

REVIEW ARTICLE

Machine Learning in Healthcare

Hafsa Habehh¹ and Suril Gohel^{1,*}¹Department of Health Informatics, Rutgers University School of Health Professions, 65 Bergen Street, Newark, NJ 07107, USA

ARTICLE HISTORYReceived: December 18, 2020
Revised: May 12, 2021
Accepted: June 04, 2021DOI:
[10.2174/1389202922666210705124359](https://doi.org/10.2174/1389202922666210705124359)

Abstract: Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) technology have brought on substantial strides in predicting and identifying health emergencies, disease populations, and disease state and immune response, amongst a few. Although, skepticism remains regarding the practical application and interpretation of results from ML-based approaches in healthcare settings, the inclusion of these approaches is increasing at a rapid pace. Here we provide a brief overview of machine learning-based approaches and learning algorithms including supervised, unsupervised, and reinforcement learning along with examples. Second, we discuss the application of ML in several healthcare fields, including radiology, genetics, electronic health records, and neuroimaging. We also briefly discuss the risks and challenges of ML application to healthcare such as system privacy and ethical concerns and provide suggestions for future applications.

Keywords: Machine learning, healthcare, support vector machine, EHR, genomics, artificial intelligence.

1. INTRODUCTION

The application of machine learning dates back to the 1950s when Alan Turing proposed the first machine that can learn and become artificially intelligent [1]. Since its advent, machine learning has been used in various applications, ranging from security services through face detection [2] to increasing efficiency and decreasing risk in public transportation [3, 4], and recently in various aspects of healthcare and biotechnology [5-10]. Artificial intelligence and machine learning have brought significant changes in business processes and have transformed day-to-day lives, and comparable transformations are anticipated in healthcare and medicine. Recent advancements in this area have displayed incredible progress and opportunity to disburden physicians and improve accuracy, prediction, and quality of care. Current machine learning advancements in healthcare have primarily served as a supportive role in a physician or analyst's ability to fulfill their roles, identify healthcare trends, and develop disease prediction models. In large medical organizations, machine learning-based approaches have also been implemented to achieve increased efficiency in the organization of electronic health records [11], identification of irregularities in the blood samples [5], organs [6-8], and bones [12] using medical imaging and monitoring, as well as in robot-assisted surgeries [9, 13]. Machine learning applications have recently enabled the acceleration of testing and hospital response in the battle against COVID-19. Hospitals have been able to organize, share, and track patients, beds, rooms, ventilators, EHRs, and even staff during the pandemic using a deep learning system by GE called the Clinical

Command Center [14]. Researchers have also used artificial intelligence for the identification of genetic sequences of SARS-CoV2 and the creation of vaccines as well as for their monitoring [15].

Many new developments emerge as the field of healthcare grows into the new world of technology. Artificial intelligence and machine learning-based approaches and applications are vital for the field's progression, including increased speed of diagnosis, accuracy, and simplicity. The purpose of this review is to highlight the advantages and disadvantages of machine learning-based approaches in the healthcare industry. As the application of new machine learning technology takes the healthcare industry by storm, we aim to provide a brief overview of the various approaches to machine learning and highlight the fields where these approaches are primarily applied. We discuss their widespread use and future advancement opportunities in healthcare. We also address the ethical and logistical risks and challenges that occur with their application.

2. OVERVIEW OF ARTIFICIAL INTELLIGENCE

Although the terms machine learning, deep learning, and artificial intelligence are typically used interchangeably, they represent different sets of algorithms and learning processes. Artificial Intelligence (AI) is the umbrella term that refers to any computerized intelligence that learns and imitates human intelligence [16]. AI is most regarded for autonomous machines such as robots and self-driving cars, but it also permeates everyday applications, such as personalized advertisements and web searches. In recent years, AI development and application have made incredible strides and have been applied to many areas due to their higher levels of decision-making, accuracy, problem-solving capability, and computational skills [17]. In generally all development of AI

*Address correspondence to this author at the Department of Health Informatics, Rutgers University School of Health Professions, 65 Bergen Street, Newark, NJ 07107, USA; Tel: 973-972-5039; E-mail: suril.gohel@rutgers.edu

algorithms, the data obtained is split into two groups, a training and test data set, to ensure reliable learning, representative populations, and unbiased predictions. As the name suggests, the training set is used for algorithm training that includes sets of characterizing data points (features) and corresponding predictions (in the case of supervised learning). The testing data set is new to the algorithm and is solely used to test the algorithm's abilities. This measure is taken to eliminate biases in the algorithm's testing by the training dataset [18]. Once an algorithm passes through a training and testing phase with acceptable results, the algorithms are implemented in healthcare settings. The application of AI is broad and has many applied sub-regions; here, we provide an overview of machine learning and deep learning, two of the several sub-regions of AI.

Machine learning encompasses several different algorithmic models and statistical methods to solve problems without specialized programming [19]. Several machine learning models are single-layered, therefore, large components of feature extraction and data processing are performed prior to inputting the data into the algorithm [20]. Without the extra layers, these machine learning algorithms require intense data preprocessing in order for the algorithms to determine accurate predictions and to avoid over-fitting or under-fitting the training dataset. Deep learning is a more elaborate sub-form of machine learning that utilizes layered artificial neural networks and provides increased accuracy and specificity with decreased interpretability [21]. The neuronal network method is characterized as the multilayer network that supports the connection between the artificial neurons, or units, in each layer with that of the layer before and after it [22]. These networks can learn, discern, and deduce from data on their own using these multilevel links for data processing, and the data are processed until the specialized results are achieved [21].

2.1. Types of Learning Approaches

Most of the machine learning and AI-based algorithms are built on different learning approaches. One subtype is supervised learning, which is used in training classification and prediction algorithms based on previous examples, or outputs. An important distinction for this learning technique is that the training set involves features and corresponding predictions, or outcomes. Simply put, the supervised learning approach generalizes information from the training set's features to construct a model that can correctly predict training-set outcomes and then uses the learned model to make predictions using the new features in the testing data set [20]. Decision Trees, Random Forest, Support Vector Machines, and Artificial Neural Networks are a few types of ML algorithms that implement supervised learning approaches. Decision tree algorithms form a decision support tool that begins with a single node and identifies the possible outcomes of that decision. The tree continues with the product of that decision and the following decisions until it reaches a final product [23]. Support Vector Machines (SVM) are known as classification algorithms that use supervised learning to classify features in two group problems by finding the

largest margin hyperplane to separate the data and providing the best fit to organize it [16, 24]. Artificial Neural Networks (ANNs) consist of an input layer, one or more hidden layers, and output layers, where functional unities/neurons in one layer are connected to every neuron in the layer before and after [25]. In healthcare, supervised machine learning approaches are widely implemented in disease prediction [26], identifying hospital outcomes [14], and image detection [27] to name a few.

Another subtype of AI-based learning approaches is unsupervised learning, which is typically used to evaluate data and to cluster applications. Unsupervised machine learning is usually purposeful in data analysis, stratification, and reduction rather than prediction. In general, unsupervised clustering methods use algorithms to group data that has not been classified or categorized into independent clusters. Although data preprocessing and feature extraction are done before the input in most forms of machine learning, this method allows for the extraction of features and explores possibilities of data clusters by identifying the underlying relationships or features in the data, then grouping them by their similarities [18]. Some unsupervised learning approaches include the k-Means algorithm, Deep Belief Networks, and Convolutional Neural Networks. The most common unsupervised learning algorithm is the k-Means algorithm that is used as a clustering method to identify the mean between groups within unlabeled datasets and create groups based on the mean [18]. A Deep Belief Network (DBN) is a multi-layer network consisting of intra-level connections useful for data retrieval that typically uses unsupervised learning and has many hidden layers tasked with feature detection and finding correlations in the data [28, 29]. A Convolutional Neural Network (CNN) is a multilayer network that relies on feature recognition and identification and is useful for anomaly detection, image recognition, and identification [25]. Many unsupervised algorithms are used for clustering due to the lack of predetermined results and homogeneity in the data, and although the unsupervised methods are useful and quick, they are only semipopular in healthcare.

