Journal of Biomedical Informatics 113 (2021) 103627

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues

Shahab Shamshirband [a,j,*], Mahdis Fathi [b], Abdollah Dehzangi [c,d], Anthony Theodore Chronopoulos [e,f], Hamid Alinejad-Rokny [g,h,i]

[a] Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
[b] Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
[c] Department of Computer Science, Rutgers University, Camden, NJ 08102, USA
[d] Center for Computational and Integrative Biology, Rutgers University, Camden, NJ 08102, USA
[e] Department of Computer Science, University of Texas at San Antonio, San Antonio, TX 78249, USA
[f] (Visiting Faculty) Department of Computer Science, University of Patras, 26500 Rio, Greece
[g] Systems Biology and Health Data Analytics Lab, The Graduate School of Biomedical Engineering, UNSW Sydney, 2052 Sydney, Australia
[h] School of Computer Science and Engineering, The University of New South Wales (UNSW Sydney), 2052 Sydney, Australia
[i] Health Data Analytics Program Leader, AI-enabled Processes (AIP) Research Centre, Macquarie University, Sydney 2109, Australia
[j] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, ROC

ARTICLE INFO

Keywords:
Machine learning
Deep neural network
Healthcare applications
Diagnostics tools
Health data analytics

ABSTRACT

In the last few years, the application of Machine Learning approaches like Deep Neural Network (DNN) models have become more attractive in the healthcare system given the rising complexity of the healthcare data. Machine Learning (ML) algorithms provide efficient and effective data analysis models to uncover hidden patterns and other meaningful information from the considerable amount of health data that conventional analytics are not able to discover in a reasonable time. In particular, Deep Learning (DL) techniques have been shown as promising methods in pattern recognition in the healthcare systems. Motivated by this consideration, the contribution of this paper is to investigate the deep learning approaches applied to healthcare systems by reviewing the cutting-edge network architectures, applications, and industrial trends. The goal is first to provide extensive insight into the application of deep learning models in healthcare solutions to bridge deep learning techniques and human healthcare interpretability. And then, to present the existing open challenges and future directions.

1. Introduction

Analysis of healthcare data is coming to play a very important role in personalized medicine. For example, personalized treatment in cancer is trying to provide the right treatment to the right patient by taking into account several types of patient's data, including genomics variants, environment, imaging genomics, current drugs, and lifestyle. In the last decade, modern technologies such as genomics, imaging, and lifetime monitoring technologies, produced huge and complex amount of health data that allow researchers to provide better treatments for patients. Despite this huge amount of data, our understanding of diseases, and how we can treat the patients is still insufficient. To analyze such data,

Abbreviations: PPMI, Parkinson's Progression Markers Initiative; SNUH, Seoul National University Hospital; FP-CIT, I-fluoropropyl carbomethoxy iodophenyl nortropane; SPECT, Single-Photon Emission Computed Tomography; PD, Parkinson Disease; SWEDD, Scans Without Evidence of Dopaminergic Deficit; AUC, Area Under the ROC Curve; DDSM, Digital Database for Screening Mammography; SDN, Stack of Denoising Autoencoder; SCDL, Sparsely Connected Deep Learning; FROC, Free-response ROC curve; DAE, Denoising Autoencoder; DNN, Deep Neural Network; SFM, Screen-Film Mammograms; DM, Digital Mammograms; UM, University of Michigan Health System; USF, University of South Florida; LIDC/IDRI, Lung Image Database Consortium and Image Database Resource Initiative; LUAD, Lung Adenocarcinoma; STAD, Stomach Adenocarcinoma; BRCA, Breast Invasive Carcinoma; PGBM, Point-wise Gated Boltzmann Machine; BRATS, Brain Tumor Image Segmentation; CAD, Computer Aided Diagnosis; DoS, Denial of Service.

* Corresponding author.
E-mail addresses: shahab.shamshirband@ntnu.no, shamshirbands@yuntech.edu.tw (S. Shamshirband).

https://doi.org/10.1016/j.jbi.2020.103627
Received 21 March 2020; Received in revised form 29 August 2020; Accepted 14 November 2020
Available online 28 November 2020

the application of Machine Learning (ML) and data mining techniques [1], including Deep Neural Networks (DNN) have become more attractive given the complexity of the data. In particular, identifying the associations between all the different types of patient's data is a fundamental problem to develop trustable diagnostic tools based on data-driven techniques and machine learning models.

During the past decade, a wide range of Artificial Intelligence (AI) and ML techniques have been used to analyze the massive data in healthcare, effectively. For example, A logistic regression-based prediction model was applied in heart disease detection for automated early detection of heart disease [2]. ML was also applied in medical imaging to provide automatic discovery of object features [3]. Among different ML models, DNN based techniques are attracting lots of attention, in particular in the analysis of big datasets. Deep Learning (DL) techniques are multiple steps feature learning techniques, in which data is filtered through a cascade of multiple layers. DNN models become more and more accurate when they process large scale data, which enables them to outperform many classical machine learning models. DNN based approaches have also demonstrated a great performance in image processing and natural language processing [4–7].

Considering the performance of DL (also widely referred as DNN) approaches in different areas and their rapid continuous methodological improvement, these models are becoming the new and exciting tools to analyze healthcare data. A wide range of initiatives have been conducted using DL models on biomedical and healthcare data. For example, Google's DeepMind [8] and IBM's Watson [9] have developed a computer-based support system to analyze healthcare data [10]. Deep Learning has been successfully used to encode a deformable model's parameters that can facilitate the left ventricle's (LV) segmentation from short-axis cardiac MRI [11]. Another deep learning model based on Restricted Boltzmann Machines (RBM) was applied in medical imaging to identify biomarkers from MRI scans [12].

In this review paper, we categorized DL methods into two groups namely, single DL (S-DL) and hybrid DL (H-DL) techniques that are used between 2015 and 2019. The single DL (S-DL) refers to those methods that solely used deep learning architecture to build their model. On the other hand, H-DL refers to those methods that use DL in conjunction with other traditional machine learning models (non-DL models). In this way, we investigate the strength and weaknesses of DL techniques compared to those models that uses traditional machine learning combined with DL architecture.

Here, the performance of each technique is discussed in more detail and the training time of existing DL- health care systems (DL-HCS) is also discussed. Recently, Deep EHR [13] surveyed specific deep learning techniques that are employed on electronic health records (EHR). They discussed the clinical applications used by specific DL models and identified several limitations of current DLs such as model interpretability, data heterogeneity, and lack of universal benchmarks. Finally, Deep EHR summarized the state-of-the-art models and identified avenues for the future direction.

Unlike Deep EHR which reviews specific deep learning techniques on electronic health records, our review paper focuses on hybrid deep learning techniques tailored to early disease detection. Deep EHR identifies several limitations of DLs being used in EHR while our survey highlights some limitations of single and hybrid DLs in Healthcare. Like Deep EHR which focused on evaluation metrics of DLs such as AUC, precision, recall, and F1 score, our survey uses the same standard classification metrics to highlight AUC, accuracy, sensitivity of S-DL and H-DL. Unlike Deep EHR, this review paper investigates the training time of existing DL-HCS. While some studies share similar investigation and evaluation metrics, results are not directly comparable due to proprietary nature of datasets.

Lan *et al.* [14] proposed a survey paper that investigates data mining and deep learning techniques for specific domain knowledge of bioinformatics. Preprocessing, classification, clustering, and optimized neural network architectures are summarized in deep learning methods and the advantages and disadvantages of methods are discussed. Different from this work, our survey paper tackles more specific deep learning studies which are combined with other conventional (also called traditional) machine learning techniques. We focus on the advantages and disadvantages of S-DL and H-DL techniques. Our survey paper also focuses on specific deep learning techniques used to identify disease detection. Also, the performance evaluation discussed in our survey paper is not covered in [14]. Altogether this paper focuses on the state-of-the-art studies that employed deep learning methods for disease detection and analysis of big data in the field of healthcare.

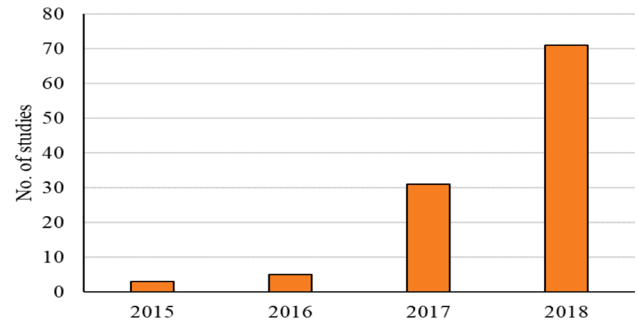


Fig. 1. The yearly distribution on DL techniques in HCS. The number of studies on DL applications in HCS has been significantly increased since 2015.

In this review, four deep learning techniques are selected by focusing on health care systems (HCS) in the period between 2015 and 2019. These four architectures are namely, Convolutional Neural Network (CNN) [15], Deep Belief Networks (DBN) [16], Auto-Encoder (AE) [17] and Recurrent Neural Network (RNN) [18]. These architectures are widely used for disease detection applications. Fig. 1 indicates the yearly distribution of DL articles in HCS. Fig. 2 (a, b) shows the distribution of DL techniques used for disease detection. Fig. 2.a indicates the distribution of different types of deep learning methods such as CNN, DBN, AE and RNN applied in the healthcare systems.

The main contributions of this paper can be summarized as follows:

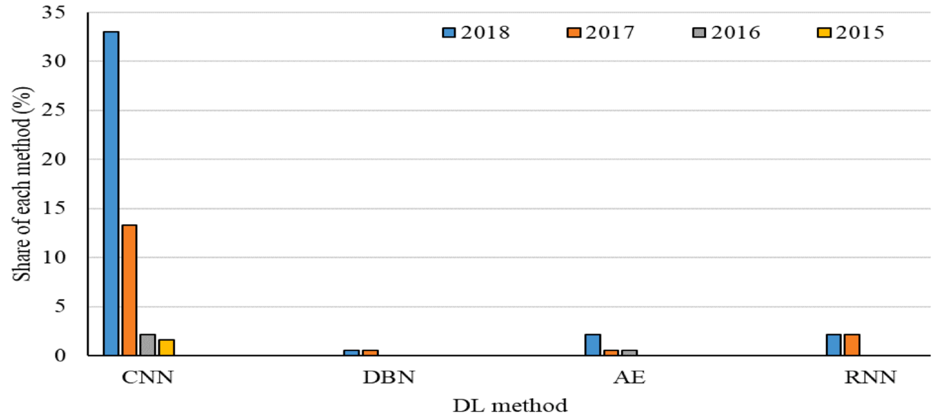
- Classification of the existing deep learning approaches in healthcare.
- Providing extensive insights into the accuracy and applicability of deep learning models in healthcare solutions.
- Discussing the core technologies which can reshape deep learning approaches in healthcare technologies.
- Presenting open issues and challenges in current deep learning models in healthcare.

The current study is organized as follows. In Section 2, a research methodology is explained. In Section 3, the definition and architecture of deep learning methods are presented. Section 4 presents deep learning methods with the best performances for disease detection. Open issues and conclusions of this paper are provided in Sections 5 and 6, respectively.

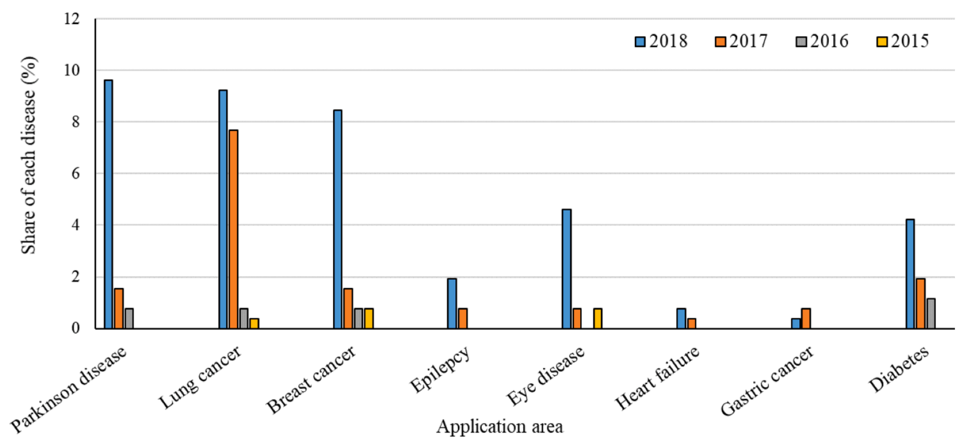
2. Deep learning: Definition

2.1. Algorithms and architecture

A Deep Learning architecture can be simply defined as an Artificial Neural Network (ANN) with two or more hidden layers aiming at enhancing the prediction accuracy [19]. Unlike traditional artificial neural networks, DL uses many more hidden layers. In a conventional Deep Neural Network (DNN), a weighted and bias-corrected input value is passed through a non-linear activation function such as ReLu and softmax function to derive an output [20]. As such, the objective of training a DNN is to optimize the weights of the network so that the loss function is minimized [21]. Each lower level features are utilized for the



(a)



(b)

Fig. 2. The distribution of DL techniques used for disease detection. a) Indicates the types of deep learning methods used in diseases detection. b) shows the most important diseases as an application of DL methods.

definition of higher-level ones. DL has been utilized in different applications such as speech recognition [22,23], image analysis [24], text mining [25], health monitoring [3,26], drug discovery [27], computer vision [28], object detection [29], and many different applications [30–33]. This paper attempts to investigate the state of the art of deep learning models used in HCS applications. Fig. 3 shows the taxonomy of common deep learning architectures for analyzing HCS data along with selected applications in HCS, especially in disease detection.

