

Use of Deep Learning in Personalized Medicine: Current Trends and the Future Perspective

Arooj Hussain, Sameena Naaz
{aroojpunjabi29@gmail.com, snaaz@jamiahamdard.ac.in}

Department of Computer Science and Engineering, School of Engineering Sciences and Technology,
Jamia Hamdard, New Delhi-110062, India

Abstract. Personalized Medicine is about to bring a paradigm shift in the way diseases are being treated across the world today. The first step to achieve Accurate Prognosis of medication however, is to attain Accurate Diagnosis of diseases. Due to this, the primary step in the direction of Personalized Treatment is to apply various Machine Learning and Deep Learning methods to predict the diseases and drug responses from various inputs such as Magnetic Resonance Images, CT Scans, PET Scans, etc. This paper aims to canvass the research studies that have been conducted in the previous 2-3 years to employ ML and DL techniques in predicting disorders as well as predicting responses to drugs from scans, images and other similar data. The disorders included are lung cancer, breast cancer, brain tumor, diabetes etc. One technique that has repeatedly been used by the researchers and which has replicated good results generally is the Convolutional Neural Network.

Keywords: Personalized Medicine, Deep Neural Network (DNN), Convolutional Neural Network (CNN), Linear Regression, Support Vector Machine, Variational AutoEncoder.

1. Introduction

With unparalleled advancements taking place in the field of Machine Learning, it has begun to play a strategic role in various subdomains of medicine such as Personalized Medicine (PM), which refers to providing individualized treatment plans to patients that may seemingly be suffering from the same disease. It paves the path for a tailor-made treatment solution for each patient and does not follow the “one-size fits all” approach. Even though the concept of providing differential treatment has been in place since the time of Hippocrates, but it is only recently that the scientific basis for the same is being studied and validated. Personalized Medicine has garnered lots of applications in the treatment of complex diseases dependent on genetics, environment and medical history.(1) Some of those diseases are diabetes (2), cancer (3), heart disease (4) and psychiatric disease (5). But a major stepping stone to using Personalized Medicine clinically is to transform the large, multimodal datasets into viable use cases and attaining high quality results and greater accuracy on pre-collected data, before the methods are being employed on patients. Accurate diagnosis is hence a prerequisite

of PM, only then Accurate Prognosis can be achieved. To achieve the goal of accurate predictions, various Machine Learning as well as Deep Learning methods have been applied on varied datasets by researchers. The techniques that have generally replicated good results belong to the domain of Deep Learning, such as Convolutional Neural Networks.

Comparing such previous work, a comprehensive study on a number of Machine Learning and Deep Learning models has been done in this survey, thereby comparing their performance and figuring out which diseases are mostly predictable. In this survey, several ML and DL models viz. Deep Neural Network (DNN), Support Vector Machine (SVM), Convolutional Neural Network (CNN), Adaboost, Linear Regression, Variational AutoEncoder (VAE), Generative Adversarial Network (GAN) have been compared. At last, the outcome of the previously conducted studies has been analyzed which hopefully would help in future advancement and research.

This remaining paper has been arranged as follows: In Section 2, the description of problem statement and the aim of the study is given. Section 3 discusses the general techniques that have been used by researchers in the selected papers. Section 4 summarizes the results that are obtained from each study. Section 5 describes the cumulative results drawn from all the studies in a tabulated and graphical form, and finally in Section 6, the conclusion obtained from the entire exercise is provided.

2. Problem Statement

This survey has been carried out to perform a review of Use Cases of Machine Learning / Deep Learning techniques in the domain of Personalized Medicine. The aim of this survey is to compare the techniques used across various pathologies and see how accurate they are in predicting the disorders or conditions related to the disorders. The long term aim of this survey is to develop an understanding of how the effective usage of ML or DL techniques can be beneficial to attain the much hyped and much needed goal of Personalized Medicine, that is, to offer treatments variable according to the patient profile and not simply prescribe one medicine/treatment to all the patients suffering from a particular disease. Although the scope of study for such a domain is infinite, this paper has been wrung out from the careful analysis of research studies conducted on more general diseases and where measurable results were obtained. Various techniques that the reader would come across while reading this paper include Convolution Neural Network, Generative Adversarial Network, Deep Neural Network, Variational AutoEncoders, SVM, Adaboost, Linear Regression, etc.