Reinforcement learning is another learning method that is neither supervised nor unsupervised learning. Similar to the mechanisms of conditioning in psychology, this learning depends on the sequences of rewards, and it forms a strategy for operation in a specific problem space. Reinforcement learning methods have the potential to influence their environment, are geared towards optimizing the error criterion, and have been described as the closest form of learning as seen in humans and animals [30]. Given the types of learning approaches, the selection of learning methods is relatively less complicated than the selection of algorithms and is usually dictated by the implementation purpose. A commonly used neural network that uses reinforcement learning is the Recurrent Neural Network (RNN). An RNN is one of the neural networks in which every artificial neuron is connected; the artificial neurons can receive inputs with delays in time and can reuse outputs from previous steps as input for a future step. It is useful for time series prediction, translation, speech recognition, rhythm learning, and music com-

position [25]. Although healthcare applications of reinforcement learning remain limited due to its need for structure, heterogeneous data, definition and implementation of rewards, and extensive computational resources, it still possesses the significant potential to bring major strides in healthcare.

Given the several types of machine learning and deep learning approaches, it is highly imperative to identify and implement an approach suitable specific to the healthcare application. Several factors, including the number of features [28], sample size [31, 32], and data distributions [33], can have significant effects on the learning and prediction processes and should be considered.

3. AI IN HEALTHCARE

In healthcare, common machine learning advances have been evolving for years. The application of AI has the capacity to assist with case triage and diagnoses [26], enhance image scanning and segmentation [34], support decision mak-

ing [11], predict the risk of disease [35, 36], and in neuroimaging [37]. Here we provide a brief overview of current advances in AI applications to specific aspects of health science. Inclusion criteria for the applications mentioned are based on the higher availability of digital data used in the ML-based approaches and their clear implementation of learning approaches with clinical applications and experiments. In the current review, we focused on ML application to healthcare in the fields of electronic health records, medical imaging, and genetic engineering. These areas also represent healthcare's "BIG" data, or the structured and unstructured data of the field, and have shown significant promise in relation to clinical applications.

Our search strategy is as follows: articles between June and December 2020, online libraries and journal databases including, but not limited to, and Academic OneFile, Gale, Nature, Sage Journals, Science Direct, PsycNet, and Pubmed were used. The compilation of articles and papers focused on the use of machine learning and artificial intelli-

Table 1. List of primary references.

Healthcare Area	Type of Machine Learning Model	Description	Applied or Experiment	References
EHRs	SVM, DT	Using EHRs for predicting diagnoses	Applied	Liang <i>et al.</i> 2014 [26]
-	RNN	Predicting post-stroke pneumonia using deep neural network approaches	Experiment	Ge <i>et al.</i> , 2019 [35]
-	LSTM, CNN	Deep EHR: Chronic Disease Prediction Using Medical Notes	Experiment	Liu, Zhang & Razavian 2018 [40]
-	ML	SRML-Mortality Predictor : A hybrid machine learning framework to predict mortality in paralytic ileus patients using Electronic Health Records (EHRs)	Experiment	Ahmad <i>et al.</i> , 2020 [41]
Medical Imaging	CNN	Dermatologist-level classification of skin cancer with deep neural networks	Experiment	Esteva <i>et al.</i> 2017 [7]
-	CNN	Chexnet : Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep-Learning	Applied	Rajpurkar <i>et al.</i> , 2017; Tsai & Tao, 2019 [8]
-	CNN	International evaluation of an AI system for breast cancer screening	Experiment	McKinney <i>et al.</i> 2020 [49]
-	Deep CNN	Deep-learning algorithm predicts diabetic retinopathy progression in individual patients	Experiment	Arcadu <i>et al.</i> 2019 [56]
-	DBN	Structural MRI classification for Alzheimer's disease detection using deep belief network	Experiment	Faturrahman <i>et al.</i> , 2017 [37]
-	Decision tree	Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction	Experiment	Patel <i>et al.</i> , 2015 [27]
Genetic Engineering & Genomics	RT	Application of machine learning models to predict tacrolimus stable dose in renal transplant recipients	Experiment	Tang <i>et al.</i> 2017 [10]
-	ML	Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2 leading to universal blueprints for vaccine designs	Applied	Malone <i>et al.</i> 2020 [15]
-	Deep CNN, Deep FFs	Off-target predictions in CRISPR-Cas9 gene editing using deep learning	Applied	Lin & Wong 2018 [76]
-	RNNs	DeepHF : Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning	Applied	Wang <i>et al.</i> , 2019 [85]
-	Random Forest	CUNE : Unlocking HDR-mediated nucleotide editing by identifying high-efficiency target sites using machine learning	Applied	O'Brien <i>et al.</i> , 2019 [86]
-	CNNs	ToxDL : deep learning using primary structure and domain embeddings for assessing protein toxicity	Applied	Pan <i>et al.</i> , 2020 [87]

Applied is defined as an algorithm or application that is currently available on a public or private platform to healthcare professionals. It also refers to applications that are currently applied in medical practices such as clinics, hospitals, *etc.* An experiment is defined as an algorithm or application that has been used in a research study. EHR: Electronic Health Records, SVM: Support Vector Machine, LSTM: Long Short-Term Memory Neural Network, CNN: Convolutional Neural Network, MLP: Multi-Layer perceptron Neural Network, RNN: Recurrent Neural network, DBN: Deep Belief Network, ANN: Artificial Neural Network, ML: Machine Learning.

gence in healthcare as well as current and potential applications. Search terms included machine learning in healthcare, artificial intelligence medical imaging, BIG data and machine learning, machine learning in genomics, electronic health records, challenges of AI in healthcare, and medical applications of AI. Variations of these terms were used to ensure exhaustive search results. Searches were not limited by year or journal Table 1.

3.1. Electronic Health Records

Electronic Health Records (EHRs), originally known as clinical information systems, were first introduced by Lockheed in the 1960s [38]. Since then, the systems have been reconstructed many times to create an industry-wide standard system. In 2009, the US federal government invested billions in promoting EHR implementation in all practices in an effort to improve the quality and efficiency of the work; this ultimately resulted in nearly 87 percent of office-based practices nationwide implementing EHRs in their systems by 2015 [39]. BIG data collected from EHR systems with structured feature data have been instrumental in deep learning applications, including medication refills and using patient history for predicting diagnoses [11]. This has resulted in significant improvement in data organization, accessibility, and quality of care and has helped physicians with diagnoses and treatments. The standardization of features across datasets has also allowed for increased access to health records for research purposes.