2.1.1. *Convolutional neural network*

Convolutional Neural Network (CNN) is a supervised deep learning architecture. It is mainly used for image analysis applications [34,35]. CNN includes three types of layers namely, convolutional layers, pooling layers, and fully connected layers. In the convolutional layer, the input image passes through kernels or filters to generate some feature maps. In the pooling layer, the size of each of the feature maps is reduced to keep the number of weights small. This process is also known as down-sampling or subsampling. There exist various kinds of pooling methods such as global pooling, max pooling, and average pooling. After these aforementioned layers, the fully connected layer is used to transform two-dimensional feature maps into a one-dimensional vector for final classification. This procedure of CNN for classifying images is illustrated in Fig. 4. Most common CNN architectures are ZFNet [36], GoogLeNet [37], VGGNet [38], AlexNet [15], ResNet [39].

Fig. 4: demonstrates a CNN architecture with two convolutional layers. Each convolutional layer followed by a pooling/subsampling layer. The output of the last pooling layer is fed to a fully connected layer and a final output layer [40]. A novel end-to-end DL called Deepr was proposed in [41] to extract important features from medical records and predict the abnormality. A convolutional neural net is applied to a sequence of discrete elements to predict unplanned readmission after discharge. Later on, Cheng *et al.* [42] employed a DL technique to explore the temporal characteristics of patient EHR. In the second layer of the proposed DL, the convolution operator performed on the time dimension of the patient EHR matrices. Early fusion, late fusion and slow fusion as a temporal fusion strategy are applied in the model to leverage the temporal smoothness of EHR into the learning process.

2.1.2. *Recurrent neural network*

Recurrent Neural Network (RNN) is utilized for pattern recognition for stream or sequential data such as speech, handwriting, and text. There is a cyclic connection in the structure of RNN. The recurrent computations are performed in these cyclic connections of hidden units to sequentially process the input data [26]. Each of the previous input data is kept in a state vector in hidden units and these state vectors are utilized to compute the outputs. Therefore, RNNs consider the current input and the previous input to compute the new output. Despite the promising performance of RNN, the vanishing gradient for data training is the main problem for this method. One solution to solve this problem is the use of Long Short Term Memory (LSTM) networks which can store sequences for a long time as well as using Gated Recurrent Units (GRUs) [43,44]. Fig. 5 shows the architecture of an RNN. Despite promising results using GRU to address the vanishing gradient problem, the effectiveness of this method is highly dependent on the input data and

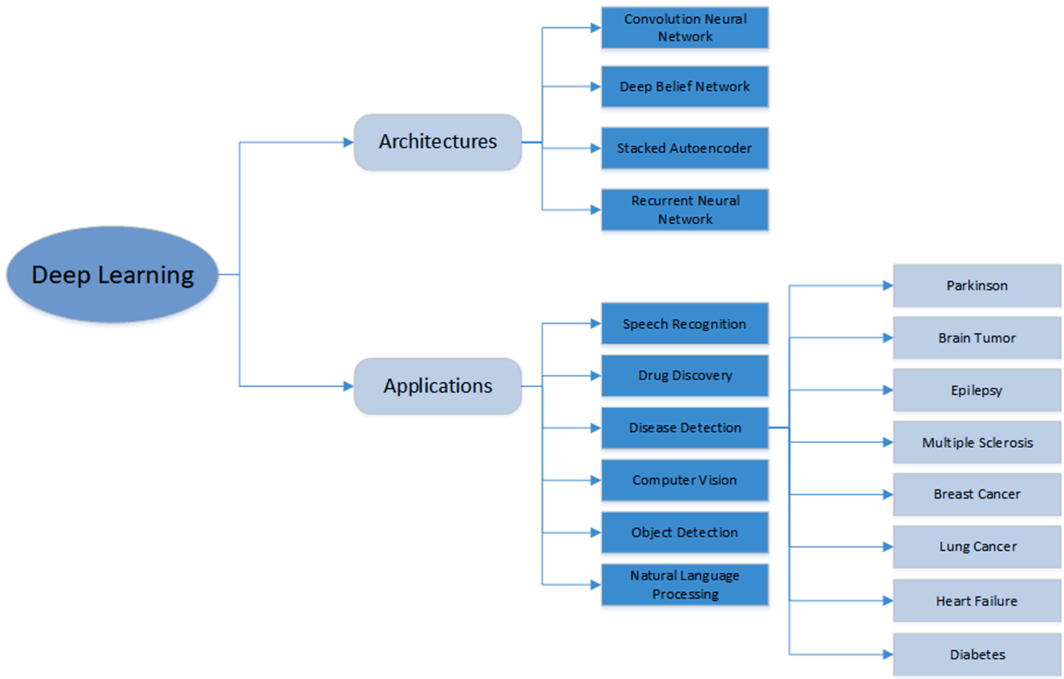


Fig. 3. Taxonomy of Deep Learning architectures applied in the health care system. The mentioned applications in this figure are among those that have been widely investigated using DL models.

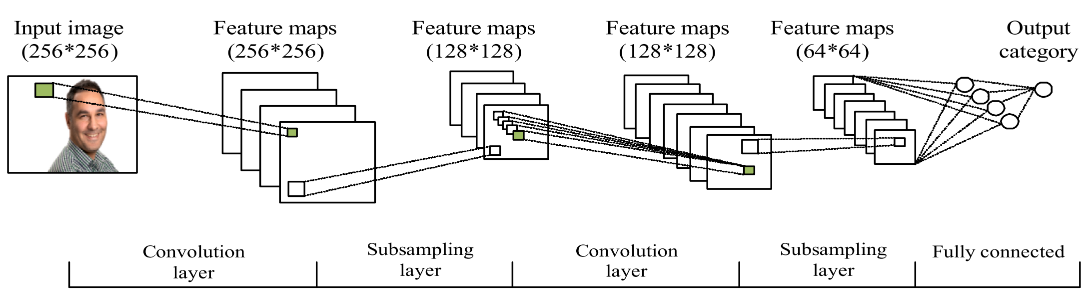


Fig. 4. CNN Architecture with two convolutional layers adapted from [40]. Each convolutional layer followed by a pooling/subsampling layer. The output of the last pooling layer id is fed to a fully connected layer and a final output layer.

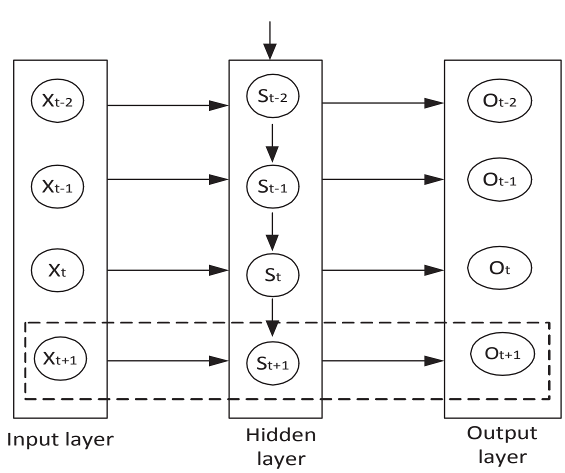


Fig. 5. Recurrent Neural Network (RNN) architecture as it is explained in [45].

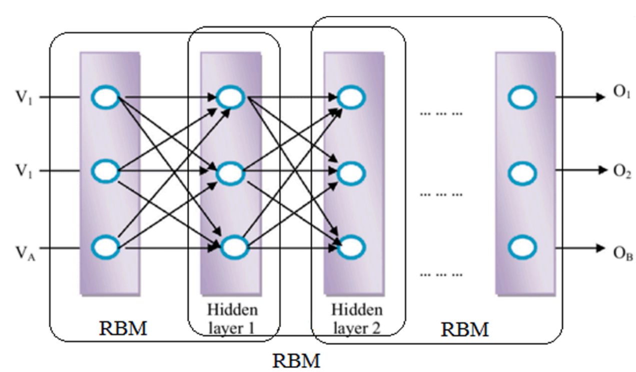


Fig. 6. Deep Belief Network (DBN) architecture as it is explained in [51].

complexity of the problem.

2.1.3. The deep belief networks

The Deep Belief Networks (DBNs) can learn high dimensional manifolds of the data. DBNs is a hybrid multi-layer neural network that contains both directed and undirected connections. The connection

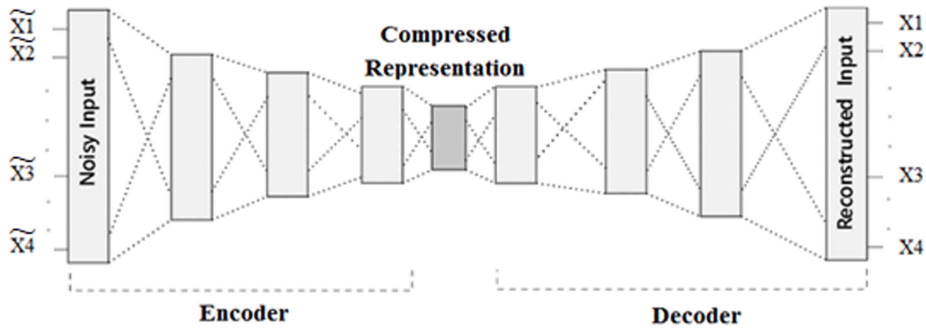


Fig. 7. Denoising Auto-Encoder (DAE) architecture as it is explained in [53].

between the top two layers is undirected and connections between all other layers are directed. DBNs can be considered as a stack of Restricted Boltzman Machines (RBMs) which are trained in a greedy manner. The layers of RBMs communicate with previous and subsequent layers [46–48]. This model consists of a feed-forward network and several layers of RBM as feature extractors [49]. A hidden layer and visible layer are only two layers of an RBM [50]. Fig. 6 presents the architecture of DBN methodology which has been adopted from [51], where (v) is the stochastic visible variable of the deep belief model.

2.1.4. Auto-Encoder

Auto-Encoder (AE) is an artificial neural network that aims at efficiently coding the data. To do this, it maps the input through an interconnected neural network to itself, thus it can be used for feature reduction or network initialization. AE is categorized as unsupervised learning and it includes SAE, VAE, and DAE. Denoising Auto-Encoder (DAE), which extended from AE, is a neural network mainly used to extract features from a noisy dataset. Typically, the DAE contains an input layer, encoding layer, and decoding layer, respectively [17]. DAE can be stacked to obtain high-level features. Stacked Denoising Auto-Encoder (SDAE) is another deep learning method that has been recently used for nonlinear dimensionality reduction. Fig. 7 presents the architecture of the DAE method as it is explained in [53].

3. Deep Learning: Algorithm and disease detection

In this section, we discuss specific types of DL algorithms applied in healthcare systems to characterize healthy versus unhealthy individuals. To do this, we discuss the models concerning a specific disease. Finally, we highlight the performance of DLs in distinguishing infected individuals from healthy individuals.

3.1. Disease categories

In this section, we investigate different methods based on disease categories. This section focuses on the most important DL methods with different performance and application in several categories of diseases. Below, we review nine categories of diseases namely, (1) EEG Imagery, (2) Multiple Sclerosis (MS), (3) Breast Cancer, (4) Brain Cancer, (5) Hybrid detections of Lung Adenocarcinoma, Stomach Adenocarcinoma, and Breast Invasive Carcinoma, (6) Epilepsy Diagnosis, (7) Heart Disease, (8) Parkinson's Disease and (9) Eye Disease. These nine diseases that are investigated in this study are selected using two main criteria. First, those that have been widely investigated using DL models. Second, those that either solved or obtained promising results using DL models. Below, DL methods used to tackle these diseases are discussed in detail.

