A Survey On The Role Of Deep Learning In 2d Transthoracic Echocardiography

Vishal Chandra, Vinay Singh, Prattay Guha Sarkar

Abstract: Deep learning technology is currently a most effective choice for medical image analysis, due to revolution in it, cardiovascular imaging has changed rapidly with huge impact. This article will elucidate the current application, role, challenges, and limitations in 2D echocardiography. It is beneficial to track developments in this technology to make track; it has a actual significant impact on medical practices for medical professionals. This review is a stepping stone in the contribution of deep learning in echocardiography. In this paper only most impact full deep learning models and processes are explained, and also a concise overview of the DL process is provided. In this review, only echocardiography modality has covered with a review of more than sixty papers.

Keywords: heart disease, assessment, automation, deep learning technique, transthoracic echocardiography

1 INTRODUCTION

In the past four years, cardiovascular diseases are the main cause of deaths 29% in India [1], 30% in the United States [2] and 45% in the European Union [3]. Due to the emergence of deep learning techniques, a stream in computer science called computer vision, mainly in object detection, segmentation, image classification has a huge impact [4-5]. Medical imaging is not apart from it and influenced significantly by it [6-7]. Due to the error-prone nature of disease diagnosis by a physician, there is a need for automated medical procedures ranging from disease diagnosis to treatment. Previous expert systems which are based on rules are proven inefficient, because they required manual feature engineering and also required vast domain knowledge to achieving significant accuracy. Machine learning computers are allow to learn from given data, not by programmed explicitly it came up with an effective and efficient way of combining and using various medical resources images, health records, pathological samples data to enhance the accuracy and reliability of various medical tasks. Contribution of machine learning is to deals with medical data, analyzing them at population-level data to concluding the condition of individual patients. Machine learning techniques provide capability of a system to work autonomously by acquiring knowledge through extracting features from large data sets [8]. This technology giving tremendous impact in all sectors of technology and science with the application in industry, autonomous vehicle driving, and recognition, spam filters, sentiment analysis, chat-bots, speech analysis and speech processing. In medical imaging, ML has demonstrated as effective as a human sonographer in validating presumptive diagnoses [9]. Machine learning’s special subdomain is Deep Neural Networks that is a well-defined structured collection of multiple layers of neurons, a neuron is the smallest unit of it. Another popular kind is the convolutional neural network. Availability of data in bulk and high-performance graphics processing units, deep learning has proven, it is the most robust solution for not only computer vision tasks like image classification [4], image segmentation [10] but also natural language processing [11][12] and genomics [13]. Advantages of DL over other traditional machine learning algorithms, to solve a particular problem it requires less domain knowledge the self-feature extracting nature makes it easy to use, higher accuracy can be achieved usually by either increasing the large number of data samples or the adjusting model parameters of the proposed network. Small learning models like Support Vector Machines (SVMs) and decision trees are inefficient, this is due to requirement of a large computation in both training and inference also required a huge number of domain knowledge for achieving generalization, it also involves significantly huge human labor for feature extraction manually for the prior domain knowledge in the model [14]. Since the easily availability of a huge amount of echocardiography images, the surge of publications that are based upon deep learning in cardiac imaging is flooded. In this review paper, our aim is to provide an introduction to basics concise neural networks, concepts deep learning techniques, and their applications and an review of deep learning application in cardiovascular image interpretation and analysis. Various reviews papers have already published dealing with the impact of deep learning on (parts of) cardiology and cardiovascular imaging (three papers). Their works by Slomka et al. [15] and Litjens et al. [7] are almost similar to the current survey. Slomka et al. [16] reviewed various cardiac imaging and discussed approaches for being autonomous, in which DL techniques played a significant role. Their detailed studies applied DL to echocardiography imaging were covered by Slomka et al. [16]. Litjens et al. [7], their detailed reviewed on deep learning applications to cardiac imaging, they only covered studies that are published before all most end of 2018, whereas various intensive works on cardiovascular imaging appeared also after that. This review article included more than 40 original research papers in which analysis and diagnosis are done by 2D transthoracic echocardiography with the of deep learning techniques. We included papers from the year 2015 to 2019 searched by keywords “Machine Learning” OR “Deep Learning” OR “2D Transthoracic Echocardiography” OR “Cardiac Imaging” on Google Scholar, PubMed and Scopus. Papers that only focused on TTE image analysis were selected by manually on the basis of title and abstract int this review. The structure of