Considering the vital role that prediction plays in providing treatment, scientists have developed deep learning models for the diagnosis and prediction of clinical conditions using EHRs. In a recent research study, Liu, Zhang, and Razavian developed a deep learning algorithm using LSTM networks (reinforcement learning) and CNNs (supervised learning) to predict the onset of diseases, such as heart failure, kidney failure, and stroke. Unlike other prediction models, this algorithm used both structured data obtained from EHR and unstructured data contained in progress and diagnosis notes. As explained by Liu and colleagues, the inclusion of unstructured data within the model resulted in significant improvements in all the baseline accuracy measures, further indicating the versatility and robustness of such algorithms [40]. In another research study using deep neural network approaches, Ge and colleagues built a model to predict post-stroke pneumonia within 7 and 14-day periods. The model returned an Area under the ROC curve (AUC, a measure of model performance by combining sensitivity and specificity of a model) value of 92.8 percent for the 7-day predictions and 90.5 percent for the 14-day predictions [35], providing a highly accurate model predicting pneumonia following a stroke. In addition, several ML-based models have also been implemented to predict mortality in ICU patients. In one of such models, Ahmad and colleagues have shown great ability to predict mortality in paralytic ileus (PI, incomplete blockage of the intestine that prohibits the passage of food, eventually leading to a build-up and complete blockage of the intestines) patients using EHRs. The algorithm, named Statistically Robust Machine Learning-based Mortality Pre-

dictor (SRML-Mortality Predictor), showed an 81.30% accuracy rate in predicting mortality in PI patients [41]. Providing patients and practitioners with predicted mortality, through the use of EHR prediction algorithms, can allow them to make more educated clinical treatment decisions.

3.2. Medical Imaging

Given the digital nature of data and the presence of structured data formats such as DICOM (Digital Imaging and Communications in Medicine), medical imaging has seen significant strides with the implementation of machine learning-based approaches to several imaging modalities, including Computed Tomography (CT), Magnetic Resonance Imaging (MRI), X-Ray, Positron Emission Tomography (PET), Ultrasound, and more. Several ML-based models have been developed to identify tumors [42, 43], lesions [44], fractures [45, 46], and tears [47, 48].

In a recent study, McKinney and colleagues have implemented a deep learning algorithm to detect tumors based on Mammograms in earlier stages of growth. In comparison to traditional screening techniques used to identify tumors, these deep learning-based screen techniques allow for the identification and location of tumors in earlier stages of breast cancer, allowing for a better rate of resection. In a direct comparison, the deep learning-based approach was able to outperform experienced radiologists by an AUC score of 11.5% [49]. Several other studies have also implemented ML-based approaches for breast cancer detection with variable success, including models by Wang and colleagues [50], Amrane and colleagues [51], and Ahmad and colleagues [52].

Similarly, in a recent study, Esteva and colleagues used CNN (unsupervised learning) to classify 2032 different skin diseases using dermoscopic images. An objective comparison of CNN classification with that of 21 board-certified dermatologists resulted in “on par” performance, further confirming the veracity of the results [7]. When implemented in conjunction with the average consumer mobile platform, this approach can result in ease of use and early diagnosis. In parallel, studies have also implemented ML-based approaches to quantify the progression of retinal diseases [51-54]. In one such study, Arcadu and colleagues applied a deep learning CNN to detect the aneurysms that cause vision loss due to the progression of Diabetic Retinopathy (DR) [55]. The CNN was also able to detect small and low contrast microaneurysms, although it was not explicitly designed to accomplish that task [55, 56]. Given that diabetic retinopathy is a common eye condition that affects around 60 percent of type 1 diabetes patients [57], it is difficult to detect in its preliminary stages. Early prediction obtained using a CNN approach has the potential to prevent and delay irreversible damage to patients' vision. X-rays have been used for decades to identify abnormalities in the chest cavity and lung disease, though an in-depth careful examination by a training radiologist is often required. In a recent study, Rajpurkar and colleagues conducted a retrospective study to explore the capacities of a 121-layer convolutional neural net-

work to examine a collection of chest x-rays with various thoracic diseases and identify irregularities in an attempt to mimic the detection by trained radiologists [8]. In comparison, CNN's performance in the accuracy of identification observed was 81%, which was 2% higher than that of the radiologists. Although applied retrospectively, this study, along with CNNs developed by Tsai and Tao [58], Asif and colleagues [59], Liang and colleagues [60], and Lee and colleagues [61], indicates incredible support that these approaches can provide in examining and diagnosing illnesses, further reducing the burden on healthcare professionals.

ML-based approaches have also been implemented to predict and diagnose disease progression of neurodegenerative diseases, including Alzheimer's disease [37, 62], Parkinson's disease [63, 64], serious mental disorders including Psychosis, [65, 66], depression [27, 67], PTSD [68], and developmental disorders, including autism [69, 70] and ADHD [71, 72]. In one such study, Faturrahman and colleagues presented a higher-level model using DBNs (unsupervised learning) for predicting Alzheimer's Disease (AD) progression using structural MRI images, resulting in 91.76% accuracy, 90.59% sensitivity, and 92.96% specificity [37]. Although there is no cure for AD, early diagnosis can help implement strategies to delay the symptoms and degeneration. Using decision tree models and feature-rich data sets consisting of functional MRI, cognitive behavior scores, and age, Patel and colleagues developed a model to predict the diagnosis and treatment response for depression. The model scored 87.27% accuracy for diagnosis and 89.47% accuracy for treatment response [27]. This predictive diagnosis can help identify patients with depression and develop personalized treatment plans based on their responses. With the current ML applications in medical imaging, it is evident that its use has valuable implications for advancing the medical field due to its pronounced advantages in accuracy, classification, sensitivity, and specificity in prediction and diagnoses.

3.3. Genetic Engineering and Genomics

The discovery of the adaptive DNA system known as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) has cultivated the field of genetic engineering [73]. This exploration of "programmable endonucleases" has simplified genetic engineering and has helped make the process of genetic modification and diagnosis easier, as well as dropping the cost of the procedure dramatically [74]. The recent application of CRISPR to Cas (CRISPR-associated protein) editing, such as Cas-9 [75] and Cas-13a [73], has changed genetic editing, though the tool is not perfect. Recently, several machine learning techniques for predicting off-target mutations in Cas9 gene editing have emerged. A new program developed by Jiecong Lin and Ka-Chun Wong has improved the quality of these machine learning predictions by using deep CNNs (AUC score: 97.2%) and deep FFs (AUC score: 97%) [76]. Considering the space for error and off-target mutations using the Cas9 tool, scientists are using Cas9 for developing activity predictors and more reliable Cas9 variants to reduce error. These models include higher accuracy and fidelity Cas9 variants [77-79], hyper-ac-

curate Cas9 variants [80], and guide RNA design tools using deep learning [81-85].

Outside of CRISPR gene editing, O'Brien and colleagues have developed a service to provide efficiency in nucleotide editing using random forest algorithms (supervised learning) to investigate how different nucleotide compositions influence the HDR (homology-directed repair) efficiency [86]. They developed the Computational Universal Nucleotide Editor (CUNE), used to find the most efficient method to identify a precise location to enter a specific point mutation and predict HDR efficiency. Additionally, Pan and colleagues have developed a model for prediction in gene editing named ToxDL that uses a CNN approach to predict protein toxicity *in-vivo* using only the sequence data [87]. Another branch of genetic engineering, pharmacogenomics, has also made significant strides in the use of AI and machine learning to determine stable doses of medications that have become popular [88-90]. In one such study, Tang and colleagues implemented an ML-based approach to determine a stable Tacrolimus dose (the immunosuppressive drug) for patients who received a renal transplant to reduce the risk of acute rejection [10]. The use of machine learning in pharmacogenomics has recently been applied in psychiatry [90], oncology [91], bariatrics [92], and neurology [93].