3.1.1. EEG imagery signals (Category 1)

The focus of this section is to show the significance of DL methods in terms of classification performance. Authors in [48] used deep belief nets to classify EEG waveforms for brain signals. The results in DBN demonstrate that the prediction task for this model takes less time compared to SVM and KNN classifiers even on raw data. Later on [52] proposed a new approach based on deep learning to classify ECG signals. In this model, after the feature learning phase, a regression layer on the top of the resulting hidden representation layer was added which creates a deep neural network. Compared to other methods the accuracy of this new method was enhanced with less expert interactions and the online retraining phase was also faster than others.

More recently, an ensemble of CNN and Stacked Auto-Encoders (SAE) methods has been employed to classify EEG signals in [54]. CNN extracts the features in the first step and then it classified through SAE. The patterns such as EEG channel and power values learned in the convolution layer and then max pooling is adopted to make CNN invariant to the time location of activation patterns. Finally, SAE is

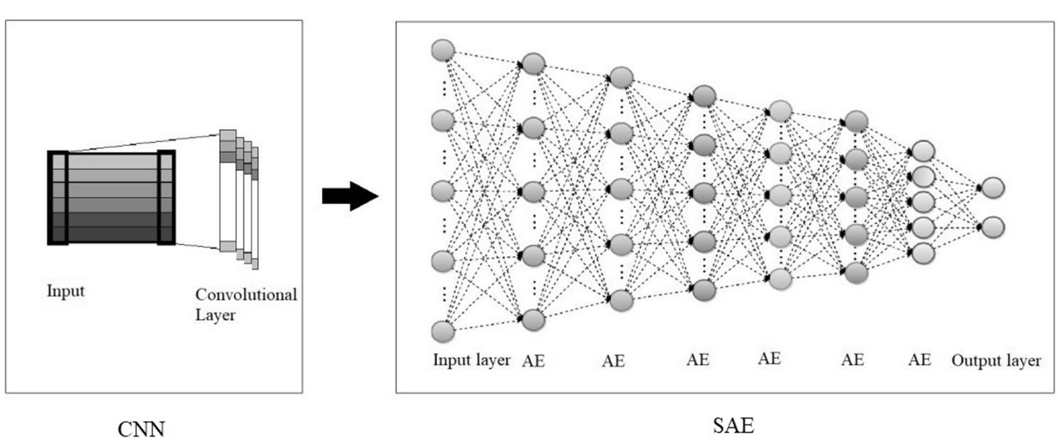


Fig. 8. Structure of CNN-SAE network as it is explained in [54]

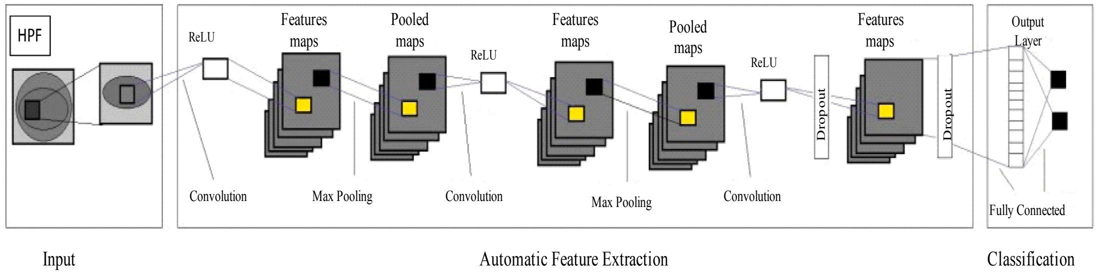


Fig. 9. The architecture of Two-Phase CNN Model that is proposed in [55]. In this model, a CNN model is applied to cope with a classification example by addressing the high similarities between mitoses and non-mitosis. In the second step, the blue ratio histogram-based K-means was proposed for under-sampling of the majority class skewness with little information loss.

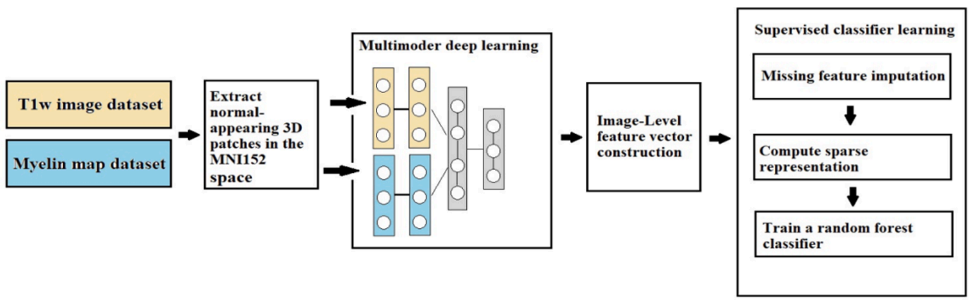


Fig. 10. The architecture of DBN Algorithm based on [58].

utilized with six hidden layers to classify it. Fig. 8 presents the structure of the CNN-SAE network as presented in [54].

A novel data-balancing technique using a CNN model was proposed for classifying mitotic and non-mitotic nuclei in breast cancer imagery in [55]. In this model, a CNN model is applied to cope with a classification example by addressing the high similarities between mitoses and non-mitosis. In the second step, the blue ratio histogram-based K-means was proposed for under-sampling of the majority class skewness with little information loss. The result reported in this study indicated that the model improved the mitosis detection rate of CNN and it reduced the training time. Fig. 9 presents the structure of this two-phase CNN model.

3.1.2. MS (Category 2)

A stacked denoising AE was developed in [56] to classify a mental workload. The accuracy of the classification was calculated for within-session and cross-session conditions. After that, this new classification was compared to traditional mental workload classifiers with different feature selection and noise corruption paradigms. Later on, a CNN method was used in [57] to extract latent multiple sclerosis (MS) lesion patterns from baseline images. A deep belief network and random forest applied to myelin and T1W images to identify MS pathology on normal-appearing brain tissue on MRI was proposed in [58]. In this model, a four-layer deep belief network (DBN) (Fig. 10) employed to 3D image patches of NAWM and NAGM to learn a latent feature representation. After that, Voxel-wise *t*-test was employed to select the image patches. Next, a feature selection was performed by the Least Absolute Shrinkage and Selection Operator (LASSO). Finally, features were used to train a random forest that would discriminate images of MS subjects from those of normal subjects.

3.1.3. Breast cancer (Category 3)

Later on, a two-layer deep learning architecture was introduced in [47] to classify the malignant and benign breast tumors with shear wave elastography. It includes a restricted Boltzmann machine and a point-wise gated Boltzmann machine. The method was compared to the statistical features quantifying image intensity and texture in terms of accuracy, specificity, AUC, and sensitivity. In [59], authors introduced a computer-aided diagnosis system with the utilization of deep learning techniques to detect, segment, and classify masses in mammograms. To do this, they considered three steps for this purpose that include detection, segmentation, and classification. A cascade of DL-methods such as CNN, DBN, and CRF was proposed for mass detection. A deep structure output was proposed for the segmentation of masses and a DL classifier was proposed for classification of masses to increase the performance of handcrafted features. CNN was trained to classify the mass in two phases. In the first phase, a regressor was estimated from the handcrafted features and the second phase fine-tuned the CNN model. The architecture of this method is demonstrated in Fig. 11.

Authors in [60] designed a CNN for region classification of semantically coherent tissues. This method was also used for the detection of masses in mammograms. The results indicate the high accuracy of the computation and classification. In [61], the authors proposed a DL-

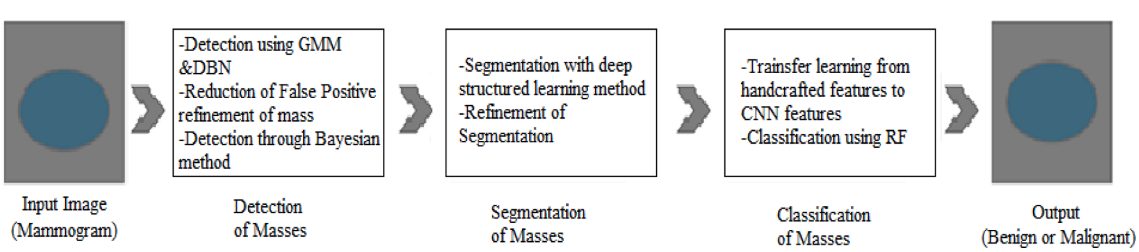


Fig. 11. The three steps deep learning architecture method proposed in [59].

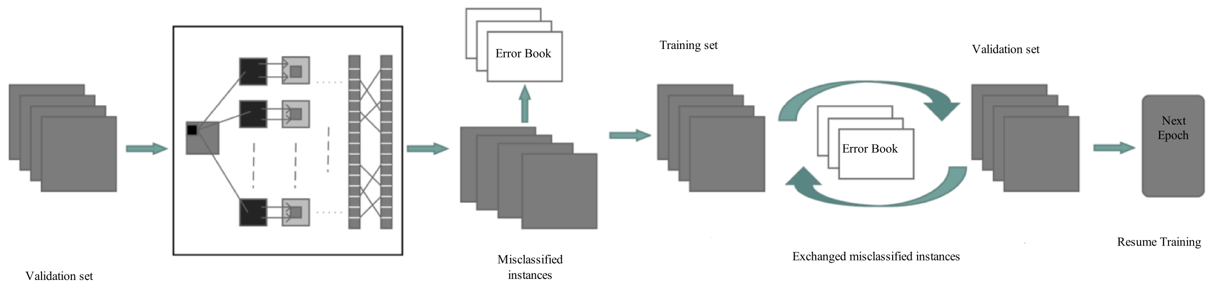


Fig. 12. The architecture of DL-method consists of parasitic metric layers and CNN layers for breast masses classification proposed in [61].

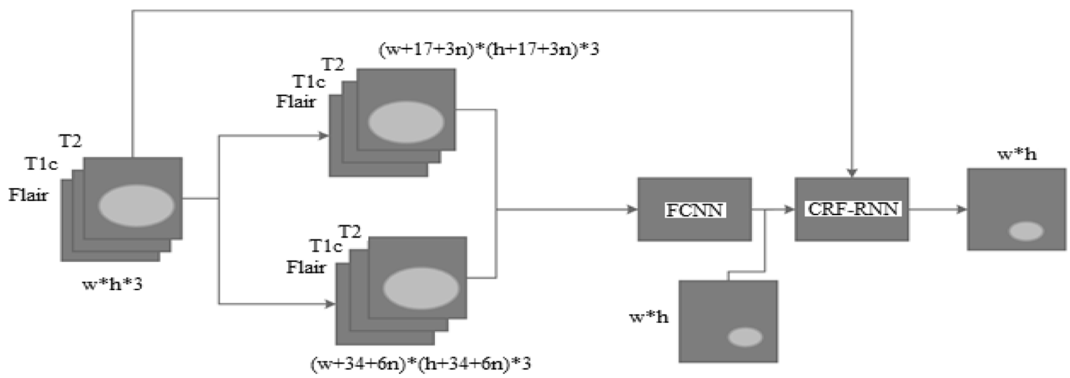


Fig. 13. The architecture of the FCNN-CRF Method based on [65].

method that consists of parasitic metric layers and CNN layers for breast masses classification. At first, CNN layers prepare basic discriminative representation, then the metric layer increases the performance of the classification. In each training cycle, misclassified instances are stored in an "errored book". The stored instance is sent to the training set if the validation error does not get reduced in a cycle. As shown in Fig. 12, these steps are repeated in the next cycle until the process is completed. A CNN was also applied to another research in breast cancer tomosynthesis (DBT) [62].

3.1.4. *Brain cancer (Category 4)*

A tissue segmentation method was proposed in [63] for brain-MRI which uses the multi-scale CNN. The method was tested with various degrees of abnormality. The results show that it can segment brain tissues, accurately. Model transparency is an important issue in the clinical domain which affects the real-world medical decision-making and patient treatments in terms of prediction [64]. In a different study, [65] proposed a new deep learning model for brain tumor segmentation by integrating Fully Convolutional Neural Networks (FCNN) and Conditional Random Fields (CRF) (Fig. 12). Such integration aimed at enhancing the robustness of the system. The model was trained in three steps. First, FCNN was trained with the use of image patches. Second, CRF-RNN was trained with the use of image slices. Finally, the image slices were utilized to fine-tune the whole network [65]. The method consists of four steps that include pre-processing, segmenting image slices using deep learning models with integrated FCNNs and CRF-RNN (from axial, coronal and sagittal views), feature extraction. and classification.

