31, no. 3, pp. 264-323, 1999.

[37] M. M. Gordon, A. M. Moser, and E. Rubin, "Unsupervised analysis of classical biomedical markers: robustness and medical relevance of patient clustering using bioinformatics tools," *PLOS One,* vol. 7, no. 3, p. e29578, 2012.

[38] S. Hussein, M. M. Chuquicusma, P. Kandel, C. W. Bolan, M. B. Wallace, and U. J. a. p. a. Bagci, "Supervised and Unsupervised Tumor Characterization in the Deep Learning Era," 2018.

[39] J. Lyons *et al.*, "Predicting backbone Calpha angles and dihedrals from protein sequences by stacked sparse auto-encoder deep neural network," *J Comput Chem,* vol. 35, no. 28, pp. 2040-6, 2014.

[40] R. Heffernan *et al.*, "Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning," *Sci Rep,* vol. 5, p. 11476, 2015.

[41] M. Spencer, J. Eickholt, and C. Jianlin, "A Deep Learning Network Approach to ab initio Protein Secondary Structure Prediction," *IEEE/ACM Trans Comput Biol Bioinform,* vol. 12, no. 1, pp. 103-12, 2015.

[42] S. P. Nguyen, Y. Shang, and D. Xu, "DL-PRO: A Novel Deep Learning Method for Protein Model Quality Assessment," *Proc Int Jt Conf Neural Netw,* vol. 2014, pp. 2071-2078, 2014.

[43] S. M. Plis *et al.*, "Deep learning for neuroimaging: a validation study," *Front Neurosci,* vol. 8, p. 229, 2014.

[44] K. L. Hua, C. H. Hsu, S. C. Hidayati, W. H. Cheng, and Y. J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *Onco Targets Ther,* vol. 8, pp. 2015-22, 2015.

[45] H. I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," *Med Image Comput Comput Assist Interv,* vol. 16, no. Pt 2, pp. 583-90, 2013.

[46] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A Cross-Modality Learning Approach for Vessel Segmentation in Retinal Images," *IEEE Trans Med Imaging,* vol. 35, no. 1, pp. 109-18, 2016.

[47] J. Xu *et al.*, "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images," *IEEE Trans Med Imaging,* vol. 35, no. 1, pp. 119-30, 2016.

[48] C. L. Chen *et al.*, "Deep Learning in Label-free Cell Classification," *Sci Rep,* vol. 6, p. 21471, 2016.

[49] M. A. van Gerven, F. P. de Lange, and T. Heskes, "Neural decoding with hierarchical generative models," *Neural Comput,* vol. 22, no. 12, pp. 3127-42, 2010.

[50] S. Koyamada, Y. Shikauchi, K. Nakae, M. Koyama, and S. Ishii, "Deep learning of fMRI big data: a novel approach to subject-transfer decoding," *arXiv: 00093,* 2015.

[51] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A deep learning method for classification of EEG data based on motor imagery," in *International Conference on Intelligent Computing*, 2014, pp. 203-210: Springer.

[52] X. Jia, K. Li, X. Li, and A. Zhang, "A novel semi-supervised deep learning framework for affective state recognition on eeg signals," in *Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on*, 2014, pp. 30-37: IEEE.

[53] M. Huanhuan and Z. Yue, "Classification of electrocardiogram signals with deep belief networks," in *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on*, 2014, pp. 7-12: IEEE.

[54] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," *J Neural Eng,* vol. 8, no. 3, p. 036015, 2011.

[55] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-44, 2015.

[56] R. Kiros *et al.*, "Skip-thought vectors," in *Advances in neural information processing systems*, 2015, pp. 3294-3302.

[57] A. Shrikumar, P. Greenside, and A. Kundaje, "Reverse-complement parameter sharing improves deep learning models for genomics," *arXiv: 103663,* 2017.

[58] O. Denas and J. Taylor, "Deep modeling of gene expression regulation in an erythropoiesis model," in *Representation Learning, ICML Workshop*, 2013.

[59] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning," *Nat Biotechnol,* vol. 33, no. 8, pp. 831-8, 2015.