To keep the study as relevant as possible, only those papers were selected for the survey that were published in the time frame of 2017-2019.

3. Machine Learning Techniques

Some of the techniques that find place in this survey are discussed briefly below:

- 3.1 Convolutional Neural Network** – A very popular Neural Network employed in image recognition tasks is the Convolutional Neural Network. The architecture usually consists of a number of convolutional layers and subsampling layers. A convolutional layer consists of a set of kernels or filters that compute the dot product of the value of the filter and the input at the particular location. The subsampling layers are used for size reduction of the computed feature maps. These feature maps are finally concatenated to fully connected layers which give rise to the output. There are numerous architectures of CNNs available in the AI world today. Some that have garnered maximum attention are: GoogleNet, AlexNet, UNet, Inception, ResNet, etc. Fig. 1 below shows a basic representation of a CNN.

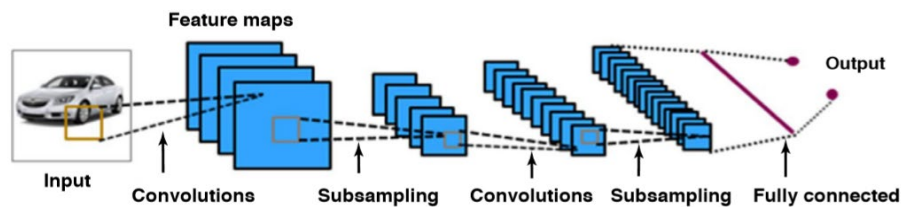


Fig. 1. Representation of Convolutional Neural Network(6)

- 3.2 Generative Adversarial Network** – A Generative Adversarial Network (GAN) is a combination of two Neural Networks, one convolutional and the other deconvolutional. The idea is to generate images using the convolutional network or generator and using the deconvolutional or discriminator network to determine whether the images have been generated artificially or are real images from the real world. This process continues as a feedback loop. The generator keeps on producing newer, higher quality images whereas the discriminator tends to become more able to segregate artificially generated images.
- 3.3 Support Vector Machine** – The Support Vector Machine (SVM) proposed by Platt et.al. (7) is a supervised learning method used for classification as well as regression problems. In a classification task, it creates two or more classes from the data values by taking into account specific parameters. The dataset may first require mapping to a higher dimension before separation into classes. It is then followed by finding the best suited hyperplane for separating the data values on the basis of distance between nearest points(8), which is the one that has maximum distance from the closest data point or maximum margin.

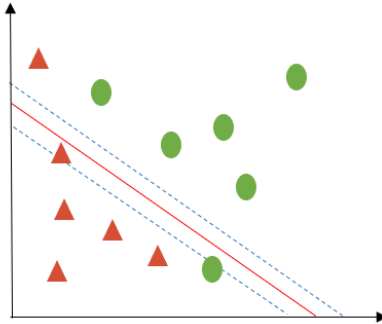


Fig. 2. Basic representation of SVM

3.4 Random Forest - Being an ensemble method of machine learning, Random Forest constructs a number of decision trees during the training phase. The training set is repeatedly broken down by splitting up the trees till a final result is obtained. Each step chooses the most favoured node on the basis of a random subset of characteristics. Fig. 3 tries to represent the basic splitting mechanism followed by a Random Forest.

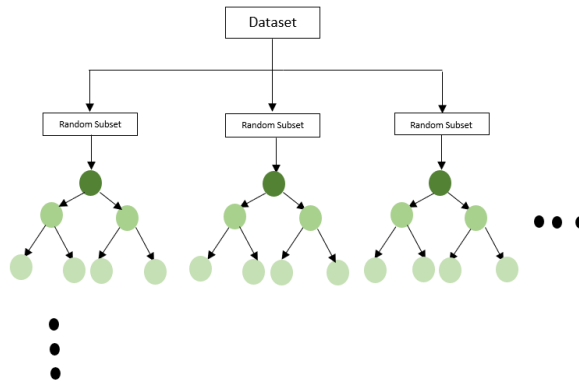


Fig. 3. Representation of Random Forest

3.5 Adaboost – or Adaptive Boosting algorithm is an ensemble classifier method that works iteratively. The basic idea behind this is first setting up the weights of classifiers, then training the data sample for every iteration in a way that accurate predictions of unusual observations are obtained. The Adaboost algorithm can be formulated by taking any basic Machine Learning algorithm as a base classifier.