The CRF method has also been formulated as RNN. The model was trained in three steps. First, FCNN was trained with the use of image patches. Second, CRF-RNN was trained with the use of image slices. Finally, the image slices were utilized to fine-tune the whole network. After these three steps, the model was applied to image slices for tumors segmentation. As shown in Fig. 13, the image slice with (w*h) size and three channels include T1c, T2 and flair scans was padded and then two larger images were created. These two images then were given to the FCNN as input. As a result, five label prediction images were produced from the FCNN procedure. Finally, an image slice and these five label prediction images were given to the CRF-RNN, and as an output, an optimized segmentation result of the first image was created.

3.1.5. *Hybrid detection: Lung Adenocarcinoma, Stomach Adenocarcinoma and Breast Invasive Carcinoma (Category 5)*

Authors in [63] used an ensemble deep learning for identification of three types of cancers namely, Lung Adenocarcinoma, Stomach Adenocarcinoma, and Breast Invasive Carcinoma. To build this model, they selected the important genes using differentiated gene expression analysis. These selected genes were then transmitted to train five CNN classifiers and ensembled them to obtain the final result. They demonstrated that their model can enhance the accuracy of cancer prediction for all the tested RNA-Seq data sets compared to a single classifier or the majority voting algorithm [66]. In [67], the authors proposed a method using a CNN for automatic classification of gastric carcinoma. In a different study [68] implemented three multichannel ROI based deep structured algorithms namely, CNN, DBN and stacked denoising AE for lung cancer diagnosis.

3.1.6. *Epilepsy diagnosis (Category 6)*

In [69], the authors provided a new computer-aided diagnosis system using CNN to analyze the encephalogram signals for an epilepsy diagnosis. This CNN method consists of 13-layers to build a complex and strong model to detect seizure (normal and preictal classes) and it is compared to other machine learning methods in terms of specificity, accuracy, and sensitivity. To the best of our knowledge, this is the only significant use of DL for an epilepsy diagnosis.

3.1.7. *Heart disease (Category 7)*

Despite the widespread of this disease, DL has not been extensively used for it. The most significant use of DL is to assist in detecting this disease as is done in [70]. They proposed an RNN model with gated

Table 1
Summary of the articles that have used DL in HCS.

Disease Category	Article	Year	Data	DL Method
Breast Cancer	Dhungel et al. [59]	2017	INbreast	CNN + DBN
	Jiao et al. [61]	2018	DDSM MIAS	CNN
	Samala et al. [62]	2016	DM,SFM-UM DM,SFM-USF	DCNN
	Wahab et al. [55]	2017	MITOS12 TUPAC16	CNN
	Zhang et al. [47]	2016	227 SWE Images of 121 Female Patients	DBN (PGBM)
Gastric Cancer	Sharma et al. [67]	2017		CNN
Lung Cancer	Sun et al. [68]	2017	LIDC/IDRI	CNN + DBN + SDAE
Brain Cancer	Zhao et al. [65]	2018	BRATS 2013,2015,2016	FCNN + CRF-RNN
	Tabar & Halici [54]	2016	dataset III from BCI Competition II dataset 2b from BCI Competition IV	CNN + SAE
Cancer	Xiao et al. [66]	2018	RNA-seq (LUAD + STAD + BRCA)	DNN
Multiple sclerosis	Yoo et al. [57]	2017	Myelin and T1w Images	CNN
	Yoo et al. [58]	2018	Myelin and T1w Images	DBN
Parkinson	Choi et al. [71]	2017	PPMI SNUH	CNN
Heart Failure	Choi et al. [70]	2016	Sutter Palo Alto Medical Foundation (Sutter-PAMF)	RNN adopted by GRU
	Al Rahhal et al. [52]	2016	MIT-BIH INCART SVDB	DNN + DAE
Epilepsy and Seizure	Acharya et al. [69]	2017	EEG signals from the Bonn University database	CNN
Diabetic Retinopathy	Orlando et al. [74]	2018	e-ophtha DIARETDB1 MESSIDOR	CNN + HCF
Mental Workload	Yin & Zhang [56]	2017	–	SDAE
Multiple Disease	Miotto et al. [73]	2016	–	SDN
	Nie et al. [64]	2015	EveryoneHealthy WebMD MedlinePlus	SCDP

recurrent units to detect heart failures by checking the relations among time-stamped events with an observation window of cases and controls.

3.1.8. Parkinson's disease (Category 8)

A deep learning-based FPCIT SPECT interpretation system was proposed in [71] to refine the Parkinson's disease. It was shown in this study that this method can overcome the interobserver variability issue. More recently, in [72], DL was used for Parkinson's disease classification. To do this, they used 9-layered CNN on the set of vocal (speech) features. They demonstrated the effectiveness of their employed classifier to solve this problem.

3.1.9. Eye disease (Category 9)

An unsupervised deep feature learning algorithm was proposed in [73] to provide a machine learning framework for augmenting clinical decision systems. An ensemble deep learning and domain knowledge was proposed in [74] to detect red lesions in fundus images. The learning process of CNN architecture was completed by incorporating handcrafted features. The performance of this combination was reported to be higher than other individual classifiers [74].

Table 1 lists the number of DL methods applied in different disease detections. For example, [70] applied an RNN for heart failure detection. It uses the Sutter-PAMF dataset. The result indicates the 77.68% accuracy of detection. Table 1 also shows the RNN method proposed in [65] to detect a brain tumor. BRATS dataset utilized clinical imaging data of 65 glioma patients, including 14 patients with low-grade gliomas (LGG) and 51 patients with high-grade gliomas (HGG). The performance of the DNN algorithm has been validated using RNA-seq data from cancers such as Lung Adenocarcinoma (LUAD), Stomach Adenocarcinoma (STAD) and Breast Invasive Carcinoma (BRCA). The gene expression data used are from The Cancer Genome Atlas (TCGA) project [66].

The result of DNN on LUAD indicates an accuracy of 98.8%, ROC of 0.988 and PR Curve of 0.994. Moreover, DNN and DAE methods were applied to MIT-BIH, INCART, and SVDB in [52]. Arrhythmia Database (MIT-BIH), consists of 48 two-lead half-hours long at 360 Hz recordings which contain annotation for both beat timing and class information. St. Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database (INCART) consists of 75 recordings extracted from 32 Holter records including 12 standard leads collected from 17 men and 15 women, aged between 18 and 80. MITBIH Supraventricular Arrhythmia Database (SVDB), consists of 78 two-lead recordings in 30 min at 128 Hz. The result of INCART database shows an accuracy of 99.91%, a sensitivity of 92.85%, a specificity of 93.85%, and a positive predictive value of 99.34%.

BCI Competition IV dataset 2b was analyzed by the SAE method. Those datasets include MI task experiments for the right hand and left-hand movements. The BCI Competition IV dataset 2b includes three sessions in the training set [54]. The results indicate an accuracy of 77.6% and a kappa value of 73.2%. A data set of images is selected via a voxel-wise *t*-test [58]. The outcome of DBN gave an accuracy of 75%, sensitivity of 78.7%, specificity of 70.4%.

3.2. Dataset used for DL methods-based healthcare

In this section, we highlight those datasets which were utilized for applying different deep learning algorithms in healthcare and disease detection. Table 2 shows the mostly used healthcare or cancer dataset for DL methods.

Table 2 indicates that the CNN algorithm has the highest number of implementations in various databases. For example, MITOS12 which is provided by the ICPR 2012 contest organizers. It comprises 5 slides from 5 patients, stained with H&E stains, and marked by an experienced pathologist [55]. The CNN has been applied to both MITOS12 and TUPAC16. DDSM database is provided by the University of South Florida (USF). Each case in DDSM includes four images of the breast, along with associated patient information [61].

4. Comprehensive comparison among DL methods applied in the healthcare system

In this section, we discuss the findings of the studies presented based on the architecture of DL's methods applied for disease detection. We then, analyze the strengths and weaknesses of each study. We first discuss each study individually and then compare it with other methods mentioned in Table 1. To evaluate the performance of DL methods, two important factors such as accuracy (ACC) and area under the curve (AUC) investigated from the total DL's manuscripts mentioned in Table 1. Finally, we highlighted those techniques which have more influences in terms of rate of detection.

Acharya *et al.*[69] employed a CNN technique with a ten-fold cross-validation strategy to analyze EEG signals in a database of Bonn University. In this study, the results were compared with those other similar studies found in the literature. Comparisons have been performed in terms of accuracy, sensitivity, or specification. Fig. 14 presents the results of the study proposed in [69]. These are comparison results of the CNN method with the other techniques developed in this database in terms of accuracy (a) and sensitivity (b).

Table 2
The dataset and the deep learning architecture of the studies in Table 1 that have used DL in HCS.

Paper number	Dataset	CNN	RNN	DNN	DAE	DBN	SDN	SAE	KNN
P1 [75]	Bonn University dataset [69]	√							
P2 [76]	PPMI, SNUH [71]	√							
P3 [71]	Electronic health record dataset [70]		√						
P4 [59]	INBreast [59]	√							
P5 [61]	DDSM [61]	√							
P6 [73]	Electronic health record dataset [77]							√	
P7 [64]	Dataset related to the popular disease concepts from Everyone Healthy, WebMD and Medline Plus [78]								√
P8 [74]	e-ophtha, DIARETDB1 and MESSIDOR [79]	√							
P9 [80]	MIT-BIH, SVDB, INCART [52]			√	√				
P10 [62]	screen-film mammograms and digital mammogram datasets [81]	√							
P11 [82]	Gastric cancer dataset [83]	√							
P12 [46]	Lung Image Database Consortium and Image Database Resource Initiative [84]	√			√	√			
P13 [80]	Dataset III from BCI, Dataset 2b from BCI [54]	√						√	
P14 [55]	MITOS12, TUPAC16 [55]	√							
P15 [66]	LUAD, STAD, BRCA [66]			√					
P16 [56]	Electroencephalogram dataset [85]				√				
P17 [86]	Myelin and T1w [57]	√				√			
P18 [58]	MRI dataset for relapsing-remitting MS [87]					√			
P19 [47]	Dataset related to shear-wave elastography experiments [88]					√			
P20 [65]	BRATS [65]	√	√						

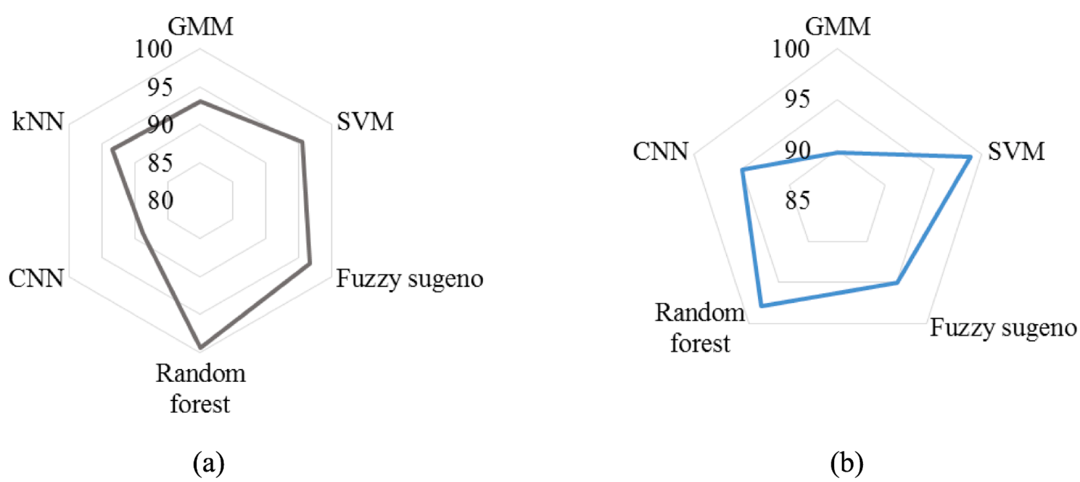


Fig. 14. Comparison between CNN and other conventional machine learning models in terms of accuracy. The results were obtained from [69].