[60] D. R. Kelley, J. Snoek, and J. L. Rinn, "Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks," *Genome Res,* vol. 26, no. 7, pp. 990-9, 2016.

[61] H. R. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging,* vol. 35, no. 5, pp. 1170-1181, 2016.

[62] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," *Med Image Comput Comput Assist Interv,* vol. 16, no. Pt 2, pp. 411-8, 2013.

[63] P. P. Ypsilantis *et al.*, "Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks," *PLoS One,* vol. 10, no. 9, p. e0137036, 2015.

[64] S. Lee, M. Choi, H.-s. Choi, M. S. Park, and S. Yoon, "FingerNet: Deep learning-based robust finger joint detection from radiographs," in *Biomedical Circuits and Systems Conference (BioCAS)*, 2015, pp. 1-4: IEEE.

[65] J. Duryea, Y. Jiang, P. Countryman, and H. Genant, "Automated algorithm for the identification of joint space and phalanx margin locations on digitized hand radiographs," *Medical Physics,* vol. 26, no. 3, pp. 453-461, 1999.

[66] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clin Neurophysiol,* vol. 120, no. 11, pp. 1927-1940, 2009.

[67] P. Baldi, S. Brunak, P. Frasconi, G. Soda, and G. Pollastri, "Exploiting the past and the future in protein secondary structure prediction," *Bioinformatics,* vol. 15, no. 11, pp. 937-46, 1999.

[68] S. Park, S. Min, H. Choi, and S. Yoon, "deepMiRGene: Deep neural network based precursor microrna prediction," *arXiv: 00017,* 2016.

[69] S. Hochreiter, M. Heusel, and K. Obermayer, "Fast model-based protein homology detection without alignment," *Bioinformatics,* vol. 23, no. 14, pp. 1728-1736, 2007.

[70] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch II, "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG," *Neurocomputing,* vol. 30, no. 1-4, pp. 201-218, 2000.

[71] Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." *Advances in Neural Information Processing Systems*. 2017.

[72] S. Sabour, "Diagnostic Value of Small Bowel Capsule Endoscopy in Isolated Ileitis: A Methodological Mistake," *Dig Dis Sci,* vol. 62, no. 4, pp. 1099-1100, 2017.

[73] M. Ammad-Ud-Din, S. A. Khan, K. Wennerberg, and T. Aittokallio, "Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression," *Bioinformatics,* vol. 33, no. 14, pp. i359-i368, 2017.

[74] B. E. Johnson *et al.*, "Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma," *Science,* vol. 343, no. 6167, pp. 189-193, 2014.

[75] M. Tan, "Prediction of anti-cancer drug response by kernelized multi-task learning," *Artif Intell Med,* vol. 73, pp. 70-77, 2016.

[76] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge data engineering,* vol. 22, no. 10, pp. 1345-1359, 2010.

[77] P. D. Lena, K. Nagata, and P. F. Baldi, "Deep spatio-temporal architectures and learning for protein structure prediction," in *Advances in neural information processing systems*, 2012, pp. 512-520.

[78] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Advances in Neural Information Processing Systems*, 2009, pp. 545-552.

[79] R. Hadsell *et al.*, "Learning long-range vision for autonomous off-road driving," *Journal of Field Robotics,* vol. 26, no. 2, pp. 120-144, 2009.

[80] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International Conference on Artificial Neural Networks*, 2011, pp. 52-59: Springer.

[81] S. Hochreiter, M. Heusel, and K. Obermayer, "Fast model-based protein homology detection without alignment," *Bioinformatics,* vol. 23, no. 14, pp. 1728-1736, 2007.

[82] O. Witkowski and T. Ikegami, "Emergence of Swarming Behavior: Foraging Agents Evolve Collective Motion Based on Signaling," *PLoS One,* vol. 11, no. 4, p. e0152756, 2016.

[83] E. Bonabeau, D. Marco, M. Dorigo, and G. Théraulaz, *Swarm Intelligence: From Natural to Artificial Systems* (no. 1). Oxford university press, 1999.