3.6 Deep Neural Network (DNN) – An Artificial Neural Network (ANN) was built to be analogous to the neural networks found in the human brain, its most basic component too being called a Neuron. A neuron is a powerful little processing unit that converts an input into output on the basis of its weight and an activation function. An individual neuron however is not of much use, it is when they combine to form multiple layers, its power can be harnessed to perform certain tasks. A minimum of three layers are required to form the most fundamental Artificial Neural Network, and more than one hidden layer in between the input and output layer is required for an ANN to be called a *Deep Neural Network (DNN)* as shown in Fig. 4.

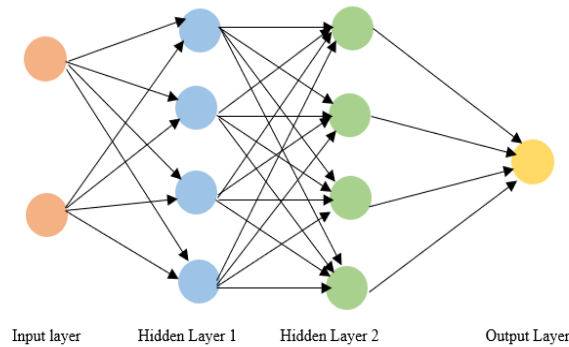


Fig. 4. Basic structure of DNN

3.7 Linear Regression – Linear Regression is one of the most basic and most used algorithms employed for prediction tasks. It works by analysing the relationship between the dependent and independent variables. If the dependent variable increases/decreases accordingly as the independent variable is increasing/ decreasing, the relation is said to be linear. Linear Regression tends to analyse the output variable after it is trained to understand and analyse numerous such cases.

3.8 Variational AutoEncoder (VAE) – A simple AutoEncoder is an unsupervised learning model made up of two parts: encoder and decoder. The encoder maps the input values into relevant patterns known as latent representations whereas the decoder reconstructs the inputs from these generated latent representations. A Variational AutoEncoder (VAE) is an AutoEncoder that gets regularised training in order to avoid overfitting and to ensure that the latent representations so generated are less prone to error and give better results.

4. Findings/ Literature Review

Bejnordi et. al. (9) conducted a competition to devise algorithms to diagnose breast cancer from whole slide images verified by immunohistochemical staining for identification and classification

of Nodal Metastases. They used percentage split to segregate the dataset into training and testing sets, consisting of 270 and 129 images respectively. The techniques that were employed by the participants in the study included some basic ML ones like Support Vector Machine, Random Forest, Adaboost to a number of Deep Learning based Convolution Neural Networks such as GoogleNet, SegNet, AlexNet, UNet, etc. To compare the results with human efficacy, 11 reputed pathologists were consulted to diagnose the said disorder from the test set of images within a time constraint and 1 expert without time constraint. AUC was used to determine the efficiency of the Metastases Classification task. The best performing algorithm (an ensemble of GoogleNet) had an AUC of 0.994 whereas the AUC of the mean performance of the 11 pathologists was measured to be 0.810.

Rajpurkar et. al. (10) developed a CNN named ChexNext to detect the presence of 14 different pathologies viz. Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pleural Thickening, Pneumonia and Pneumothorax concurrently from chest radiographs. The dataset used was ChestX-ray 8 which is an extensively sourced dataset of 112,120 frontal-view chest radiographs of 30,805 unique patients, which was divided into three parts for training, testing and validation. The validation set consisted of 420 radiographs to evaluate ChexNext as well as the radiologists with whom the performance of the model was to be compared. Metrics such as AUC, Sensitivity, Specificity, etc. were compared to conclude that ChexNext performed equally well on 10 pathologies, better on 1 (Atelectasis), and worse on 3 (Cardiomegaly, Emphysema, Hiatal Hernia).