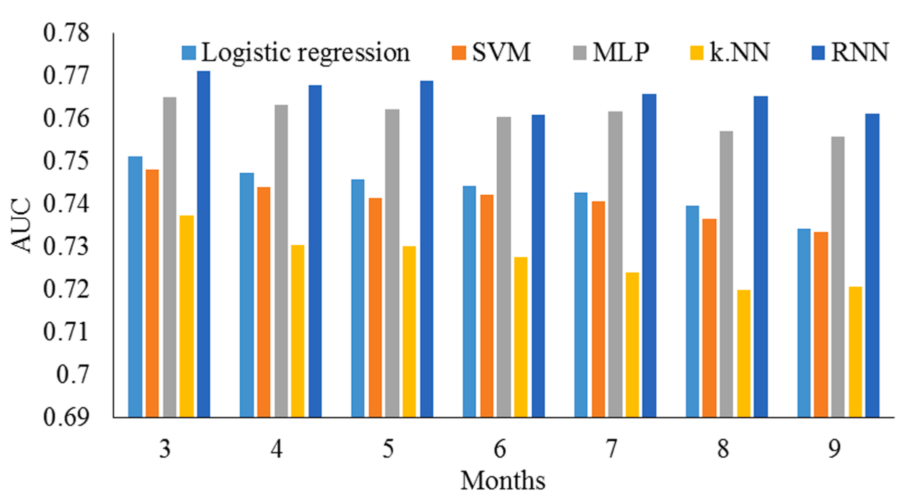


Fig. 15. Comparison of AUC between SVM, MLP, KNN and RNN in terms of prediction between 3 and 9 months. The results were obtained from [70].

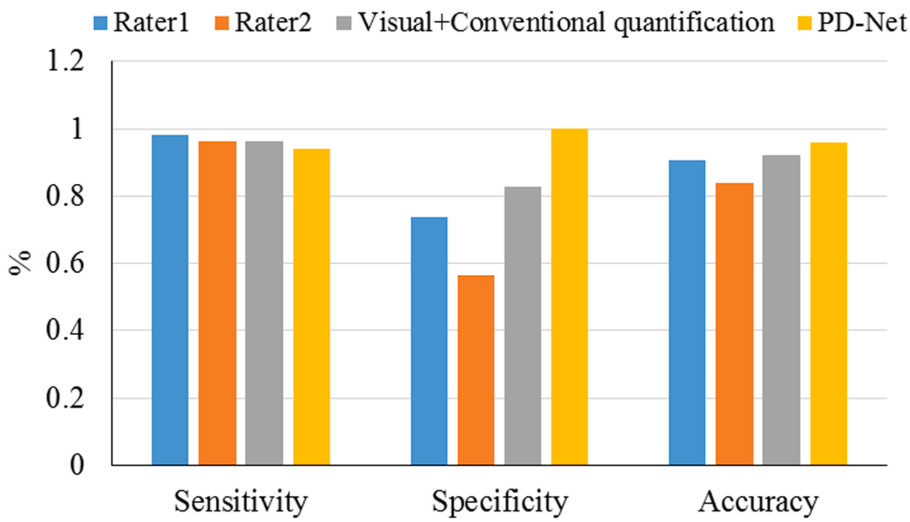


Fig. 16. Comparison of accuracy between PDNet and other methods on the re-classification of the SWENN group. As the plot shows PDNet shows an acceptable accuracy. The results were obtained from Choi *et al.* [71].

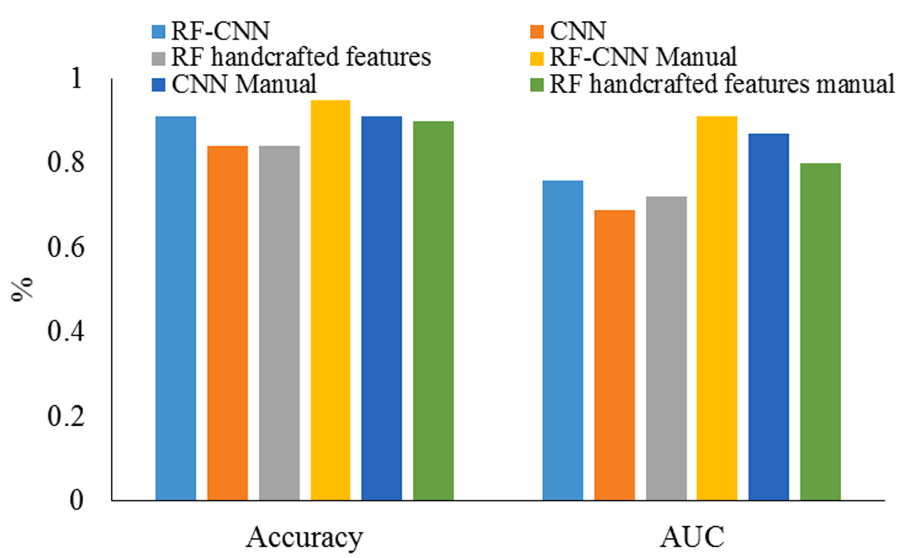


Fig. 17. The accuracy comparison between Random Forest based CNNs for different setups. The results were obtained from [59]. As shown in this figure, RF-CNN Manual configuration obtains the best results in terms of AUC and accuracy.

Based on Fig. 14 (a), CNN presented in [69] achieves the lowest accuracy among the other techniques presented using the same database. Fig. 14 (b) indicates that Fuzzy Sugeno and Random Forest provides the highest sensitivity values. The main reason for the poor performance of the CNN technique is the size of the benchmark. Given the small size of this database, the CNN method was unable to perform well, as the application of this model is to extract features for large datasets. Therefore, we hypothesize that the use of this model in such a database could not produce the desired accuracy.

Choi *et al.* [70] developed an RNN based on gated recurrent units (GRUs) to analyze time-stamped events. The performance of the model was compared with the ordinary neural network, K-nearest neighbor classifier, logistic regression, and support vector machine using the area under the curve (AUC) values as the performance measurement. Fig. 15 shows the AUC of RNN in terms of prediction between 3 and 9 months in comparison with other methods.

As is demonstrated in Fig. 15, in each month, the RNN method provides a higher AUC compared to other methods found in the literature [70].

Choi *et al.* [71] developed an automated CNN technique for Parkinson's disease diagnosis [71]. The visual interpretation and PD Net were compared in terms of accuracy, specificity, and sensitivity. Two readers (Rater 1 and Rater 2) visually reviewed images of the PPMI test set blinded to the diagnosis and clinical information. Images were visually labeled with 'normal' and 'abnormal' DAT binding. The accuracy of PD Net was compared with that of readers. Fig. 16 presents the visualized results.

Based on the results presented in Fig. 16, the PDNet method indicates an acceptable accuracy which makes it reliable for the reclassification of the SWEDD group. This makes the method more reliable to be applied in the future works in the application to imaging interpretation in various diseases.

Dhungel *et al.* [59] developed a Random Forest-based CNN (RF-CNN) for the application to the INbreast dataset into two setups including

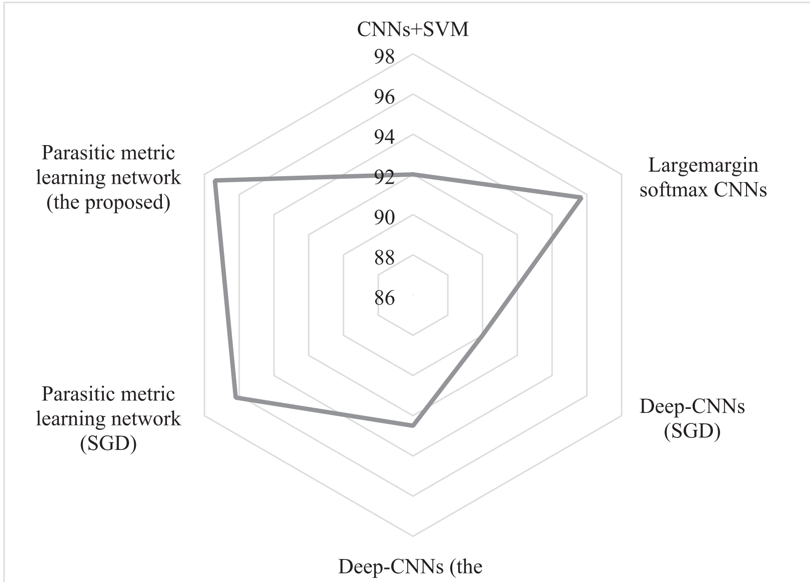


Fig. 18. Comparison of parasitic metric learning net for different CNN-based techniques for breast mass cancer classification obtained from [61]. As shown in this figure, large margin softmax CNN and parasitic metric learning network obtain the best results compared to other classifiers. Further discussion is provided in [61].

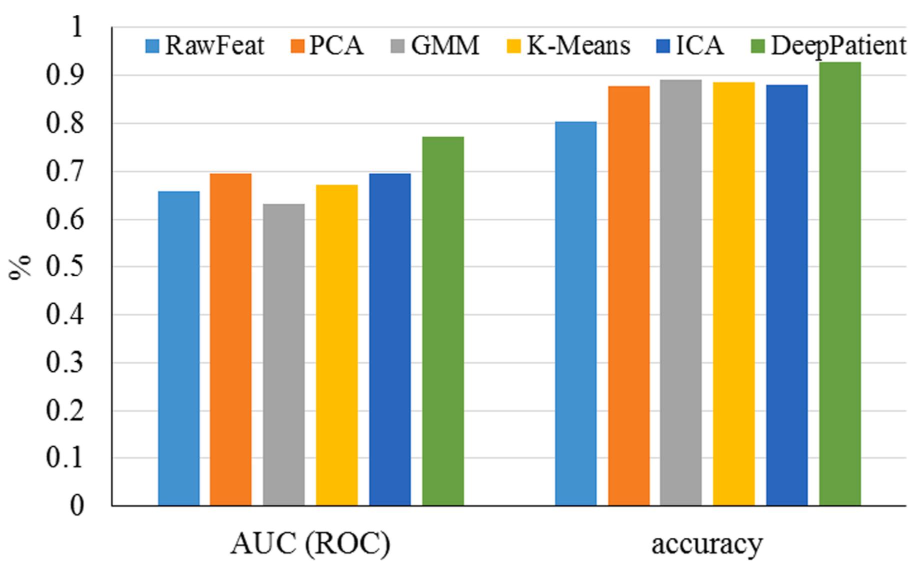


Fig. 19. Results achieved using URPD (Deep Patient) compared to similar approaches found in the literature.

minimal interaction and manual. Fig. 17 shows the results in terms of accuracy and the area under the curve.

Fig. 17 indicates high accuracy of detection for the random forest-based CNN technique in the manual setup followed by random forest-based CNN technique in minimum user interaction setup. This means that the RF has a positive effect on the CNN method because of its ability in classification using an ensemble learning method. Therefore, this ability increases the CNN technique's accuracy [59].

In the study by Jiao *et al.*, [61]. a parasitic metric learning method compared with a CNN technique for breast mass classification. CNN's method was compared with CNN's-SVM and the Stochastic Gradient Descent (SGD) based CNN techniques developed by other researchers in the same dataset size in terms of accuracy. The results are shown in Fig. 18.

As it is shown in Fig. 18, both large margin CNN and parasitic metric learning obtained a high accuracy compared to those other similar

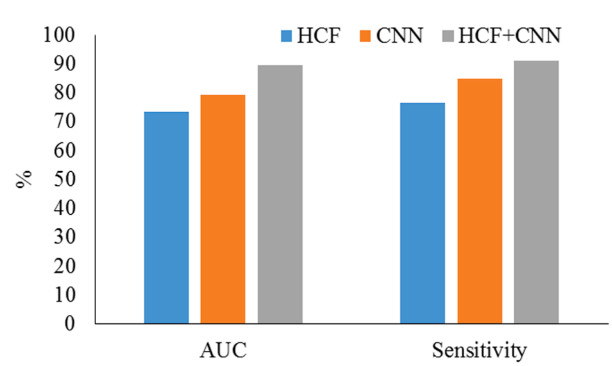


Fig. 20. Comparison of hand-crafted features (HCF), CNN, and hybrid CNN-HCF deep learning methods developed in detecting a red lesion in fundus images. The results were obtained from Orlando *et al.* [74].