[84] I. D. Couzin, "Collective cognition in animal groups," *Trends Cogn Sci,* vol. 13, no. 1, pp. 36-43, 2009.

[85] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithms and theory," *IEEE Transactions on Automatic Control,* vol. 51, no. 3, pp. 401-420, Mar 2006.

[86] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *Ieee Transactions on Automatic Control,* vol. 49, no. 9, pp. 1465-1476, Sep 2004.

[87] S. Murata, A. Konagaya, S. Kobayashi, H. Saito, and M. Hagiya, "Molecular robotics: a new paradigm for artifacts," *New Generation Computing,* vol. 31, no. 1, pp. 27-45, 2013.

[88] S. Murata and H. Kurokawa, "Self-organizing robots," *Self-Organizing Robots,* vol. 77, pp. 1-252, 2012.

[89] K. Lund *et al.*, "Molecular robots guided by prescriptive landscapes," *Nature,* vol. 465, no. 7295, pp. 206-10, 2010.

[90] H. Hess and J. L. Ross, "Non-equilibrium assembly of microtubules: from molecules to autonomous chemical robots," *Chem Soc Rev,* vol. 46, no. 18, pp. 5570-5587, 2017.

[91] J. Howard, "Mechanics of Motor Proteins and the Cytoskeleton," 2001.

[92] H. Hess and V. Vogel, "Molecular shuttles based on motor proteins: active transport in synthetic environments," *J Biotechnol,* vol. 82, no. 1, pp. 67-85, 2001.

[93] A. Agarwal and H. Hess, "Biomolecular motors at the intersection of nanotechnology and polymer science," *Progress in Polymer Science,* vol. 35, no. 1-2, pp. 252-277, 2010.

[94] H. Hess, J. Clemmens, D. Qin, J. Howard, and V. Vogel, "Light-controlled molecular shuttles made from motor proteins carrying cargo on engineered surfaces," *Nano Letters,* vol. 1, no. 5, pp. 235-239, 2001.

[95] R. Tucker, A. K. Saha, P. Katira, M. Bachand, G. D. Bachand, and H. Hess, "Temperature compensation for hybrid devices: kinesin's k-m is temperature independent," *Small,* vol. 5, no. 11, pp. 1279-1282, 2009.

[96] H. Hess, "Self-assembly driven by molecular motors," *Soft Matter,* vol. 2, no. 8, pp. 669-677, 2006.

[97] A. T. Lam, V. VanDelinder, A. M. Kabir, H. Hess, G. D. Bachand, and A. Kakugo, "Cytoskeletal motor-driven active self-assembly in in vitro systems," *Soft Matter,* vol. 12, no. 4, pp. 988-97, 2016.

[98] J. J. Keya *et al.*, "DNA-assisted swarm control in a biomolecular motor system," *Nat Commun,* vol. 9, no. 1, p. 453, 2018.

[99] F. Ruhnow, D. Zwicker, and S. Diez, "Tracking single particles and elongated filaments with nanometer precision," *Biophysical Journal,* vol. 100, no. 11, pp. 2820-2828, 2011.

[100] H. Hess *et al.*, "Molecular self-assembly of "nanowires" and "nanospools" using active transport," *Nano Letters,* vol. 5, no. 4, pp. 629-633, 2005.

[101] V. Schaller, C. Weber, C. Semmrich, E. Frey, and A. R. Bausch, "Polar patterns of driven filaments," *Nature,* vol. 467, no. 7311, pp. 73-7, 2010.

[102] Y. Sumino *et al.*, "Large-scale vortex lattice emerging from collectively moving microtubules," *Nature,* vol. 483, no. 7390, pp. 448-452, 2012.

[103] M. Hüttenrauch, A. Šošić, and G. Neumann, "Learning Complex Swarm Behaviors by Exploiting Local Communication Protocols with Deep Reinforcement Learning," *arXiv: 07224,* 2017.

[104] Searls, D. B. (2016). Computational biology. Retrieved from https://www.britannica.com/science/computational-biology

[105] M. K. Leung, H. Y. Xiong, L. J. Lee, and B. Frey, "Deep learning of the tissue-regulated splicing code," *Bioinformatics,* vol. 30, no. 12, pp. i121-i129, 2014.