Machine learning applications of genetic engineering have also been instrumental in the fight against COVID 19. In a recent study, Malone and colleagues utilized software based on machine learning algorithms to "predict which antigens have the required features of HLA-binding, processing, presentation to the cell surface, and the potential to be recognized by T cells to be good clinical targets for immunotherapy" [15]. The use of immunogenicity predictions from this software, along with the presentation of antigen to infected host-cells, allowed the team to successfully profile the "entire SARS-CoV2 proteome" as well as epitope hotspots. These discoveries help predict blueprints for designing universal vaccines against the virus that can be adapted across the global population.

4. RISKS AND CHALLENGES

While machine learning-based applications in healthcare present unique and progressive opportunities, they also raise unique risk factors, challenges, and healthy skepticism. Here we discuss the main risk factors including the probability of error in prediction and its impact, the vulnerability of the systems' protection and privacy, and even the lack of data availability to obtain reproducible results. Some of the challenges include ethical concerns, loss of the personal element of healthcare, and the interpretability and practical application of the approaches to bedside setting.

One of the most important risks of machine learning-based algorithms is the reliance on the probabilistic distribution and the probability of error in diagnosis and prediction. This also gives rise to a healthy skepticism related to the validity and veracity of predictions from ML-based approaches. Even though the probability of error and reliance on probability is deep-rooted in the various aspects of health

care, the implications of ML-based approaches resulting in a human fatality are severe. One solution is to subject these machine learning-based approaches to strict institutional and legal approval by several organizations before their application [94, 95]. Another approach that can be implemented is the human intervention and oversight from an experienced healthcare worker in highly sensitive applications to avoid false-positive or false-negative diagnoses (*e.g.*, diagnosis of depression or breast cancer). The inclusion of present healthcare professionals in developing and implementing these approaches may increase adaptation rates and decrease concerns related to fewer employment opportunities for humans or the shrinking of the workforce [96].

Another risk associated with the application of ML and deep learning algorithms to health care is the availability of high-quality training and testing data with large enough sample sizes to ensure high reliability and reproducibility of the predictions. Given that the ML and deep learning-based approaches 'learn' from data, the importance of quality data cannot be stressed enough. In addition, the large amounts of feature-rich data required for these learning networks and approaches are not readily available and may also represent a narrow distribution of the population sample. Moreover, in several healthcare segments, data collected are incomplete, heterogeneous, and have a significantly higher number of features than the number of samples. These challenges should be taken into great consideration when developing and interpreting the results of ML-based approaches. The open science and recent push towards research data sharing may assist in overcoming such challenges. One should also consider the risk associated with privacy as well as ethical implications of the application of ML-based approaches to healthcare. With the understanding that these approaches require large-scale, easily expandable data storage, and significantly high computing power, several ML-based approaches are developed and implemented using cloud-based technologies. Given the sensitive nature of healthcare data along with privacy concerns, increased data security and accountability should be one of the first aspects to be considered well before model development.

With respect to ethical concerns, researchers working on applying ML-based approaches to healthcare can readily learn from the field of genetic engineering which has undergone extensive ethical debate. The controversy surrounding the use of genetic engineering to create long-lasting genetic advancements and treatments is a continuous discourse. Identification and editing of injurious genetic mutations, such as the HTT mutation that causes Huntington's disease, may provide life-altering treatment for harmful diseases [97]. Contrarily, creating treatments that alter the individual's genome, as well as that of their offspring, while it is still inaccessible due to costs, may worsen the socio-economic divide for populations that are unable to afford such care [98]. Recently, there has been an emergence of guidelines for the development of AI machinery. In 2019, Singapore proposed a Model Artificial Intelligence Governance Framework to guide private sector organizations on developing and using AI ethically [99]. The US Administration has also released

an executive order to regulate AI development and "maintain American leadership in artificial intelligence" [100]. These guidelines and regulations, though strict, have been put forth to ensure ethical research conduct and development.

An important challenge with ML application to healthcare is associated with the interpretation and clinical applicability of the results. Given the complex structure of ML-based approaches, especially deep learning-based methods, it becomes incredibly complex to distinguish and identify the original features' contribution towards the prediction. Although this may not present significant concern in other applications of ML (such as web searches), lack of transparency has created a huge barrier for the adaptability of ML-based approaches in healthcare. As clearly understood in healthcare, the solution strategy is as important as the solution itself. There must be a systematic shift towards identifying and quantifying underlying data features used for prediction. The involvement of physicians and healthcare professionals in the development, implementation, and testing of ML-based approaches may also help improve the adoption rates. Additionally, although there is healthy skepticism related to the potential of a decreased personal relationship between a patient and PCP due to increased implementation of ML-based approaches, they represent a unique opportunity to increase engagement. Studies have shown that the physician-patient relationship has already become a fading concept, and nearly 25 percent of Americans do not have a PCP [101]. Here, ML can provide unique opportunities to increase engagement where patients discuss the results of potential diagnoses and increase the efficiency of outreach programs. Early prognosis due to ML-based approaches may also help patients develop a healthy lifestyle in consultations with their PCPs. Finally, a physician-focused survey found that 56 percent of physicians were spending 16 minutes or less with their patients, and 5 percent of them spent less than 9 minutes [102]. The application of AI approaches in diagnoses and symptom monitoring can ease stress and give physicians more personal time with their patients, thus improving patient satisfaction and outcomes.

CONCLUSION

While the overview demonstrates how much progress has been achieved with machine learning, there continues to be potential for widescale advancement in the future. Many of the current machine learning advancements in healthcare aim to support the physician's or specialist's ability to provide a more effective treatment to patients with increased quality, speed, and precision. The challenges of developing ML algorithms can be solved by developing and implementing improvements in data collection, storage, and dissemination or by creating algorithms to process unstructured data to address the lack of data availability. Future applications can also bring forth inexpensive forms of medical imaging and affordable medical examinations, potentially ending health disparities and creating more accessible services for countries and lower-income populations. Scientists expect advancement in the prediction of personalized drug response, optimization of medication selection and dosage, and an ap-

plication of genetic modification to provide treatment for genetic disorders and mutations [103]. With its application, ML can augment the role of physicians and redefine patient care. While the risks and challenges of the future application are addressed and corrected, the current ML algorithms can provide an excellent framework for future advancements and applications of ML in healthcare.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