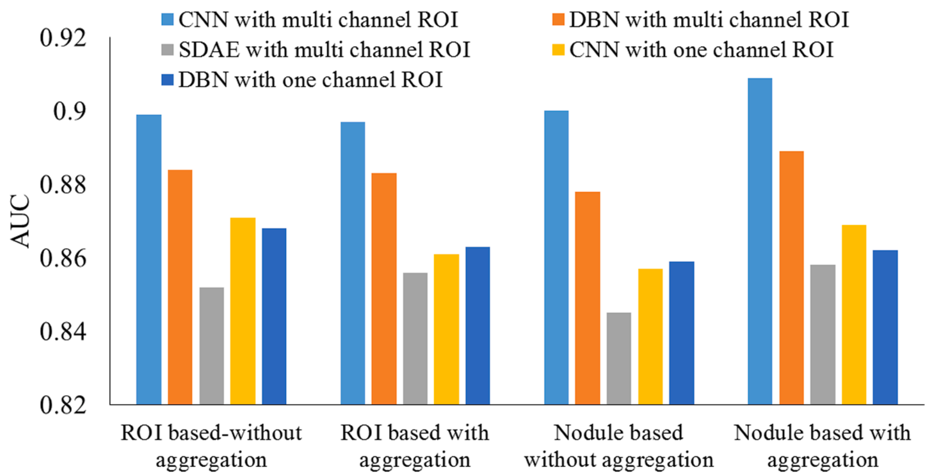


Fig. 21. AUC comparison of CNN, SDAE, and DBN with one and multi-channel ROI for lung cancer detection. the results were obtained from Sun *et al.* [68].

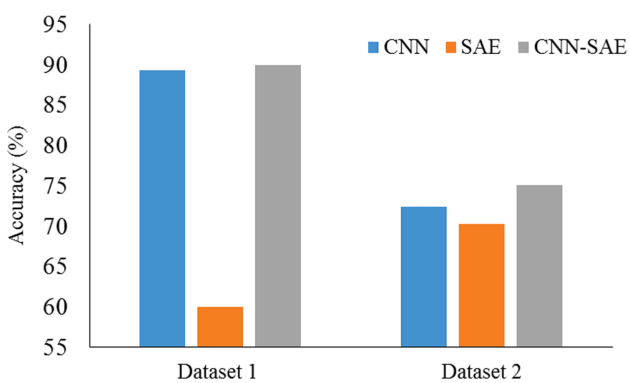


Fig. 22. AUC comparison between CNN, SAE, and CNN-SAE models in the classification of signals for imagery of the EEG motor. The results were obtained from Tabar & Halici [54].

studies found in the literature.

Miotto *et al*, [73] proposed an unsupervised representation deep patient (URDP) method called the deep patient to estimate patients' survival rate using the electronic health records. The URDP method compared with other techniques developed in this field in terms of ACC and AUC of the ROC. The results are presented in Fig. 19.

URDP provides dependencies in the data to make a general-purpose and compact set of patient features [73]. This employs an effective prediction in clinical applications. It can be hypothesized that employing a deep learning method for a nonlinear transformation in terms of pre-processing, increases the understanding ability of the machine in extracting the information embedded in the EHRs. Pre-processing EHR data with deep learning also improves ad-hoc frameworks which can be considered as the new achievement for clinical predictive modeling.

In the study by Orlando *et al.* [74] an ensemble deep learning method was developed for detecting a red lesion in fundus images. Results of the hand-crafted features (HCF), CNN, and hybrid CNN-HCF in terms of sensitivity and AUC are shown in Fig. 20.

Based on the results, this study points out the importance of hybrid methods as it integrates several robust models to build a more accurate classifier [74]. As it is shown in Fig. 20, the hybrid method provides higher sensitivity and AUC compared to those for single methods. This study emphasizes the use of hybrid methods in deep learning [74].

Sun *et al.* [68] developed CNN with one and multi-channel ROI and presented a comparison among CNN techniques and other methods for lung cancer diagnosis in the term of AUC. Their results are presented in Fig. 21.

Based on these results shown in Fig. 21, deep algorithms generate the desired performance in lung cancer diagnosis successfully [68]. Large enough datasets, including well-tuned parameters, help deep learning algorithms to produce better performance compared with that for the current popular CAD. This is the main feature of deep learning techniques. This study also confirms the highest ability of deep learning methods for adopting the huge data set and building an accurate model compared to the similar data preprocessing procedures [68]. This study

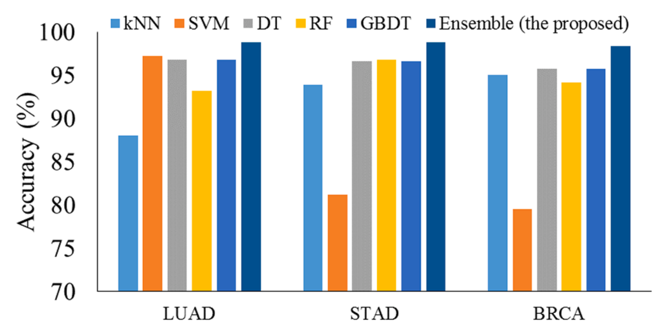


Fig. 23. AUC comparison between the multi-model ensemble method proposed by Xiao *et al.* [66] and other methods in cancer prediction. The results were obtained from Xiao *et al.* [66].

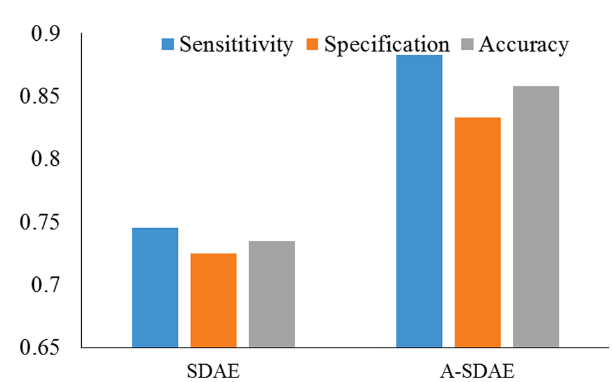


Fig. 24. A performance comparison between SDAE and A-SDAE for classification of the mental workload levels. The results were obtained from Yin & Zhang [56].

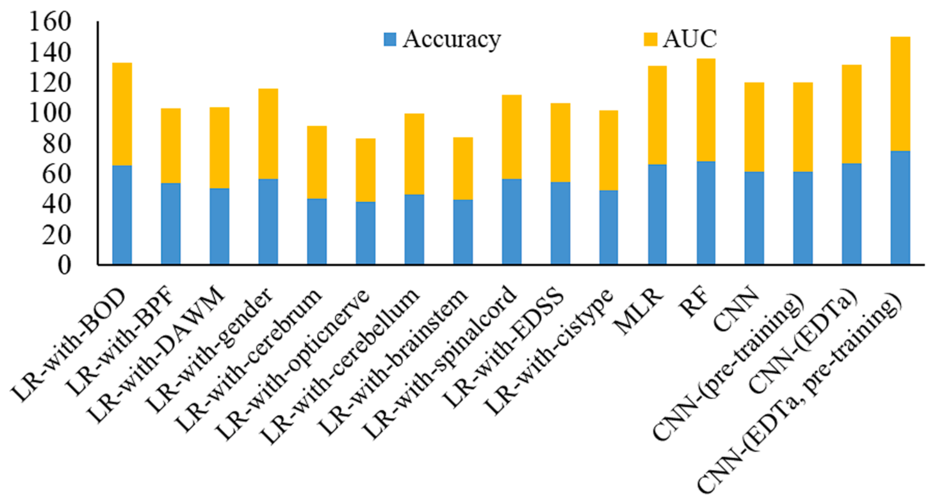


Fig. 25. Results achieved using CNN based techniques compared with other methods in terms of accuracy and AUC in [57].

introduces another ability for deep learning techniques to be used in other medical image analysis areas as well [68].

In the study by Tabar *et al.* [54]. CNN, SAE, and CNN-SAE techniques employed to classify signals for imagery of the EEG motor. The performances of these methods were compared using the accuracy factor which is shown in Fig. 22.

Fig. 22 describes two different classifications of deep learning, i.e. single and ensemble. Two single methods (CNN and SAE) developed and compared with their hybrid method counterpart. Both single methods have their advantages and disadvantages. When they were employed in their single form, their accuracies achieve lower than those for their hybrid form. In fact, in the case of using their hybrid form, their advantages joined together and increased the accuracy and cover each other's weaknesses. This research, similar to the study by Dhungel *et al.* [59], emphasize the importance of hybrid methods. Similar to the previous study, it needs to be restructured.

Xiao *et al.* [66] proposed a multi-model ensemble method based on a deep learning technique for the prediction of Cancer [66]. The database includes informative gene data that were selected from differential gene expression analysis. The results achieved for this method compared to five different classification models in the term of accuracy are shown in Fig. 23.

As shown in Fig. 23, the ensemble method tested on three datasets which includes three kinds of cancers including Breast Invasive Carcinoma (BRCA), Stomach Adenocarcinoma (STAD) and lung Adenocarcinoma (LUAD). It can be seen from the results that the ensemble method increases the prediction accuracy of cancer for all datasets.

Yin & Zhang [56] developed Stacked Denoising Auto-Encoder (SDAE) and Adaptive SDAE methods for classification of the mental workload levels. The methods were trained and tested using EEG signals which have been recorded on separate days. Results reported in this study in terms of sensitivity, specification and accuracy are shown in Fig. 24.

Based on the results presented in Fig. 24, adaptive SDAE (A-SDAE) produces better results compared to SDAE in dealing with the cross-session EEG features [56]. By analyzing the results presented in this article, the superiority of A-SDAE is also evident in cases of the computational cost for iterative tuning, the data augmentation scheme, and optimal step length. This method is implemented as an online web server [56].

In the study by Yoo *et al.*, [57]. CNN based methods developed for the production of multiple sclerosis using brain lesion patterns and user-defined clinical and MRI features. The results achieved by CNN compared to other methods used to tackle this problem are shown, in terms of accuracy, sensitivity, specificity, and AUC in Fig. 25.

Based on Fig. 25, the DL method for the prediction of MS from brain lesion patterns can be combined with user-defined measurements. This will help to predict the short-term risk for MS patients [57].

Zhang *et al.* [47] presented a DL structure based on the RBM technique for extracting image features from the shear-wave elastography automatically. RBM compared to the other state-of-the-art methods in terms of accuracy, sensitivity, and specificity are shown in Fig. 26.a. In Fig. 26.b the time complexity of those methods is also recorded and compared.

As it is shown in Fig. 26.a, the best method is PGBM-RBM-SVM with the highest accuracy and sensitivity compared with other methods. This method is considered as a ternary hybrid method. This approach increases the accuracy compared with other hybrid methods and single methods. On the other hand, based on Fig. 26.b the processing time for the ternary hybrid methods is significantly higher than that for the other methods. Considering that the first aim is to be able to accurately determine breast cancer tumor, then the high accuracy makes up for the extra time.

To highlights the strength and weakness of each DL's methods applied in disease detection, in this section we compared different DL techniques (single DL and hybrid-DL) individually, based on the type of dataset. Table 3 presents results, dataset, and method for each reference reviewed in this article. Fig. 27 shows the comparison results of DL techniques in terms of AUC (%) and Fig. 28 indicates comparing the results of DL techniques in terms of accuracy (%).

Table 3 indicates demonstrates 20 highly cited research papers related to the application of DL techniques for the disease detection that are comprehensively compared and analyzed here in terms of accuracy, specificity and sensitivity, and AUC.

Figs. 27–29 present DL techniques to demonstrate modeling productivity and efficiency in terms of AUC, accuracy, and sensitivity, respectively. It can be implied that the CNN method followed by DBN has made the highest share of contribution in this field. This can be due to the nature and characteristics of these methods. These two methods also have the highest tendency to be used as a hybrid with other methods.