[106] H. Y. Xiong *et al.*, "RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease," *Science,* vol. 347, no. 6218, p. 1254806, 2015.

[107] J. Zhou and O. G. Troyanskaya, "Predicting effects of noncoding variants with deep learning-based sequence model," *Nat Methods,* vol. 12, no. 10, pp. 931-4, 2015.

[108] G. E. Dahl, N. Jaitly, and R. Salakhutdinov, "Multi-task neural networks for QSAR predictions," *arXiv: 1406,* 2014.

[109] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics,* vol. 32, no. 12, pp. i52-i59, 2016.

[110] A. Ferrari, S. Lombardi, and A. J. P. R. Signoroni, "Bacterial colony counting with convolutional neural networks in digital microbiology imaging," vol. 61, pp. 629-640, 2017.

[111] T. Lee and S. Yoon, "Boosted categorical restricted Boltzmann machine for computational prediction of splice junctions," in *International Conference on Machine Learning*, 2015, pp. 2483-2492.

[112] S. Zhang *et al.*, "A deep learning framework for modeling structural features of RNA-binding protein targets," *Nucleic Acids Research,* vol. 44, no. 4, pp. e32-e32, 2015.

[113] Y. Chen, Y. Li, R. Narayan, A. Subramanian, and X. Xie, "Gene expression inference with deep learning," *Bioinformatics,* vol. 32, no. 12, pp. 1832-1839, 2016.

[114] Y. Li, W. Shi, and W. W. Wasserman, "Genome-wide prediction of cis-regulatory regions using supervised deep learning methods," *BMC Bioinformatics,* vol. 19, no. 1, p. 202, 2018.

[115] Y. Li, W. Shi, and W. W. Wasserman, "Genome-wide prediction of cis-regulatory regions using supervised deep learning methods," *BMC Bioinformatics,* vol. 19, no. 1, p. 202, 2018.

[116] J. Lanchantin, R. Singh, Z. Lin, and Y. Qi, "Deep motif: Visualizing genomic sequence classifications," *arXiv: 01133,* 2016.

[117] H. Zeng, M. D. Edwards, G. Liu, and D. K. Gifford, "Convolutional neural network architectures for predicting DNA-protein binding," *Bioinformatics,* vol. 32, no. 12, pp. i121-i127, 2016.

[118] B. Lee, T. Lee, B. Na, and S. Yoon, "DNA-level splice junction prediction using deep recurrent neural networks," *arXiv: 05135,* 2015.

[119] B. Lee, J. Baek, S. Park, and S. Yoon, "deepTarget: end-to-end learning framework for microRNA target prediction using deep recurrent neural networks," in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2016, pp. 434-442: ACM.

[120] E. Asgari and M. R. Mofrad, "Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics," *PLoS One,* vol. 10, no. 11, p. e0141287, 2015.

[121] S. Hochreiter, M. Heusel, and K. Obermayer, "Fast model-based protein homology detection without alignment," *Bioinformatics,* vol. 23, no. 14, pp. 1728-1736, 2007.

[122] S. K. Sønderby, C. K. Sønderby, H. Nielsen, and O. Winther, "Convolutional LSTM networks for subcellular localization of proteins," in *International Conference on Algorithms for Computational Biology*, 2015, pp. 68-80: Springer.

[123] R. R. Edelman and S. Warach, "Magnetic resonance imaging," *New England Journal of Medicine,* vol. 328, no. 11, pp. 785-791, 1993.

[124] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proc Natl Acad Sci U S A,* vol. 87, no. 24, pp. 9868-72, 1990.

[125] Hsieh J. Computed tomography: principles, design, artifacts, and recent advances. In: *SPIE* Bellingham, WA, 2009.

[126] D. Chapman *et al.*, "Diffraction enhanced x-ray imaging," *Physics in Medicine Biology,* vol. 42, no. 11, p. 2015, 1997.

[127] D. L. Bailey, M. N. Maisey, D. W. Townsend, and P. E. Valk, *Positron emission tomography*. Springer, 2005.