Ding et. al. (11) trained a Convolutional Neural Network of Inception V3 Architecture on 90% of 2019 18F-FDG PET brain images from Alzheimer's Disease Neuroimaging Initiative (ADNI) and tested it on the remaining 10% as well as on a retrospective independent test set of 40 images separately collected. The idea behind the study was to predict the final diagnosis of Alzheimer's Disease, mild cognitive impairment, or neither and to compare the results with those of the radiologic readers who generally read and conclude on such reports. The analysis of the model was done by checking various metrics such as AUC, specificity, sensitivity, saliency map etc. The AUC achieved for the model was 0.98 (95% Confidence Interval), 82% specificity at 100% sensitivity at an average 75.8 months before actual diagnosis of the disease whereas the radiological readers got a specificity of 91% and sensitivity of 57%.

Causey et. al. (12) again devised a CNN named NoduleX which was trained to predict lung nodule malignancy in patients from CT scans. The dataset they chose consisted of 1065 images of lung nodules from Lung Image Database Consortium (LIDC/IRDI cohort). Validation was performed on 20% of the dataset that was well balanced to contain both "likely malignant" and "likely benign" nodules. The AUC of the model for nodule malignancy classification was found out to be 0.99.

Madani et. al. (13) developed a twin fold approach to predict cardiac disease from Medical Imaging Assessment, using pipeline supervised models followed by semi-supervised GAN. Both the models were utilized to conduct two tasks – 15 view still-image echocardiographic (ECG) view classification and binary left ventricular hypertrophy (LVH) classification. The dataset used for the first task consisted of 267 images (240 training and 27 testing) whereas for second task consisted of

465 images (405 training and 50 testing). The supervised pipeline model employed an ensemble of CNNs for segmentation and classification. The results obtained were as follows: the pipeline models obtained an accuracy of 94.4% ECG view classification and 91.2% for LVH classification and the semi-supervised GAN achieved 80% accuracy for the former and 92.3% for the latter.

Li et al. (14) developed a Deep Learning based Radiomics approach to detect mutation status in isocitrate dehydrogenase-1 (IDH1) gene in patients suffering from Low-Grade Glioma using information from Magnetic Resonance Images. The dataset consisted of MRI images from 151 patients fed to a CNN of 6 convolutional layers and a fully connected layer of 4096 neurons. The accuracy achieved by the model was 92% as compared to the traditional radiomics method which achieved only 86%. The accuracy of DLR was further improved to 95% when multiple modality MR Images were used.

Bibault et al. (15) created three systems to check for response of patients suffering from T2-4 N0-1 rectal adenocarcinoma after neo-adjuvant chemoradiation. They employed a Deep Neural Network (DNN), a Linear Regression Model and an SVM to compare the results on 95 patients; 49 males and 46 females. Out of 95, 22 patients showed complete response to the chemoradiation which was predicted by the DNN with an accuracy of 80%, the Linear Regression model with 69.5% and SVM with 71.58%.

Miotto et al. (16) presented an unsupervised deep learning method “Deep Patient” for general purpose patient representation from electronic health records (EHR) data. The main motive of the study was to devise a system that could predict the future condition of patients. The dataset they chose for the study was a 76,214-patient strong dataset consisting of people suffering from 78 different diseases. The method performed exceptionally well for diabetes, certain cancers and schizophrenia, etc. The best AUC was achieved for Diabetes Mellitus with complications at 0.907.

Rampasek et al. (17) developed two Variational Autoencoder based generative models for better prediction of drug response prediction. The two models viz. Perturbation Variational Autoencoder (PertVAE) and its semi-supervised counterpart, Drug Response Variational Autoencoder (Dr.VAE), are trained to figure out the changes in gene states before and after the application of drug treatment. The work was performed on two datasets, one each for drug response in cancer cell lines and drug-induced perturbations, in which responses to 19 drugs were recorded. The results were 3 to 11% better in terms of AUC as compared to previously published models (VAE and Semi-Supervised VAE (SSVAE)).

Akkus et al. (18) developed a multi scale Convolution Neural Network architecture to detect the link between the deletion of chromosome arms 1p/19q and positive response to treatment and longer survival in patients suffering from low grade glioma (LGG). The dataset consisted of 159 LGGs with 3 image slices each of which was made to go through a three-step process of image registration, tumor segmentation and classification of chromosome status. The accuracy achieved by the model was 87.7% with a specificity of 82.22% and sensitivity of 93.3%.