5. Conclusion

As an emerging technique, deep learning has demonstrated tremendous potential in tackling challenging problems in healthcare. In

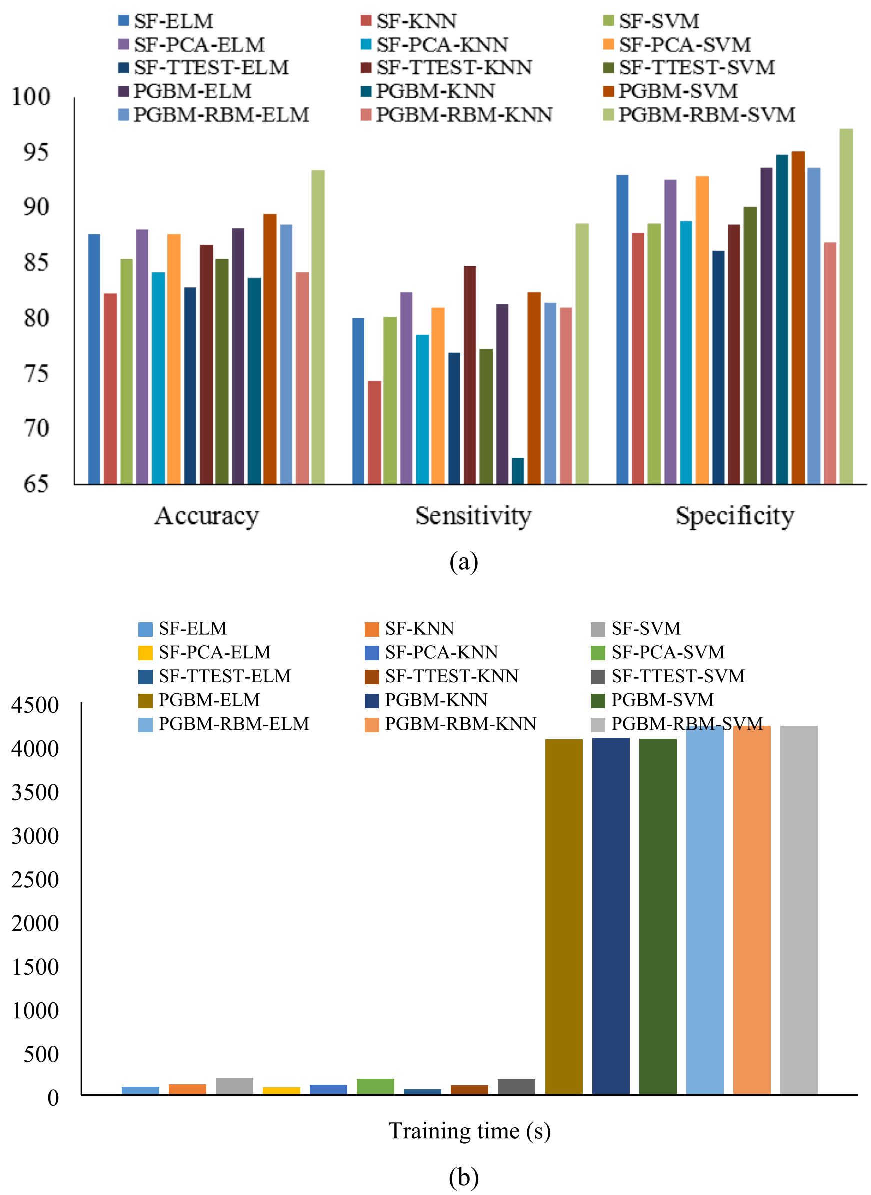


Fig. 26. Results achieved using RBM compared to similar methods found in the literature in term of accuracy (a) and with respect to training time (b) presented in [47].

this study, we focused on those problems in healthcare that have been addressed using deep learning with promising results. Deep learning-based techniques that have been shown as powerful tools in dealing with disease detection in preprocessing, feature extraction, feature selection, classification, and clustering steps. The technical aspects of ML and DL architectures were evaluated in this paper by focusing on disease detection in healthcare systems. The performance of these methods was discussed in terms of the accuracy of disease detection and algorithm parameters. Finally, the top architectures of DL methods applied to healthcare were analyzed and discussed.

As a summary, we can conclude that hybrid and ensemble methods based on DL produce better accuracy results compared to single techniques. These methods lead to an improvement of the process by combining two or more methods simultaneously and independent of the type of the datasets.

On the negative side, deep learning approaches are memory and time-consuming. Hence, designing and applying optimal methods in healthcare systems is an important challenge.

In the future, researchers may also put their efforts into developing and integrating efficient technologies to fulfill the hardware requirements of decision making systems including those with a deep neural network structure [89]. As discussed in this paper, to build more effective systems, it is also necessary to improve the structures of the current neural network models. Therefore, to solve complex problems in the healthcare systems, it is necessary to create well-defined architectures that are general and work with different types of health data [90]. It is also important to implement deep learning models in terms of Explainable Artificial Intelligence (XAI) to apply in distributed systems which can significantly improve the processing time [90].

To conclude, we believe that there is tremendous potential for the deep learning models and their applications in medicine and healthcare systems especially considering the size and complexity of health data.

Table 3
Results of different studies using DL that are covered in this article in terms of accuracy, specificity, and sensitivity.

Paper number	Method	Dataset	Accuracy (%)	Specificity (%)	Sensitivity (%)	AUC (%)
P1 [75]	DL Method	EEG signals from the Bonn University database	88.7	90	95	N.A.
P2 [76]	CNN	Sutter-PAMF	N.A.	N.A.	N.A.	77.68
P3 [71]	RNN adopted by GRU	PPMI-SNUH	98.8	100	96.6	N.A.
P4 [59]	CNN	INbreast	95	N.A.	N.A.	91
P5 [61]	CNN + DBN	DDSM	97.40	N.A.	N.A.	N.A.
		MIAS	96.7	N.A.	N.A.	N.A.
P6 [73]	CNN	–	92.9	N.A.	N.A.	77.3
P7 [64]	SDN	WebMD	98.21	N.A.	N.A.	N.A.
	SCDP	Medline Plus	91.48	N.A.	N.A.	N.A.
P8 [74]	CNN + HCF	e-ophtha DIARETDB1 MESSIDOR	N.A.	N.A.	91.09	89.32
P9 [80]	DNN + DAE	MIT-BIH INCART SVDB	99.91	93.85	92.85	N.A.
P10 [62]	DCNN	DM,SFM-UM,DM,SFM-USF	N.A.	N.A.	91	80
P11 [82]	CNN	–	81.44	N.A.	N.A.	N.A.
P12 [46]	CNN + DBN + SDAE	LIDC/IDRI	N.A.	N.A.	N.A.	90.9
P13 [80]	CNN + SAE	dataset III from BCI Competition II dataset 2b from BCI Competition IV	77.6	N.A.	N.A.	N.A.
P14 [55]	CNN	MITOS12 TUPAC16	83	N.A.	N.A.	N.A.
P15 [66]	DNN	RNA-seq (LUAD + STAD + BRCA)	98.8	N.A.	N.A.	N.A.
P16 [56]	SDAE	–	85.79	83.28	88.3	N.A.
P17 [86]	CNN	Myelin and T1w images	75	70.4	78.7	74.6
P18 [58]	DBN	Myelin and T1w images	87.9	88.6	87.3	88
P19 [47]	DBN(PGBM)	227 SWE Images of 121 Female Patients	93.4	97.1	88.6	94.7
P20 [65]	FCNN + CRF-RNN	BRATS 2013,2015,2016	N.A.	N.A.	0.83	N.A.

N.A.: Not-Available.

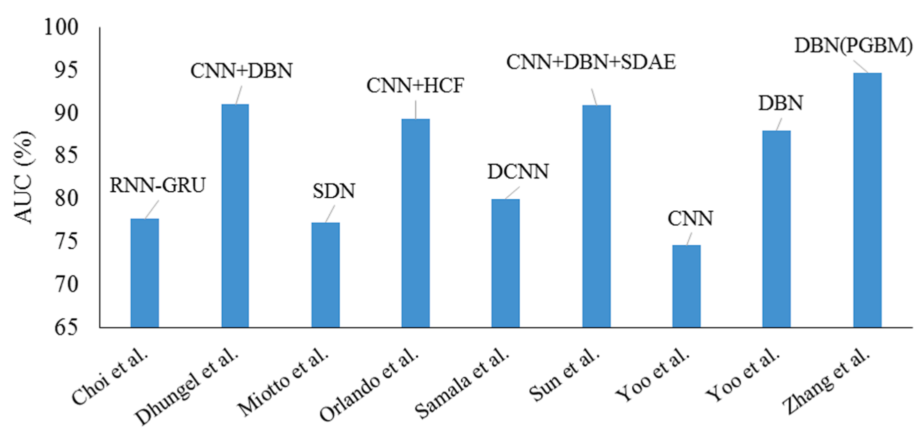


Fig. 27. Comparing the results of different DL techniques in term of AUC (%).

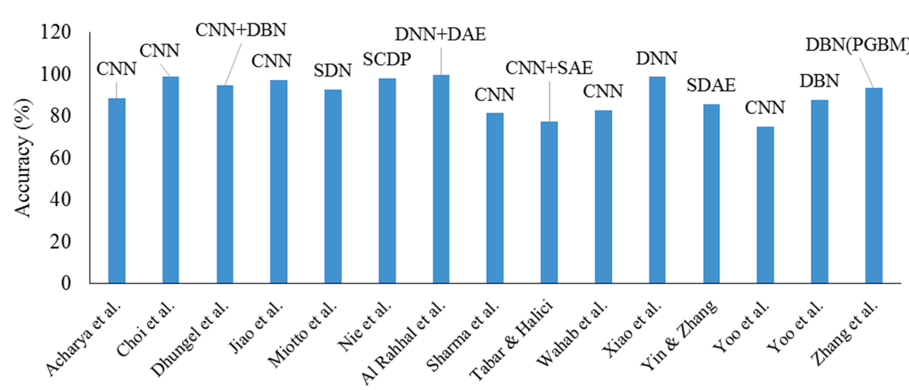


Fig. 28. Comparing the results of different DL techniques in term of Accuracy (%).

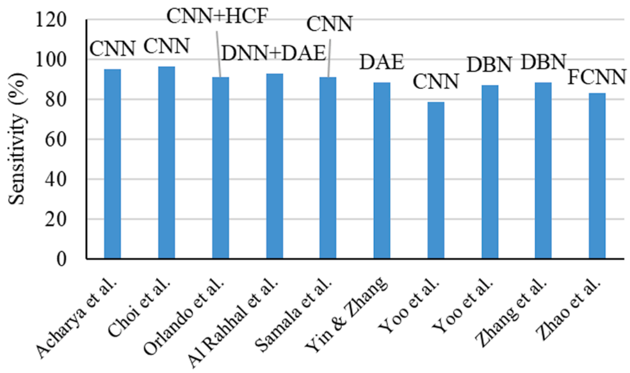


Fig. 29. Comparing the results of different DL techniques in term Sensitivity (%).

Author contributions

HAR and SS designed the study. SS analyzed and visualized the results. SS, MF, AD and ATC wrote the paper. AD and ATC revised the paper. HAR wrote the abstract and introduction sections and has not been involved in the writing, figures, and tables generation of other sections.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank Keith Downing and Agnar Aamodt for valuable technical comments. This work has been supported by the Norwegian Open AI Lab.