[128] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: a review," *IEEE Rev Biomed Eng,* vol. 2, pp. 147-71, 2009.

[129] H. R. Roth, J. Yao, L. Lu, J. Stieger, J. E. Burns, and R. M. Summers, "Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications," in *Recent advances in computational methods and clinical applications for spine imaging*: Springer, 2015, pp. 3-12.

[130] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, 2014, pp. 844-848: IEEE.

[131] T. Zeng, R. Li, R. Mukkamala, J. Ye, and S. Ji, "Deep convolutional neural networks for annotating gene expression patterns in the mouse brain," *BMC Bioinformatics,* vol. 16, p. 147, 2015.

[132] A. A. Cruz-Roa, J. E. Arevalo Ovalle, A. Madabhushi, and F. A. Gonzalez Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," *Med Image Comput Comput Assist Interv,* vol. 16, no. Pt 2, pp. 403-10, 2013.

[133] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," in *Medical Imaging 2015: Computer-Aided Diagnosis*, 2015, vol. 9414, p. 94140V: International Society for Optics and Photonics.

[134] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, and P. E. Barbano, "Toward automatic phenotyping of developing embryos from videos," *IEEE Trans Image Process,* vol. 14, no. 9, pp. 1360-71, 2005.

[135] S. C. Turaga *et al.*, "Convolutional networks can learn to generate affinity graphs for image segmentation," *Neural Computation,* vol. 22, no. 2, pp. 511-538, 2010.

[136] M. Helmstaedter, K. L. Briggman, S. C. Turaga, V. Jain, H. S. Seung, and W. Denk, "Connectomic reconstruction of the inner plexiform layer in the mouse retina," *Nature,* vol. 500, no. 7461, pp. 168-74, 2013.

[137] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843-2851.

[138] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *International conference on medical image computing and computer-assisted intervention*, 2013, pp. 246-253: Springer.

[139] M. Havaei *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Med Image Anal,* vol. 35, pp. 18-31, 2017.

[140] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation," in *Advances in Neural Information processing systems*, 2015, pp. 2998-3006.

[141] C. L. Chen *et al.*, "Deep Learning in Label-free Cell Classification," *Sci Rep,* vol. 6, p. 21471, 2016.

[142] J. Cho, K. Lee, E. Shin, G. Choy, and S. Do, "Medical image deep learning with hospital PACS dataset," *arXiv:06348,* 2015.

[143] H. R. Roth *et al.*, "Anatomy-specific classification of medical images using deep convolutional nets," *arXiv:.04003,* 2015.

[144] E. Niedermeyer and F. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[145] G. Buzsaki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes," *Nat Rev Neurosci,* vol. 13, no. 6, pp. 407-20, 2012.