This work was partly supported by NIH/NCATS UL1TR003017. This work is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NCATS.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Turing, A. Computing machinery and intelligence. *Mind*, **1950**, *LIX*(236), 433-460. <http://dx.doi.org/10.1093/mind/LIX.236.433>
- [2] Wati, D.A.R.; Abadianto, D. Design of face detection and recognition system for smart home security application. *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia, November 1-2, **2018**, pp. 342-347.
- [3] Ellis, K.; Godbole, S.; Marshall, S.; Lanckriet, G.; Staudenmayer, J.; Kerr, J. Identifying active travel behaviors in challenging environments using GPS, accelerometers, and machine learning algorithms. *Front. Public Health*, **2014**, *2*, 36. <http://dx.doi.org/10.3389/fpubh.2014.00036> PMID: 24795875
- [4] Omrani, H. Predicting travel mode of individuals by machine learning. *Transp. Res. Procedia*, **2015**, *10*, 840-849. <http://dx.doi.org/10.1016/j.trpro.2015.09.037>
- [5] Siddiqui, M.K.; Morales-Menendez, R.; Huang, X.; Hussain, N. A review of epileptic seizure detection using machine learning classifiers. *Brain Inform.*, **2020**, *7*(1), 5. <http://dx.doi.org/10.1186/s40708-020-00105-1> PMID: 32451639
- [6] Woldaregay, A.Z.; Årsand, E.; Botsis, T.; Albers, D.; Mamykina, L.; Hartvigsen, G. Data-driven blood glucose pattern classification and anomalies detection: Machine-learning applications in type 1 diabetes. *J. Med. Internet Res.*, **2019**, *21*(5), e11030. <http://dx.doi.org/10.2196/11030> PMID: 31042157
- [7] Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, **2017**, *542*(7639), 115-118. <http://dx.doi.org/10.1038/nature21056> PMID: 28117445
- [8] Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv*, **2017**, 3-9.
- [9] Kaouk, J.H.; Garisto, J.; Eltemamy, M.; Bertolo, R. Robot-assisted surgery for benign distal ureteral strictures: step-by-step technique using the SP[®] surgical system. *BJU Int.*, **2019**, *123*(4), 733-739. <http://dx.doi.org/10.1111/bju.14635> PMID: 30499629
- [10] Tang, J.; Liu, R.; Zhang, Y.L.; Liu, M.Z.; Hu, Y.F.; Shao, M.J.; Zhu, L.J.; Xin, H.W.; Feng, G.W.; Shang, W.J.; Meng, X.G.; Zhang, L.R.; Ming, Y.Z.; Zhang, W. Application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients. *Sci. Rep.*, **2017**, *7*, 42192. <http://dx.doi.org/10.1038/srep42192> PMID: 28176850
- [11] Rao, S.R.; Desroches, C.M.; Donelan, K.; Campbell, E.G.; Miralles, P.D.; Jha, A.K. Electronic health records in small physician practices: availability, use, and perceived benefits. *J. Am. Med. Assoc.*, **2011**, *18*(3), 271-275. <http://dx.doi.org/10.1136/amiajnl-2010-000010> PMID: 21486885
- [12] Tian, L.; Zhang, D.; Bao, S.; Nie, P.; Hao, D.; Liu, Y. Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas. *Clin. Radiol.*, **2020**, *76*(2), 158.e19-158.e25. PMID: 33293024
- [13] Lanfranco, A.R.; Castellanos, A.E.; Desai, J.P.; Meyers, W.C. Robotic surgery: A current perspective. *Ann. Surg.*, **2004**, *239*(1), 14-21. <http://dx.doi.org/10.1097/01.sla.0000103020.19595.7d> PMID: 14685095
- [14] Roth, M. COVID-19: Oregon hospitals share data, create real-time bed capacity system. *Health Leaders Media*, **2020**. Available from: <https://www.healthleadersmedia.com/innovation/covid-19-oregon-hospitals-share-data-create-real-time-bed-capacity-system>
- [15] Malone, B.; Simovski, B.; Moliné, C.; Cheng, J.; Gheorghe, M.; Fontenelle, H.; Vardaxis, I.; Tennøe, S.; Malmberg, J.A.; Stratford, R.; Clancy, T. Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2 leading to universal blueprints for vaccine designs. *Sci. Rep.*, **2020**, *10*(1), 22375. <http://dx.doi.org/10.1038/s41598-020-78758-5> PMID: 33361777
- [16] Lee, E.J.; Kim, Y.H.; Kim, N.; Kang, D.W. Deep into the brain: Artificial intelligence in stroke imaging. *J. Stroke*, **2017**, *19*(3), 277-285. <http://dx.doi.org/10.5853/jos.2017.02054> PMID: 29037014
- [17] Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinform.*, **2018**, *19*(6), 1236-1246. <http://dx.doi.org/10.1093/bib/bbx044> PMID: 28481991
- [18] Alloghani, M.; Al-Jumeily, D.; Mustafina, J.; Hussain, A.; Aljaaf, A.J. A systematic review on supervised and unsupervised machine learning algorithms for data science. In: *Supervised and unsupervised learning for data science*; Berry, M.; Mohamed, A.; Yap, B., Eds.; Springer, US, **2020**; pp. 3-21. http://dx.doi.org/10.1007/978-3-030-22475-2_1
- [19] Samuel, A.L. Some studies in machine learning using the game of checkers. *IBM J. Res. Develop.*, **2000**, *44*(1-2), 207-219. <http://dx.doi.org/10.1147/rd.441.0206>
- [20] Alpaydin, E. Introduction to machine learning. MIT Press, **2020**. Available from: <https://books.google.com/books?hl=en&lr=&id=tZnSDwAAQBAJ&oi=fnd&pg=PR7&dq=Introduction+to+machine+learning&ots=F3RWwXcwwf&sig=50DHyeJhVdDtmX1xZ0C4tXOGsdw>
- [21] Bhattacharjee, B.; Boag, S.; Doshi, C.; Dube, P.; Herta, B.; Ishakian, V. IBM deep learning service. *arXiv*, **2017**, 1-22.
- [22] Qin, Z.; Ye, H.; Li, G.Y.; Juang, B.H.F. Deep learning in physical layer communications. *IEEE Wirel. Commun.*, **2019**, *26*(2), 93-99. <http://dx.doi.org/10.1109/MWC.2019.1800601>
- [23] Kamiński, B.; Jakubczyk, M.; Szufel, P. A framework for sensitivity analysis of decision trees. *Cent. Eur. J. Oper. Res.*, **2018**, *26*(1), 135-159. <http://dx.doi.org/10.1007/s10100-017-0479-6> PMID: 29375266
- [24] Gaonkar, B.; Davatzikos, C. Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. *Neuroimage*, **2013**, *78*, 270-283. <http://dx.doi.org/10.1016/j.neuroimage.2013.03.066> PMID: 23583748
- [25] Leijnen, S.; van Veen, F. The neural network zoo. *Proceedings.*, **2020**, *47*(1), 9. <http://dx.doi.org/10.3390/proceedings2020047009>
- [26] Liang, Z.; Zhang, G.; Huang, J.X.; Hu, Q.V. Deep learning for healthcare decision making with EMRs. *Proc - 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Belfast, UK, November 2-5, **2014**, pp. 556-559. <http://dx.doi.org/10.1109/BIBM.2014.6999219>