• Vishal Chandra, Vinay Singh, Prattay Guha Sarkar
  Faculty of Computing and Information Technology, Usha Martin University, Ranchi Rajendra Institute of Medical Sciences, Ranchi
this review paper is, A short history of Cardiovascular Imaging for historical background on cardiovascular image analysis after that we introduced neural networks, deep learning methods. Basis of DL has presented with some important concepts and techniques in deep learning with their capabilities. In the later section, we presented a review of recent deep learning articles for functional, anatomical, and intraoperative cardiac imaging with only TTE modality. Finally, the discussion and conclusion section contains an overview of the current and future prospective of deep learning on cardiovascular image analysis and diagnosis in clinical practice.

History of cardiac imaging
Before the emergence of DL techniques, various different approaches and techniques had been developed to extract desirable clinical information from cardiac images. Early intelligent models required huge manual tuning for the conversion of the input image to the required output image in both training and inference [17]. Recent techniques have a higher level of Automation capability. Algorithms like level sets methods for segmentation can directly use image intensities [18] [19] and vessel centerline extraction uses minimum cost. Algorithms [20].

Other methods are based upon manual feature extraction from echocardiography image and input it to a statistical model for prediction of some diseases like support vector machine [21]. Extracted features describe those characteristics on which prediction depends such as shape, color and texture shape [22–24]. Recently, researches have changed the direction of image processing and interpretation in which thousands of features can be extracted automatically. The statistical classifier has to choose which features are relevant for the specific task. The common thing between these approaches, the features are designed by humans and also fed to a prediction model. Opposite to it, deep learning models capable to learn by automatically extracted of required features and to predict the output image from a given input, this is also called end-to-end learning. Most of works have focused on cardiac anatomical structures image segmentation for quantitative and visualization analysis. The success of traditional methods cannot be ruled out, they have huge contribution [25]. Few examples, determination of left ventricle ejection fraction by contouring left ventricle in short-axis MR image. For the detection of luminal stenosis, atherosclerotic plaque in vessel segmentation detection and detection technique for lumen [26]. Their results are promising but not robust for clinical practices.

Artificial NEURAL NETWORKS
ANNs are a special branch of machine learning technique it was initially inspired by human brain nerves which are the smallest components of human brain. They work on the principle of function approximation in which the x is input, they maybe a text, sound, image, signal, video, 3D volume or maybe combination of these and their output from the same set of input x but with a more meaningful desired information. In the mathematical context, the neural network objective is to identify the set of parameters \( \theta \) (where b biases and w weights):

\[
f(x; \theta) = y
\]

where \( f \) is an evaluation function and \( y \) is the prediction value based on provided input and evaluation function. \( J(\theta) \) is a cost function and \( \theta \) must be as lowest as possible it is the constraint with \( \theta \). The perceptron is the unit of artificial neural networks and perceptron structure as shown in Fig.1, first proposed by Rosenblatt [27]. An ANNs is well designed well-structured collection of a large number of specifically interconnected neurons. An example of an ANNs is shown in Figure 2. An artificial neuron may have a set of multiple inputs with a single output. Each input value is multiplied by some value, which is known as weight, and after that, all multiplied inputs are added. Most of the time some extra value is added to the previously added value, is known as bias. After that, the added value is forwarded via a nonlinear function, known as the activation function. The strength of ANNs depending upon the specifically well-designed collection of multiple numbers of neurons with relationships between the input value and output are highly nonlinear to get desire accurate predictions. The first application of ANNs in medical imaging was done in 1995[28], but they never gained the success as having now these days, due to unavailability of huge computation power, its bonded researchers to use only depth of few layers. To identify the relationships between image and its desire output requires large complex architecture of ANNs. In last few decades Development in computing technologies have been enabled researchers to build ANNs of thousands of neurons with complex architecture. These have enabled for direct operation on medical images without manual feature extraction. FCNs (Fully Connected Networks) are consist of multiple layers perceptions stacked in depth and width, which means every unit of neurons in each layer is directly and dedicatedly connected to every unit of neuron in the very next layers. Based on researches it has evidence, [29] that using only a single layer of FNNs with an adequate number of hidden layers of neurons can be used universal function approximators, but they are computationally inefficient for the
fitting of complex functions. Stacked Restricted Boltzmann Machines to form Deep Belief Networks in which each layer has encoded with statistical dependencies with units of the previous layer, for maximizing the likelihood in training data. An example of simple Convolutional Neural Networks (CNNs), model as shown in Fig. 3.