[146] H. J. L. Marriott and G. S. Wagner, *Practical electrocardiography*. Williams & Wilkins, 1988.
[147] C. De Luca, "The use of surface electromyography in biomechanics," *Journal of Applied Biomechanics,* vol. 13, no. 2, pp. 135-163, 1997.
[148] L. R. Young and D. Sheena, "Eye-movement measurement techniques," *Am Psychol,* vol. 30, no. 3, pp. 315-30, 1975.
[149] R. Barea, L. Boquete, M. Mazo, and E. López, "System for assisted mobility using eye movements based on electrooculography," *IEEE Transactions on Neural Systems Rehabilitation Engineering,* vol. 10, no. 4, pp. 209-218, 2002.
[150] Z. V. Freudenburg, N. F. Ramsey, M. Wronkiewicz, W. D. Smart, R. Pless, and E. C. Leuthardt, "Real-time naive learning of neural correlates in ECoG electrophysiology," *International Journal of Machine Learning,* vol. 1, no. 3, p. 269, 2011.
[151] K. Li, X. Li, Y. Zhang, and A. Zhang, "Affective state recognition from EEG with deep belief networks," in *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*, 2013, pp. 305-310: IEEE.
[152] W.-L. Zheng, H.-T. Guo, and B.-L. Lu, "Revealing critical channels and frequency bands for emotion recognition from EEG with deep belief network," in *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on*, 2015, pp. 154-157: IEEE.
[153] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal,* vol. 2014, 2014.
[154] S. Stober, D. J. Cameron, and J. A. Grahn, "Classifying EEG Recordings of Rhythm Perception," in *ISMIR*, 2014, pp. 649-654.
[155] S. Stober, D. J. Cameron, and J. A. Grahn, "Using Convolutional Neural Networks to Recognize Rhythm Stimuli from Electroencephalography Recordings," in *Advances in Neural Information Processing Systems*, 2014, pp. 1449-1457.
[156] H. Cecotti and A. Graeser, "Convolutional neural network with embedded Fourier transform for EEG classification," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008, pp. 1-4: IEEE.
[157] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans Pattern Anal Mach Intell,* vol. 33, no. 3, pp. 433-45, 2011.
[158] M. Soleymani, S. Asghari-Esfeden, M. Pantic, and Y. Fu, "Continuous emotion detection using EEG signals and facial expressions," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, 2014, pp. 1-6: IEEE.
[159] Z. Wang, S. Lyu, G. Schalk, and Q. Ji, "Deep Feature Learning Using Target Priors with Applications in ECoG Signal Decoding for BCI," in *IJCAI*, 2013, pp. 1785-1791.
[160] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Deep feature learning for EEG recordings," *arXiv: 04306,* 2015.
[161] J. Turner, A. Page, T. Mohsenin, and T. Oates, "Deep belief networks used on high resolution multichannel electroencephalography data for seizure detection," in *2014 AAAI Spring Symposium Series*, 2014.
[162] Y. Zhao and L. He, "Deep learning in the EEG diagnosis of Alzheimer's disease," in *Asian Conference on Computer Vision*, 2014, pp. 340-353: Springer.
[163] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep stage classification using unsupervised feature learning," *Journal of Advances in Artificial Neural Systems,* vol. 2012, p. 5, 2012.
[164] P. R. Davidson, R. D. Jones, and M. T. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans Biomed Eng,* vol. 54, no. 5, pp. 832-9, 2007.
[165] A. L. Barabasi and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nat Rev Genet,* vol. 5, no. 2, pp. 101-13, 2004.
[166] A. Bordbar and B. Palsson, "Using the reconstructed genome-scale human metabolic network to study physiology and pathology," *Journal of Internal Medicine,* vol. 271, no. 2, pp. 131-141, 2012.
[167] H. Y. Chuang, E. Lee, Y. T. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis," *Mol Syst Biol,* vol. 3, p. 140, 2007.
[168] K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabasi, "The human disease network," *Proc Natl Acad Sci U S A,* vol. 104, no. 21, pp. 8685-90, 2007.
[169] L. R. Yates *et al.*, "Subclonal diversification of primary breast cancer revealed by multiregion sequencing," *Nat Med,* vol. 21, no. 7, pp. 751-9, 2015.