- [27] Patel, M.J.; Andreescu, C.; Price, J.C.; Edelman, K.L.; Reynolds, C.F., III; Aizenstein, H.J. Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. *Int. J. Geriatr. Psychiatr.*, **2015**, *30*(10), 1056-1067.
<http://dx.doi.org/10.1002/gps.4262> PMID: 25689482
- [28] Coates, A.; Lee, H.; Ng, A.Y. An analysis of single-layer networks in unsupervised feature learning. *J. Mach. Learn. Res.*, **2011**, *15*, 215-223.
- [29] Ravi, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; Yang, G.Z. Deep learning for health informatics. *IEEE J. Biomed. Health Inform.*, **2017**, *21*(1), 4-21.
<http://dx.doi.org/10.1109/JBHI.2016.2636665> PMID: 28055930
- [30] Sutton, R.S.; Barto, A. *Reinforcement Learning: An Introduction*; MIT Press, **2018**.
- [31] Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS One*, **2019**, *14*(11), e0224365.
<http://dx.doi.org/10.1371/journal.pone.0224365> PMID: 31697686
- [32] Cui, Z.; Gong, G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage*, **2018**, *178*, 622-637.
<http://dx.doi.org/10.1016/j.neuroimage.2018.06.001> PMID: 29870817
- [33] Gu, L.; Zeng, D.; Li, P.; Guo, S. Cost minimization for big data processing in geo-distributed data centers. *IEEE Trans. Emerg. Top. Comput.*, **2014**, *2*(3), 314-323.
<http://dx.doi.org/10.1109/TETC.2014.2310456>
- [34] Li, W.; Jia, F.; Hu, Q. Automatic segmentation of liver tumor in ct images with deep convolutional neural networks. *J. Comput. Commun.*, **2015**, *03*(11), 146-151.
<http://dx.doi.org/10.4236/jcc.2015.311023>
- [35] Ge, Y.; Wang, Q.; Wang, L.; Wu, H.; Peng, C.; Wang, J. Predicting post-stroke pneumonia using deep neural network approaches. *Int. J. Med. Inform.*, **2019**, *132*, 103986.
<http://dx.doi.org/10.1016/j.ijmedinf.2019.103986>
- [36] Nguyen, B.P.; Pham, H.N.; Tran, H.; Nghiem, N.; Nguyen, Q.H.; Do, T.T.T.; Tran, C.T.; Simpson, C.R. Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. *Comput. Methods Programs Biomed.*, **2019**, *182*, 105055.
<http://dx.doi.org/10.1016/j.cmpb.2019.105055> PMID: 31505379
- [37] Faturrahman, M.; Wasito, I.; Hanifah, N.; Mufidah, R. Structural MRI classification for Alzheimer's disease detection using deep belief network. *2017 11th International Conference on Information Communication Technology and System (ICTS)*, **2017**, pp. 37-42.
<http://dx.doi.org/10.1109/ICTS.2017.8265643>
- [38] Atherton, J. Development of the electronic health record. *Virtual Mentor*, **2011**, *13*(3), 186-189. Available from: <http://virtualmentor.ama-assn.org/2011/03/msoc1-1103.html> PMID: 23127323
- [39] Yang, N.; Hing, E. National electronic health records survey: 2015 specialty and overall physicians electronic health record adoption summary tables. **2017**, (28) Available from: http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm
- [40] Liu, J.; Zhang, Z.; Razavian, N. Deep EHR: Chronic disease prediction using medical notes. *arXiv*, **2018**, arXiv:1808.04928.
- [41] Ahmad, F.S.; Ali, L. A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (EHRs). *J. Ambient Intell. Humaniz. Comput.*, **2020**, *12*(2), 3283-3293.
<http://dx.doi.org/10.1007/s12652-020-02509-7>
- [42] Tang, T.T.; Zawaski, J.A.; Francis, K.N.; Qutub, A.A.; Gaber, M.W. Image-based classification of tumor type and growth rate using machine learning: A preclinical study. *Sci. Rep.*, **2019**, *9*(1), 12529.
<http://dx.doi.org/10.1038/s41598-019-48738-5> PMID: 31467303
- [43] Sharma, K.; Kaur, A.; Gujral, S. Brain tumor detection based on machine learning algorithms. *Int. J. Comput. Appl.*, **2014**, *103*(1), 7-11.
- [44] Karimi, D.; Peters, J.M.; Ouaalam, A.; Prabhu, S.P.; Sahin, M.; Krueger, D.A. Learning to detect brain lesions from noisy annotations. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, **2020**, pp. 1910-4.
<http://dx.doi.org/10.1109/ISBI45749.2020.9098599>
- [45] Ferizi, U.; Besser, H.; Hysi, P.; Jacobs, J.; Rajapakse, C.S.; Chen, C.; Saha, P.K.; Honig, S.; Chang, G. Artificial intelligence applied to osteoporosis: A performance comparison of machine learning algorithms in predicting fragility fractures from mri data. *J. Magn. Reson. Imaging*, **2019**, *49*(4), 1029-1038.
<http://dx.doi.org/10.1002/jmri.26280> PMID: 30252971
- [46] Zuccon, G.; Waghlikar, A.S.; Nguyen, A.N.; Butt, L.; Chu, K.; Martin, S. Automatic classification of free-text radiology reports to identify limb fractures using machine learning and the snomed ct ontology. *AMIA Jt. Summits Transl. Sci.*, **2013**, *2013*, pp. 300-304.
- [47] Chang, P.D.; Wong, T.T.; Rasiej, M.J. Deep learning for detection of complete anterior cruciate ligament tear. *J. Digit. Imaging*, **2019**, *32*(6), 980-986.
<http://dx.doi.org/10.1007/s10278-019-00193-4> PMID: 30859341
- [48] Lao, Y.; Jia, B.; Yan, P.; Pan, M.; Hui, X.; Li, J.; Luo, W.; Li, X.; Han, J.; Yan, P.; Yao, L. Diagnostic accuracy of machine-learning-assisted detection for anterior cruciate ligament injury based on magnetic resonance imaging: Protocol for a systematic review and meta-analysis. *Medicine (Baltimore)*, **2019**, *98*(50), e18324.
<http://dx.doi.org/10.1097/MD.00000000000018324> PMID: 31852123
- [49] McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.S.; Darzi, A.; Etemadi, M.; Garcia-Vicente, F.; Gilbert, F.J.; Halling-Brown, M.; Hassabis, D.; Jansen, S.; Karthikesalingam, A.; Kelly, C.J.; King, D.; Ledam, J.R.; Melnick, D.; Mostofi, H.; Peng, L.; Reichert, J.J.; Romera-Paredes, B.; Sidebottom, R.; Suleyman, M.; Tse, D.; Young, K.C.; De Fauw, J.; Shetty, S. International evaluation of an AI system for breast cancer screening. *Nature*, **2020**, *577*(7788), 89-94.
<http://dx.doi.org/10.1038/s41586-019-1799-6> PMID: 31894144
- [50] Wang, D.; Li, J.R.; Zhang, Y.H.; Chen, L.; Huang, T.; Cai, Y.D. Identification of differentially expressed genes between original breast cancer and xenograft using machine learning algorithms. *Genes (Basel)*, **2018**, *9*(3), 1-16.
<http://dx.doi.org/10.3390/genes9030155> PMID: 29534550
- [51] Amrane, M.; Oukid, S.; Gagaoua, I.; Ensar, İ.T. Breast cancer classification using machine learning. *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, **2018**, pp. 1-4.
<http://dx.doi.org/10.1109/EBBT.2018.8391453>
- [52] Ahmad, L.; Eshlaghy, A.; Poorebrahimi, A.; Ebrahimi, M.; Razaivi, A. Using three machine learning techniques for predicting breast cancer recurrence. *J. Health Med. Inform.*, **2013**, *04*(02), 2-4.
- [53] Huck, Y.C.H.; Liu, F.; Huang, J.H.; Tian, M.; Hung, I.; Lin, M.D.; Liu, Y.C. Auto-classification of retinal diseases in the limit of sparse data using a two-streams machine learning model. *Lect Notes Comput. Sci.*, **2018**, pp. 323-338.
- [54] Rampasek, L.; Goldenberg, A. Learning from everyday images enables expert-like diagnosis of retinal diseases. *Cell*, **2018**, *172*(5), 893-895.
<http://dx.doi.org/10.1016/j.cell.2018.02.013> PMID: 29474917
- [55] Alqudah, A.M. AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Med. Biol. Eng. Comput.*, **2020**, *58*(1), 41-53.
<http://dx.doi.org/10.1007/s11517-019-02066-y> PMID: 31728935
- [56] Arcadu, F.; Benmansour, F.; Maunz, A.; Willis, J.; Haskova, Z.; Prunotto, M. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit. Med.*, **2019**, *2*(1), 92.
- [57] Jansson, R.W.; Hufthammer, K.O.; Krohn, J. Diabetic retinopathy in type 1 diabetes patients in Western Norway. *Acta Ophthalmol.*, **2018**, *96*(5), 465-474.
<http://dx.doi.org/10.1111/aos.13654> PMID: 29369506
- [58] Tsai, M.; Tao, Y. Machine learning based common radiologist-level pneumonia detection on chest x-rays. *2019 13th International Conference on Signal Processing and Communication Systems*