5. Results and Discussion

In this comparative study of Use Cases of Machine Learning methods in Personalized Medicine, differing values for accuracy, sensitivity, Area under the ROC Curve (AUC) has been achieved. To improve the ease of readability, two tabular as well as graphical representations of the results have been provided in this section. Table 1 below accumulates the inferences from all the studied papers and compares the AUC obtained while carrying out the studies whereas Table 2 compares the obtained Accuracy as chosen by several researchers to validate their models.

Table 1. Comparison of AUCs

Paper	Dataset	Pathological Condition	Model/s Used	AUC
Bejnordi et.al. (2017)	499 images of Nodal Metastases	Breast Cancer	SVM Random Forest Adaboost GoogleNet SegNet AlexNet UNet Many others	Highest 0.994 for GoogleNet
Rajpurkar et. al. (2018)	112,120 frontal-view chest radiographs	14 chest diseases	CNN named ChexNext	Highest 0.862 for Atelectasis
Ding et. al. (2019)	8F-FDG PET brain images from Alzheimer's Disease Neuroimaging Initiative (ADNI), 40 separately collected images	Alzheimer's Disease, mild cognitive impairment	CNN of Inception V3 Architecture	0.98
Causey et. al. (2018)	1065 images of lung nodules from Lung Image Database Consortium	Lung Cancer	CNN named NoduleX	0.99
Miotto et. al. (2016)	EHR of 76,214 patients	78 different diseases	Deep Patient – an unsupervised deep learning method	Highest 0.907 for Diabetes Mellitus
Rampasek et. al. (2017)	Two varied datasets for cancer cell lines and drug-induced perturbations	Gene status in cancer treatment	PertVAE Dr.VAE	3 to 11% > VAE and SSSVAE

Table 2. Comparison of Accuracies

Paper	Dataset	Pathological Condition	Model/s Used	Accuracy
Madani et. al. (2018)	267 and 465 ECG images	Cardiac Disease	Pipeline supervised models of CNN <ul style="list-style-type: none"> • ECG View • LVH Supervised GAN <ul style="list-style-type: none"> • ECG View • LVH 	94.4% 91.2% 80% 92.3%
Li et. al. (2017)	MRI images from 151 patients	Low-Grade Glioma	CNN based Deep Learning Radiomics	92%
Bibault et. al. (2018)	95 patient responses	T2-4N0-1 Rectal Adenocarcinoma	DNN Linear Regression SVM	80% 69.5% 71.58%
Akkus et. al. (2017)	159 LGGs with 3 image slices each	Low-Grade Glioma	CNN	87.7%

For further comparison and ease of readability, the data from Table 1 and Table 2 has been represented in the form of charts in Fig. 5 and Fig. 6