References

[1] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, A. T. Chronopoulos, Computational intelligence approaches for classification of medical data: state-of-the-art, future challenges and research directions, Neurocomputing 276 (2018) 2–22.
[2] P.M. Kumar, U.D. Gandhi, A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases, Comput. Electr. Eng. 65 (2018) 222–235.
[3] D. Ravı, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, et al., Deep learning for health informatics, IEEE J. Biomed. Health. Inf. 21 (2017) 4–21.
[4] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, D. Yu, Convolutional neural networks for speech recognition, IEEE/ACM Trans. Audio Speech Lang. Process. 22 (2014) 1533–1545.
[5] L. Deng, X. Li, Machine learning paradigms for speech recognition: an overview, IEEE Trans. Audio Speech Lang. Process. 21 (2013) 1060–1089.
[6] K. Cho, A. Courville, Y. Bengio, Describing multimedia content using attention-based encoder-decoder networks, IEEE Trans. Multimedia 17 (2015) 1875–1886.
[7] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, et al., Deep speech: Scaling up end-to-end speech recognition, 2014, arXiv preprint arXiv: 14125567.
[8] DeepMind. DeepMind Technologies Limited, 2020.
[9] Watson, Watson IBM, 2020, p. 1-16.
[10] M. Cabrita, H. op den Akker, M. Tabak, H.J. Hermens, M.M. Vollenbroek-Hutten, Persuasive technology to support active and healthy ageing: An exploration of past, present, and future, J. Biomed. Inform. 84 (2018) 17-30.
[11] M. Avendi, A. Kheradvar, H. Jafarkhani, A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI, Med. Image Anal. 30 (2016) 108–119.
[12] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, J. Li, A robust deep model for improved classification of AD/MCI patients, IEEE J. Biomed. Health. Inf. 19 (2015) 1610–1616.
[13] B. Shickel, P.J. Tighe, A. Bihorac, P. Rashidi, Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis, IEEE J. Biomed. Health. Inf. 22 (2017) 1589–1604.
[14] K. Lan, D-t Wang, S. Fong, L-s Liu, K.K. Wong, N. Dey, A survey of data mining and deep learning in bioinformatics, J. Med. Syst. (2018;42:139.).
[15] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, p. 1097-105.
[16] G. Chen, A. Tsoi, H. Xu, W.J. Zheng, Predict effective drug combination by deep belief network and ontology fingerprints, J. Biomed. Inform. 85 (2018) 149–154.
[17] L. Macías-García, J.M. Luna-Romera, J. García-Gutiérrez, M. Martínez-Ballesteros, J.C. Riquelme-Santos, R. González-Cámpora, A study of the suitability of autoencoders for preprocessing data in breast cancer experimentation, J. Biomed. Inform. 72 (2017) 33–44.
[18] E. Laksana, M. Aczon, L. Ho, C. Carlin, D. Ledbetter, R. Wetzel, The impact of extraneous features on the performance of recurrent neural network models in clinical tasks, J. Biomed. Inform. 102 (2020) 103351.
[19] L. Deng, D. Yu, Deep learning: methods and applications. Foundations and Trends®, Signal Process. 7 (2014) 197–387.
[20] J. Schmidhuber, Deep learning in neural networks: an overview, Neural networks. 61 (2015) 85–117.
[21] W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, et al., Federated learning in mobile edge networks: a comprehensive survey, IEEE Commun. Surv. Tutorials (2020).
[22] D. Yu, L. Deng, Automatic speech recognition: a deep learning approach, Springer, 2014.
[23] A. Graves, A.-r. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks. Acoustics, speech and signal processing (icassp), 2013 ieee international conference on: IEEE; 2013. p. 6645-9.
[24] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M,. Ghafoorian et al., A survey on deep learning in medical image analysis. arXiv preprint arXiv: 170205747, 2017.
[25] J. Bian, B. Gao, T.-Y. Liu, Knowledge-powered deep learning for word embedding, Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2014, pp. 132-148.
[26] S. Min, B. Lee, S. Yoon, Deep learning in bioinformatics, Briefings Bioinf. 18 (2017) 851–869.
[27] E. Gawehn, J.A. Hiss, G. Schneider, Deep learning in drug discovery, Mol. Inf. 35 (2016) 3–14.
[28] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: a review, Neurocomputing 187 (2016) 27–48.
[29] X. Chen, S. Xiang, C.-L. Liu, C.-H. Pan, Vehicle detection in satellite images by hybrid deep convolutional neural networks, IEEE Geosci. Remote Sens. Lett. 11 (2014) 1797–1801.
[30] I. Lenz, H. Lee, A. Saxena, Deep learning for detecting robotic grasps, Int. J. Robotics Res. 34 (2015) 705–724.
[31] M. Längkvist, L. Karlsson, A. Loutfi, A review of unsupervised feature learning and deep learning for time-series modeling, Pattern Recogn. Lett. 42 (2014) 11–24.
[32] S. Gao, Y. Zhang, K. Jia, J. Lu, Y. Zhang, Single sample face recognition via learning deep supervised autoencoders, IEEE Trans. Inf. Forensics Secur. 10 (2015) 2108–2118.
[33] S.E. Kahou, X. Bouthillier, P. Lamblin, C. Gulcehre, V. Michalski, K. Konda, et al., Emonets: multimodal deep learning approaches for emotion recognition in video, J. Multimodal User Interfaces 10 (2016) 99–111.
[34] S. Lawrence, C.L. Giles, A.C. Tsoi, A.D. Back, Face recognition: a convolutional neural-network approach, IEEE Trans. Neural Networks 8 (1997) 98–113.
[35] S. Chen, H. Xu, D. Liu, B. Hu, H. Wang, A vision of IoT: applications, challenges, and opportunities with china perspective, IEEE Internet Things J. 1 (2014) 349–359.
[36] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, European conference on computer vision, Springer, 2014, p. 818-33.
[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al. Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, p. 1-9.
[38] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:14091556, 2014.
[39] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, p. 770-8.
[40] J. Cong, B. Xiao, Minimizing computation in convolutional neural networks, International conference on artificial neural networks, Springer, 2014, p. 281-90.
[41] P. Nguyen, T. Tran, N. Wickramasinghe, S. Venkatesh, $\mathtt {Deepr}$: a convolutional net for medical records, IEEE J. Biomed. Health. Inf. 21 (2017) 22–30.
[42] Y. Cheng, F. Wang, P. Zhang, J. Hu, Risk prediction with electronic health records: A deep learning approach, in: Proceedings of the 2016 SIAM International Conference on Data Mining, SIAM, 2016, pp. 432–440.
[43] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[44] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modelling, arXiv preprint arXiv:14123555, 2014.
[45] J.A. Fries, Brundlefly at SemEval-2016 Task 12: Recurrent neural networks vs. joint inference for clinical temporal information extraction. arXiv preprint arXiv: 160601433, 2016.
[46] W. Sun, B. Zheng, W. Qian, Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis, Comput. Biol. Med. 89 (2017) 530–539.
[47] Q. Zhang, Y. Xiao, W. Dai, J. Suo, C. Wang, J. Shi, et al., Deep learning based classification of breast tumors with shear-wave elastography, Ultrasonics 72 (2016) 150–157.

[48] D. Wulsin, J. Gupta, R. Mani, J. Blanco, B. Litt, Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement, J. Neural Eng. 8 (2011) 036015.
[49] J. Patterson, A. Gibson, Deep Learning: A Practitioner's Approach, O'Reilly Media, Inc, 2017.
[50] S. Vieira, W.H. Pinaya, A. Mechelli, Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications, Neurosci. Biobehav. Rev. 74 (2017) 58–75.
[51] K. Mannepalli, P.N. Sastry, M. Suman, A novel adaptive fractional deep belief networks for speaker emotion recognition, Alexandria Eng. J. 56 (2017) 485–497.
[52] M.M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, R.R. Yager, Deep learning approach for active classification of electrocardiogram signals, Inf. Sci. 345 (2016) 340–354.
[53] A. Saeed, S. Trajanovski, M. Van Keulen, J. Van Erp, Deep physiological arousal detection in a driving simulator using wearable sensors, in: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), IEEE, 2017, pp. 486–493.
[54] Y.R. Tabar, U. Halici, A novel deep learning approach for classification of EEG motor imagery signals, J. Neural Eng. 14 (2016) 016003.
[55] N. Wahab, A. Khan, Y.S. Lee, Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection, Comput. Biol. Med. 85 (2017) 86–97.
[56] Z. Yin, J. Zhang, Cross-session classification of mental workload levels using EEG and an adaptive deep learning model, Biomed. Signal Process. Control 33 (2017) 30–47.
[57] Y. Yoo, L.Y. Tang, D.K. Li, L. Metz, S. Kolind, A.L. Traboulsee, et al., Deep learning of brain lesion patterns and user-defined clinical and MRI features for predicting conversion to multiple sclerosis from clinically isolated syndrome, Computer Methods Biomech. Biomed. Engineering: Imaging Visualization (2017) 1–10.
[58] Y. Yoo, L.Y. Tang, T. Brosch, D.K. Li, S. Kolind, I. Vavasour, et al., Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls, NeuroImage: Clinical. 17 (2018) 169–178.
[59] N. Dhungel, G. Carneiro, A.P. Bradley, A deep learning approach for the analysis of masses in mammograms with minimal user intervention, Med. Image Anal. 37 (2017) 114–128.
[60] A. Dubrovina, P. Kisilev, B. Ginsburg, S. Hashoul, R. Kimmel, Computational mammography using deep neural networks, Computer Methods Biomech. Biomed. Eng.: Imaging Visualization (2016) 1–5.
[61] Z. Jiao, X. Gao, Y. Wang, J. Li, A parasitic metric learning net for breast mass classification based on mammography, Pattern Recogn. 75 (2018) 292–301.
[62] R.K. Samala, H.P. Chan, L. Hadjiiski, M.A. Helvie, J. Wei, K. Cha, Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography, Med. Phys. 43 (2016) 6654–6666.
[63] P. Moeskops, J. de Bresser, H.J. Kuijf, A.M. Mendrik, G.J. Biessels, J.P. Pluim, et al., Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI, NeuroImage: Clinical. 17 (2018) 251–262.
[64] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, T.-S. Chua, Disease inference from health-related questions via sparse deep learning, IEEE Trans. Knowl. Data Eng. 27 (2015) 2107–2119.
[65] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, Y. Fan, A deep learning model integrating FCNNs and CRFs for brain tumor segmentation, Med. Image Anal. 43 (2018) 98–111.
[66] Y. Xiao, J. Wu, Z. Lin, X. Zhao, A deep learning-based multi-model ensemble method for cancer prediction, Comput. Methods Programs Biomed. 153 (2018) 1–9.
[67] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, P. Hufnagl, Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology, Comput. Med. Imaging Graph. (2017).
[68] W. Sun, B. Zheng, W. Qian, Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis, Comput. Biol. Med. (2017).
[69] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. (2017).
[70] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, J. Am. Med. Inform. Assoc. 24 (2016) 361–370.
[71] H. Choi, S. Ha, H.J. Im, S.H. Paek, D.S. Lee, Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging, NeuroImage: Clinical. 16 (2017) 586–594.
[72] H. Gunduz, Deep learning-based Parkinson's disease classification using vocal feature sets, IEEE Access 7 (2019) 115540–115551.
[73] R. Miotto, L. Li, B.A. Kidd, J.T. Dudley, Deep patient: an unsupervised representation to predict the future of patients from the electronic health records, Sci. Rep. 6 (2016) 26094.
[74] J.I. Orlando, E. Prokofyeva, M. del Fresno, M.B. Blaschko, An ensemble deep learning based approach for red lesion detection in fundus images, Comput. Methods Programs Biomed. 153 (2018) 115–127.
[75] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. 100 (2018) 270–278.
[76] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, J. Am. Med. Inform. Assoc. 24 (2017) 361–370.
[77] R. Miotto, L. Li, B.A. Kidd, Dudley JTJSr. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records, 6 (2016) 1-10.
[78] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, Chua T-SJITok, et al., Disease inference from health-related questions via sparse deep learning. 27 (2015) 2107-2119.
[79] J.I. Orlando, E. Prokofyeva, M. del Fresno, M.B.J.C.m. Blaschko, biomedicine pi An ensemble deep learning based approach for red lesion detection in fundus images 153 (2018)115-127.
[80] M.M.A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, R.R. Yager, Deep learning approach for active classification of electrocardiogram signals, Inf. Sci. 345 (2016) 340–354.
[81] R.K. Samala, H.P. Chan, L. Hadjiiski, M.A. Helvie, J. Wei, K.J.M.p. Cha, Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. 2016 (43) 6654-6666.
[82] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, P. Hufnagl, Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology, Comput. Med. Imaging Graph. 61 (2017) 2–13.
[83] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, P.J.C.M.I. Hufnagl, Graphics. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology61 (2017) 2-13.
[84] W. Sun, B. Zheng, W.J.C.i.b. Qian, Medicine. Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis, 89 (2017) 530-9.
[85] Z. Yin, J.J.B.S.P. Zhang, Control. Cross-session classification of mental workload levels using EEG and an adaptive deep learning model. 33 (2017) 30-47.
[86] Y. Yoo, L.Y.W. Tang, D.K.B. Li, L. Metz, S. Kolind, A.L. Traboulsee, et al., Deep learning of brain lesion patterns and user-defined clinical and MRI features for predicting conversion to multiple sclerosis from clinically isolated syndrome, Comput. Methods Biomech. Biomed. Eng.: Imaging Visualization 7 (2019) 250–259.
[87] Y. Yoo, L.Y. Tang, D.K. Li, L. Metz, S. Kolind, A.L. Traboulsee, et al., Deep learning of brain lesion patterns and user-defined clinical and MRI features for predicting conversion to multiple sclerosis from clinically isolated syndrome 7 (2019) 250-259.
[88] Q. Zhang, Y. Xiao, W. Dai, J. Suo, C. Wang, J. Shi et al., Deep learning based classification of breast tumors with shear-wave elastography, 72 (2016) 150-157.
[89] F. Firouzi, A.M. Rahmani, K. Mankodiya, M. Badaroglu, G.V. Merrett, P. Wong, et al., Internet-of-Things and big data for smarter healthcare: from device to architecture, applications and analytics, Elsevier, 2018.
[90] M. Shafique, T. Theocharides, C.-S. Bouganis, M.A. Hanif, F. Khalid, R. Hafız, et al., An overview of next-generation architectures for machine learning: Roadmap, opportunities and challenges in the IoT era. Design, Automation & Test in Europe Conference & Exhibition (DATE), IEEE, 2018, p. 827-32.