[170] K. Michailidou *et al.*, "Large-scale genotyping identifies 41 new loci associated with breast cancer risk," *Nat Genet,* vol. 45, no. 4, pp. 353-61, 361e1-2, 2013.
[171] P. Y. Lum *et al.*, "Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes," vol. 97, pp. 50-62, 2006.
[172] A. Chatr-Aryamontri *et al.*, "The BioGRID interaction database: 2017 update," *Nucleic Acids Res,* vol. 45, no. D1, pp. D369-D379, 2017.
[173] J. Brennecke *et al.*, "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila," *Cell,* vol. 128, no. 6, pp. 1089-103, 2007.
[174] A. De la Fuente, "From 'differential expression' to 'differential networking'–identification of dysfunctional regulatory networks in diseases," *Journal of Trends in Genetics,* vol. 26, no. 7, pp. 326-333, 2010.
[175] R. Mall, L. Cerulo, H. Bensmail, A. Iavarone, and M. Ceccarelli, "Detection of statistically significant network changes in complex biological networks," *BMC Syst Biol,* vol. 11, no. 1, p. 32, 2017.
[176] B. Chen and A. J. Butte, "Leveraging big data to transform target selection and drug discovery," *Clin Pharmacol Ther,* vol. 99, no. 3, pp. 285-97, 2016.
[177] S. Kim *et al.*, "PubChem Substance and Compound databases," *Nucleic Acids Res,* vol. 44, no. D1, pp. D1202-13, 2016.
[178] D. S. Wishart *et al.*, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic Acids Res,* vol. 34, no. Database issue, pp. D668-72, 2006.
[179] T. Sterling and J. J. Irwin, "ZINC 15--Ligand Discovery for Everyone," *J Chem Inf Model,* vol. 55, no. 11, pp. 2324-37, 2015.
[180] D. E. Cameron, C. J. Bashor, and J. J. Collins, "A brief history of synthetic biology," *Nat Rev Microbiol,* vol. 12, no. 5, pp. 381-90, 2014.
[181] C. C. Guet, M. B. Elowitz, W. Hsing, and S. Leibler, "Combinatorial synthesis of genetic networks," *Science,* vol. 296, no. 5572, pp. 1466-70, 2002.
[182] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature,* vol. 403, no. 6767, pp. 339-42, 2000.
[183] S. Mukherji and A. van Oudenaarden, "Synthetic biology: understanding biological design from synthetic circuits," *Nat Rev Genet,* vol. 10, no. 12, pp. 859-71, 2009.
[184] P. E. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems," *Nat Rev Mol Cell Biol,* vol. 10, no. 6, pp. 410-22, 2009.
[185] D. Fabry, E. Sugiono, and M. Rueping, "Online monitoring and analysis for autonomous continuous flow self-optimizing reactor systems," *Journal of Reaction Chemistry Engineering,* vol. 1, no. 2, pp. 129-133, 2016.
[186] R. D. King *et al.*, "Functional genomic hypothesis generation and experimentation by a robot scientist," *Nature,* vol. 427, no. 6971, pp. 247-52, 2004.
[187] K. Williams *et al.*, "Eve: Integration of machine learning with compound testing in a Robot Scientist," 2015.
[188] G. Q. Chen, "Omics Meets Metabolic Pathway Engineering," *Cell Syst,* vol. 2, no. 6, pp. 362-3, 2016.
[189] J. M. Monk *et al.*, "Multi-omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes," *Cell Syst,* vol. 3, no. 3, pp. 238-251 e12, 2016.
[190] J. Utrilla *et al.*, "Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-scale Mechanism of Adaptive Evolution," *Cell Syst,* vol. 2, no. 4, pp. 260-71, 2016.
[191] F. Ceroni, R. Algar, G. B. Stan, and T. Ellis, "Quantifying cellular capacity identifies gene expression designs with reduced burden," *Nat Methods,* vol. 12, no. 5, pp. 415-8, 2015.
[192] T. M. Hoehler and B. B. Jorgensen, "Microbial life under extreme energy limitation," *Nat Rev Microbiol,* vol. 11, no. 2, pp. 83-94, 2013.
[193] T. Nishizaki, K. Tsuge, M. Itaya, N. Doi, and H. Yanagawa, "Metabolic engineering of carotenoid biosynthesis in Escherichia coli by ordered gene assembly in Bacillus subtilis," *Appl Environ Microbiol,* vol. 73, no. 4, pp. 1355-61, 2007.
[194] D. Chicco, P. Sadowski, and P. Baldi, "Deep autoencoder neural networks for gene ontology annotation predictions," in *Proceedings of*