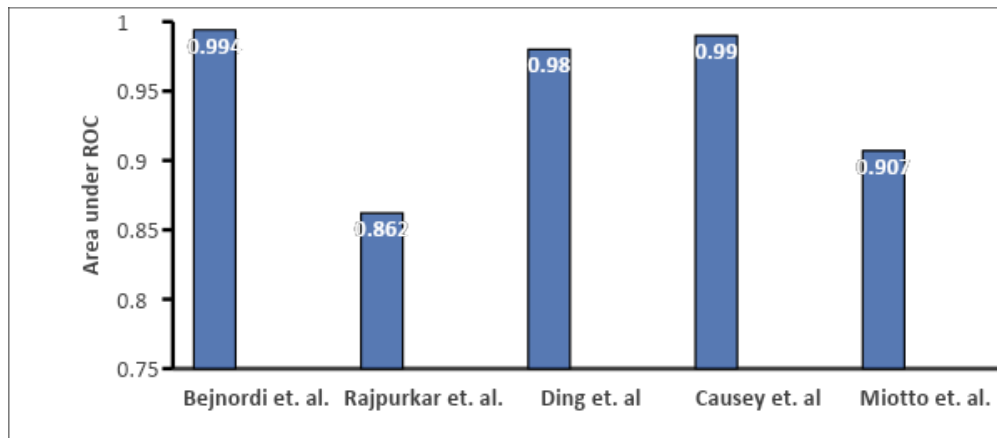


Fig. 5. Comparison of best achieved AUC

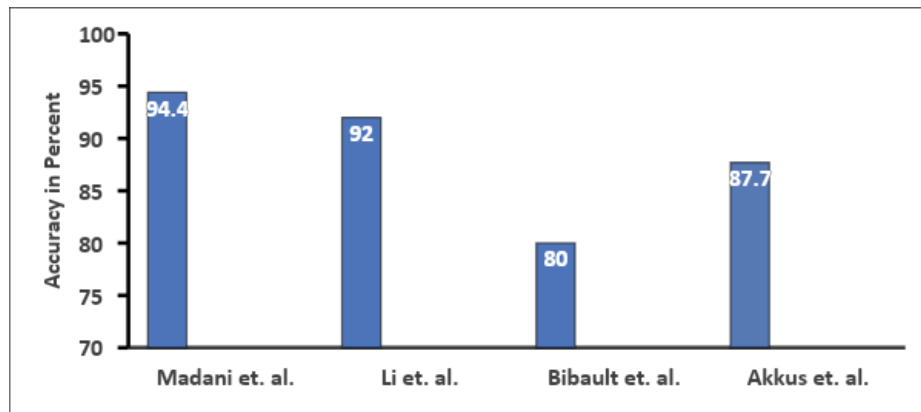


Fig 6. Comparison of best achieved Accuracy

6. Conclusion

Personalized Medicine is the way forward in the field of Medicine, which has the power to revolutionize the entire way this industry works. From predicting the likelihood of the occurrence of a disease to the prediction of response of a particular treatment, Personalized Medicine encompasses it all. This survey tried to summarize the various Use Cases where such prediction was performed during the recent years using well popularized techniques from Machine Learning and Deep Learning. The disorders which were studied included some cancers such as of the lung, breast, adenocarcinoma, and other diseases like Low-Grade Glioma, Diabetes etc. Most researchers have preferred to use deep learning based Convolutional Neural Networks (CNN) over the other discussed techniques, which in the best case has achieved an AUC of 0.99.

References

1. Zhang S, Bamakan SMH, Qu Q, Li S. Learning for Personalized Medicine: A Comprehensive Review From a Deep Learning Perspective. *IEEE Rev Biomed Eng.* 2018;12:194–208.
2. Wei Z, Wang K, Qu H-Q, Zhang H, Bradfield J, Kim C, et al. From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. *PLoS Genet.* 2009;5(10):e1000678.
3. Kim J, Shin H. Breast cancer survivability prediction using labeled, unlabeled, and pseudo-labeled patient data. *J Am Med Informatics Assoc.* 2013;20(4):613–8.
4. Antman EM, Loscalzo J. Precision medicine in cardiology. *Nat Rev Cardiol.* 2016;13(10):591.
5. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2018;3(3):223–30.

6. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discov Today*. 2018;23(6):1241–50.
7. Platt J. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. 1998;(MSR-TR-98-14).
8. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. *Neural Process Lett*. 1999;9(3):293–300.
9. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA - J Am Med Assoc*. 2017;318(22):2199–210.
10. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med*. 2018;15(11):e1002686.
11. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, et al. A Deep learning model to predict a diagnosis of alzheimer disease by using 18F-FDG PET of the brain. *Radiology*. 2018;290(2):456–64.
12. Causey JL, Zhang J, Ma S, Jiang B, Qualls JA, Politte DG, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans. *Sci Rep*. 2018;8(1):9286.
13. Madani A, Ong JR, Tibrewal A, Mofrad MRK. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *npj Digit Med*. 2018;1(1):59.
14. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep*. 2017;7(1):5467.
15. Bibault J-E, Giraud P, Durdux C, Taieb J, Berger A, Coriat R, et al. Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep*. 2018;8(1):12611.
16. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep*. 2016;6:26094.
17. Rampasek L, Hidru D, Smirnov P, Haibe-Kains B, Goldenberg A. Dr. vae: Drug response variational autoencoder. *arXiv Prepr arXiv170608203*. 2017;
18. Akkus Z, Ali I, Sedlár J, Agrawal JP, Parney IF, Giannini C, et al. Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. *J Digit Imaging*. 2017;30(4):469–76.