- (ICSPCS), 2019, pp. 1-7.
<http://dx.doi.org/10.1109/ICSPCS47537.2019.9008684>
- [59] Asif, S.; Wenhui, Y.; Jin, H.; Tao, Y.; Jinhai, S. Classification of covid-19 from chest radiography images using deep convolutional neural network. *J. Xidian Univ.*, **2020**, *14*(8), 565-569.
- [60] Liang, S.; Tang, F.; Huang, X.; Yang, K.; Zhong, T.; Hu, R.; Liu, S.; Yuan, X.; Zhang, Y. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. *Eur. Radiol.*, **2019**, *29*(4), 1961-1967.
<http://dx.doi.org/10.1007/s00330-018-5748-9> PMID: 30302589
- [61] Lee, K.S.; Kim, J.Y.; Jeon, E.T.; Choi, W.S.; Kim, N.H.; Lee, K.Y. Evaluation of scalability and degree of fine-tuning of deep convolutional neural networks for covid-19 screening on chest x-ray images using explainable deep-learning algorithm. *J. Pers. Med.*, **2020**, *10*(4), 1-14.
<http://dx.doi.org/10.3390/jpm10040213> PMID: 33171723
- [62] Lee, G.; Nho, K.; Kang, B.; Sohn, K.A.; Kim, D.; Weiner, M.W. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci. Rep.*, **2019**, *9*(1), 1952.
<http://dx.doi.org/10.1038/s41598-018-37769-z> PMID: 30760848
- [63] Mandal, I.; Sairam, N. New machine-learning algorithms for prediction of Parkinson's disease. *Int. J. Syst. Sci.*, **2014**, *45*(3), 647-666.
<http://dx.doi.org/10.1080/00207721.2012.724114>
- [64] Ahmadi, R.D.; Ho, N.; Halliday, G.M.; Dzakmo, N. Parkinson's progression prediction using machine learning and serum cytokines. *NPJ Park Dis.*, **2019**, *5*(1), 1-8.
- [65] Ramyeed, A.; Studerus, E.; Kometer, M.; Uttinger, M.; Gschwandtner, U.; Fuhr, P.; Riecher-Rössler, A. Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients. *World J. Biol. Psychiatry*, **2016**, *17*(4), 285-295.
<http://dx.doi.org/10.3109/15622975.2015.1083614> PMID: 26453061
- [66] Corcoran, C.M.; Carrillo, F.; Fernández-Slezak, D.; Bedi, G.; Klim, C.; Javitt, D.C.; Bearden, C.E.; Cecchi, G.A. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*, **2018**, *17*(1), 67-75.
<http://dx.doi.org/10.1002/wps.20491> PMID: 29352548
- [67] Kessler, R.C.; van Loo, H.M.; Wardenaar, K.J.; Bossarte, R.M.; Brenner, L.A.; Cai, T.; Ebert, D.D.; Hwang, I.; Li, J.; de Jonge, P.; Nierenberg, A.A.; Petukhova, M.V.; Rosellini, A.J.; Sampson, N.A.; Schoevers, R.A.; Wilcox, M.A.; Zaslavsky, A.M. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol. Psychiatry*, **2016**, *21*(10), 1366-1371.
<http://dx.doi.org/10.1038/mp.2015.198> PMID: 26728563
- [68] Kim, Y.W.; Kim, S.; Shim, M.; Jin, M.J.; Jeon, H.; Lee, S.H.; Im, C.H. Riemannian classifier enhances the accuracy of machine-learning-based diagnosis of PTSD using resting EEG. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, **2020**, *102*(April), 109960.
<http://dx.doi.org/10.1016/j.pnpbp.2020.109960> PMID: 32376342
- [69] Wall, D.P.; Kosmicki, J.; Deluca, T.F.; Harstad, E.; Fusaro, V.A. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl. Psychiatry*, **2011**, *2012*, 2. PMID: 22832900
- [70] Thabtah, F.; Peebles, D. A new machine learning model based on induction of rules for autism detection. *Health Informatics J.*, **2020**, *26*(1), 264-286.
<http://dx.doi.org/10.1177/1460458218824711> PMID: 30693818
- [71] Mueller, A.; Candrian, G.; Kropotov, J.D.; Ponomarev, V.A.; Baschera, G.M. Classification of ADHD patients on the basis of independent ERP components using a machine learning system. *Nonlinear Biomed. Phys.*, **2010**, *4*(Suppl. 1), S1.
<http://dx.doi.org/10.1186/1753-4631-4-S1-S1> PMID: 20522259
- [72] Tenev, A.; Markovska-simoska, S.; Kocarev, L.; Pop-jordanov, J. Machine learning approach for classification of ADHD adults. *Int. J. Psychophysiol.*, **2014**, *93*(1), 162-166.
- [73] Gootenberg, J.S.; Abudayyeh, O.O.; Lee, J.W.; Essletzbichler, P.; Dy, A.J.; Joung, J. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science (80-)*, **2017**, *356*(6336), 438-442.
- [74] DNA sequencing costs: Data. National Human Genome Research Institute, 2020. Available from: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>
- [75] Charpentier, E.; Doudna, J.A. Rewriting a genome An accurate distance to the nearest galaxy. *Nat. Biotechnol.*, **2013**, *495*, 50-51. Available from: <https://www.nature.com/articles/495050a>
- [76] Lin, J.; Wong, K.C. Off-target predictions in CRISPR-Cas9 gene editing using deep learning. *Bioinformatics*, **2018**, *34*(17), i656-i663.
<http://dx.doi.org/10.1093/bioinformatics/bty554> PMID: 30423072
- [77] Casini, A.; Olivieri, M.; Petris, G.; Montagna, C.; Reginato, G.; Maule, G.; Lorenzin, F.; Prandi, D.; Romanel, A.; Demichelis, F.; Inga, A.; Cereseto, A. A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat. Biotechnol.*, **2018**, *36*(3), 265-271.
<http://dx.doi.org/10.1038/nbt.4066> PMID: 29431739
- [78] Lee, J.K.; Jeong, E.; Lee, J.; Jung, M.; Shin, E.; Kim, Y.H.; Lee, K.; Jung, I.; Kim, D.; Kim, S.; Kim, J.S. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.*, **2018**, *9*(1), 3048.
<http://dx.doi.org/10.1038/s41467-018-05477-x> PMID: 30082838
- [79] Vakulskas, C.A.; Dever, D.P.; Rettig, G.R.; Turk, R.; Jacobi, A.M.; Collingwood, M.A.; Bode, N.M.; McNeill, M.S.; Yan, S.; Camarena, J.; Lee, C.M.; Park, S.H.; Wiebking, V.; Bak, R.O.; Gomez-Ospina, N.; Pavel-Dinu, M.; Sun, W.; Bao, G.; Porteus, M.H.; Behlke, M.A. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.*, **2018**, *24*(8), 1216-1224.
<http://dx.doi.org/10.1038/s41591-018-0137-0> PMID: 30082871
- [80] Chen, J.S.; Dagdas, Y.S.; Kleinstiver, B.P.; Welch, M.M.; Sousa, A.A.; Harrington, L.B.; Sternberg, S.H.; Joung, J.K.; Yildiz, A.; Doudna, J.A. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*, **2017**, *550*(7676), 407-410.
<http://dx.doi.org/10.1038/nature24268> PMID: 28931002
- [81] Rahman, M.K.; Rahman, M.S. CRISPRpred: A flexible and efficient tool for sgRNAs on-target activity prediction in CRISPR/Cas9 systems. *PLoS One*, **2017**, *12*(8), e0181943.
<http://dx.doi.org/10.1371/journal.pone.0181943> PMID: 28767689
- [82] Chari, R.; Yeo, N.C.; Chavez, A.; Church, G.M. sgRNA scorer 2.0: A species-independent model to predict crispr/cas9 activity. *ACS Synth. Biol.*, **2017**, *6*(5), 902-904.
<http://dx.doi.org/10.1021/acssynbio.6b00343> PMID: 28146356
- [83] Kim, H.K.; Min, S.; Song, M.; Jung, S.; Choi, J.W.; Kim, Y.; Lee, S.; Yoon, S.; Kim, H.H. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat. Biotechnol.*, **2018**, *36*(3), 239-241.
<http://dx.doi.org/10.1038/nbt.4061> PMID: 29431740
- [84] Chuai, G.; Ma, H.; Yan, J.; Chen, M.; Hong, N.; Xue, D.; Zhou, C.; Zhu, C.; Chen, K.; Duan, B.; Gu, F.; Qu, S.; Huang, D.; Wei, J.; Liu, Q. DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol.*, **2018**, *19*(1), 80.
<http://dx.doi.org/10.1186/s13059-018-1459-4> PMID: 29945655
- [85] Wang, D.; Zhang, C.; Wang, B.; Li, B.; Wang, Q.; Liu, D.; Wang, H.; Zhou, Y.; Shi, L.; Lan, F.; Wang, Y. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat. Commun.*, **2019**, *10*(1), 4284.
<http://dx.doi.org/10.1038/s41467-019-12281-8> PMID: 31537810
- [86] O'Brien, A.R.; Wilson, L.O.W.; Burgio, G.; Bauer, D.C. Unlocking HDR-mediated nucleotide editing by identifying high-efficiency target sites using machine learning. *Sci. Rep.*, **2019**, *9*(1), 2788.
<http://dx.doi.org/10.1038/s41598-019-39142-0> PMID: 30808944
- [87] Pan, X.; Zuallaert, J.; Wang, X.; Shen, H-B.; Campos, E.P.; Marushchak, D.O.; De Neve, W. ToxDL: Deep learning using primary structure and domain embeddings for assessing protein toxicity. *Bioinformatics*, **2020**, *36*(21), 5159-5168.
<http://dx.doi.org/10.1093/bioinformatics/btaa656> PMID: 32692832
- [88] Cosgun, E.; Limdi, N.A.; Duarte, C.W. High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African Americans. *Bioinformatics*, **2011**, *27*(10), 1384-1389.
<http://dx.doi.org/10.1093/bioinformatics/btr159> PMID: 21450715
- [89] Sharabiani, A.; Bress, A.; Douzali, E.; Darabi, H. Revisiting war-