the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2014, pp. 533-540: ACM.

[195] A. P. Arkin *et al.*, "The DOE systems biology knowledgebase (KBase)," p. 096354, 2016.

[196] Y. Jeune-Smith, A. Agarwal, and H. Hess, "Cargo loading onto kinesin powered molecular shuttles," *J Vis Exp,* no. 45, 2010.

[197] G. L. Masala, B. Golosio, M. Tistarelli, and E. Grosso, "2D recurrent neural networks for robust visual tracking of non-rigid bodies," *Engineering Applications of Neural Networks, EANN 2016,* vol. 629, pp. 18-34, 2016.

[198] T. Ngo and B. S. Manjunath, "Saccade gaze prediction using a recurrent neural network," *2017 24th IEEE International Conference on Image Processing (ICIP),* pp. 3435-3439, 2017.

[199] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. J. S. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," vol. 17, no. 7, p. 1501, 2017.

[200] S. J. D. Hochreiter, Technische Universität München, "Untersuchungen zu dynamischen neuronalen Netzen," vol. 91, no. 1, 1991.

[201] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks,* vol. 5, no. 2, pp. 157-166, 1994.

[202] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput,* vol. 9, no. 8, pp. 1735-80, 1997.

[203] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research,* vol. 15, pp. 1929-1958, 2014.

[204] H. Hess *et al.*, "Molecular shuttles operating undercover: A new photolithographic approach for the fabrication of structured surfaces supporting directed motility," *Nano Letters,* vol. 3, no. 12, pp. 1651-1655, 2003.

[205] Ngiam, Jiquan, et al. "Multimodal deep learning." *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011.

[206] Y. Cao *et al.*, "Medical Image Retrieval: A Multimodal Approach," *Cancer Inform,* vol. 13, no. Suppl 3, pp. 125-36, 2014.

[207] http://scholarpedia.org/article/Multiscale_modeling

[208] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research,* vol. 15, pp. 1929-1958, 2014.

[209] H. Hess *et al.*, "Molecular shuttles operating undercover: A new photolithographic approach for the fabrication of structured surfaces supporting directed motility," *Nano Letters,* vol. 3, no. 12, pp. 1651-1655, 2003.

[210] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor ai: Predicting clinical events via recurrent neural networks," in *Machine Learning for Healthcare Conference*, 2016, pp. 301-318.

[211] S. Mathew and M. D. Madhu, "Multimodal Deep Learning."

[212] N. Neverova, C. Wolf, G. W. Taylor, and F. Nebout, "Multi-scale deep learning for gesture detection and localization," in *Workshop at the European conference on computer vision*, 2014, pp. 474-490: Springer.

[213] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *Computer Vision - Eccv 2014, Pt I,* vol. 8689, pp. 818-833, 2014.

[214] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM,* vol. 60, no. 6, pp. 84-90, 2017.

[215] X. Y. Liu, J. Wu, and Z. H. Zhou, "Exploratory under-sampling for class-imbalance learning," *IEEE Trans Syst Man Cybern B Cybern,* vol. 39, no. 2, pp. 539-50, 2009.

[216] M. Nakamura, Y. Kajiwara, A. Otsuka, and H. Kimura, "LVQ-SMOTE - learning vector quantization based synthetic minority over-sampling technique for biomedical data," *BioData Min,* vol. 6, no. 1, p. 16, 2013.

[217] T. Jo, N. Japkowicz, "Class imbalances versus small disjuncts," *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 40-49, 2004

[218] D. Erhan, Y. Bengio, A. Courville, P. A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *Journal of Machine Learning Research,* vol. 11, pp. 625-660, 2010.

[219] S. S. Azab and H. A. Hefny, "Swarm Intelligence in Semi-Supervised Classification," *arXiv:.00998,* 2017.

[220] X.-S. Yang and X. He, "Swarm intelligence and evolutionary computation: overview and analysis," in *Recent Advances in Swarm Intelligence and Evolutionary Computation*: Springer, 2015, pp. 1-23.

[221] X.-S. Yang and M. Karamanoglu, "Swarm intelligence and bio-inspired computation: an overview," in *Swarm Intelligence and Bio-Inspired Computation*: Elsevier, 2013, pp. 3-23.

[222] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 689-696.

[223] M. Soleymani, S. Asghari-Esfeden, M. Pantic, and Y. Fu, "Continuous emotion detection using EEG signals and facial expressions," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, 2014, pp. 1-6: IEEE.

[224] T. Chien, A. Doshi, and T. Danino, "Advances in bacterial cancer therapies using synthetic biology," *Current Opinion in Systems Biology,* vol. 5, pp. 1-8, 2017.