- farin dosing using machine learning techniques. *Comput. Math Methods Med.*, **2015**, *2015*, 560108. <http://dx.doi.org/10.1155/2015/560108>
- [90] Athreya, A.P.; Neavin, D.; Carrillo-Roa, T.; Skime, M.; Biernacka, J.; Frye, M.A.; Rush, A.J.; Wang, L.; Binder, E.B.; Iyer, R.K.; Weinsilboum, R.M.; Bobo, W.V. Pharmacogenomics-driven prediction of antidepressant treatment outcomes: A machine-learning approach with multi-trial replication. *Clin. Pharmacol. Ther.*, **2019**, *106*(4), 855-865. <http://dx.doi.org/10.1002/cpt.1482> PMID: 31012492
- [91] Chiu, Y.-C.; Chen, H.-I.H.; Gorthi, A.; Mostavi, M.; Zheng, S.; Huang, Y. Deep learning of pharmacogenomics resources: Moving towards precision oncology. *Brief. Bioinform.*, **2019**, *00*(July), 1-18. <http://dx.doi.org/10.1093/bib/bbz144> PMID: 31813953
- [92] Romero, A.; Carrier, P.L.; Erraqabi, A.; Sylvain, T.; Auvolat, A.; Dejoie, E. *Diet* network: thin parameters for fat genomic. *arXiv*, **2017**, v3, 1-11.
- [93] Jain, S.; Danesh, A.; Laroche, M.; Bakhtiari, S.; Maxwell, C.; Al-Khalili, Y. (335) Using pharmacogenomics as an effective tool in minimizing drug related side effects in patients with Chronic regional pain syndrome: A feasibility study. *J. Pain*, **2017**, *18*(4), S58-S59. <http://dx.doi.org/10.1016/j.jpain.2017.02.228>
- [94] Cseko, G.C.; Tremaine, W.J. The role of the institutional review board in the oversight of the ethical aspects of human studies research. *Nutr. Clin. Pract.*, **2013**, *28*(2), 177-181. <http://dx.doi.org/10.1177/08884533612474042> PMID: 23447409
- [95] Finlayson, S.G.; Bowers, J.D.; Ito, J.; Zittrain, J.L.; Beam, L.; Kohane, I.S. Emerging vulnerabilities demand new conversations. *Science* (80), **2019**, *363*(6433), 1287-1290. Available from: <http://science.sciencemag.org/content/363/6433/1287>
- [96] Meskó, B.; Hetényi, G.; Gyórfy, Z. Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Serv. Res.*, **2018**, *18*(1), 545. <http://dx.doi.org/10.1186/s12913-018-3359-4> PMID: 30001717
- [97] NYU School of Medicine. High school bioethics project - compassionate use. **2015**, *15* Available from: [https://med.nyu.edu/highschoolbioethics/sites/default/files/highschoolbioethics/Genetic Editing Module.pdf](https://med.nyu.edu/highschoolbioethics/sites/default/files/highschoolbioethics/Genetic%20Editing%20Module.pdf) https://med.nyu.edu/highschoolbioethics/sites/default/files/highschoolbioethics/CompassionateUse_Module.pdf
- [98] Harding, J.R. Beyond abortion: human genetics and the new eugenics. *Pepperdine Law Rev.*, **1991**, *18*(1), 471-512. PMID: 11659992
- [99] Ngiam, K.Y.; Khor, I.W. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.*, **2019**, *20*(5), e262-e273. [http://dx.doi.org/10.1016/S1470-2045\(19\)30149-4](http://dx.doi.org/10.1016/S1470-2045(19)30149-4) PMID: 31044724
- [100] Executive Office of the President. Maintaining american leadership in artificial intelligence: Executive order 13859. **2019**, 3967-3972. Available from: <https://www.govinfo.gov/content/pkg/FR-2019-02-14/pdf/2019-02544.pdf>
- [101] Levine, D.M.; Linder, J.A.; Landon, B.E. Characteristics of Americans with primary care and changes over time, 2002-2015. *Ann. Intern. Med.*, **2019**, *169*(1), 36-43.
- [102] Kane, L. Medscape Physician Compensation Report 2018. **2018**. Available from: <https://www.medscape.com/slideshow/2018-compensation-overview-6009667#32>
- [103] Kalinin, A.A.; Higgins, G.A.; Reamaroon, N.; Reza Soroushmehr, S.M.; Allyn-Feuer, A.; Dinov, I.D. Deep learning in pharmacogenomics: From gene regulation to patient stratification. *arXiv*, **2018**, *19*, 629-650.