For analysis of the image dataset, the CNNs model with well-organized and well-defined architecture is required. Due to weight sharing, capability CNNs, they have a smaller number of weights than standard ANNs. The name convolutional neural network is due to its convolution operation for feature extraction. Convolutional neural networks are come with two types of layers convolutional layer and pooling layers. Pooling layers have no weights, but they Down-sample the image. In this process resolution of the image reduces but on the other hand in subsequent layers have increased field of view, this makes CNN integrate more contextual information. CNNs are widely used in image classification and regression with last fully connected layer. Most of convolutional neural networks are used for classification of image or regression with last layers as fully connected layers. These FCL are normal collection of neural networks as shown in figure 2 and used as a summarizing of all feature information into one prediction, which is good classifiers. Weights sharing feature of CNNs make much faster and efficient than FCL. Convolutional layer-based classifiers are VGG [30] is simplest one CNN architecture, having (3 x 3) convolutional filters, vgg16 and vgg19 are two versions of VGG. GoogleNet [31] is another CNN architecture uses the inception module. It has multiple parallel units of convolutional layers at each level result is produced by concatenation of its parallel layers, this makes the inception module to learn multiple level features. ResNet [32] is another CNN architecture has new kind of layer, having residual functions learning with reference to the input layer, it facilitates training with deeper networks. Autoencoders (AEs) are different architecture consists of two parts encoder and decoder, prime objective is to copy useful property from input data to output while training. Encoder acts as a down-sampler while Decoder acts as upsampler to the original image dimension. A simple and common Encoder-Decoder architecture is Stacked Denoised AE (SDAE), having objective to reconstruct original denoised clean image from noised corrupted image [33]. Another similar architecture is Unet [10], which is used for biomedical image segmentation. Unet used skip connections that concatenate the layers of the encoder with the corresponding to the decoder one. In contrast to previously described architectures, Recurrent Neural Networks (RNNs) have feedback loops that are used in the internal state to process the input. Due to vanishing gradients problem in RNNs Long-Short Term Memory (LSTM) was proposed which is capable to store information over long time.

Challenges in dealing with training deep models
For achieving the excellent training performance, a deep learning model requires a large number of labeled training data, thousands of labeled data. The requirement of large datasets is difficult in current medical imaging analysis; also, expert annotation of images is very much expensive and some diseases (congenital mitral stenosis) are rare. Thus, how to get excellent performance from a deep learning model with a limited number of training samples is still an unsolved question in medical imaging. The reason is model gets easily over-fit when it goes training with limited number of training data. Overfitting is a phenomenon in which model shows high training performance but when test with unknown data performance becomes poor. There are two ways to reduce the model overfitting problem, model optimization and transfer learning. There are few strategies which can be useful for model optimization such as proper kernel initialization, stochastic gradient descent (SGD) and its variants (e.g. Adagrad [34]), efficient activation functions (e.g. RELU), and also some powerful intermediate regularization techniques (e.g., batch normalization) have proposed and drastically improved in few recent years.

List as follows [8]:

1. Proper kernel initialization/momentum techniques [35] it is the utilization of initialization randomly and controlling the momentum parameter on iteration while training the model.
2. Use of efficient activation functions, like RELU [36,37], work as a nonlinear operation in the convolutional layer.
3. Dropout [38] is another technique which randomly deactivates the units/neurons in a network architecture at a particular rate (e.g., 0.5) on each iteration in the training process.
4. Batch normalization, it performs the normalization operation for each mini-batch training process

Supporting Hardware and Software
Easily and widespread availability of graphics processing units (GPUs) makes capable, deep learning so popular and easy to use both in sense of efficiency and affordability, Nvidia provides libraries for these GPUs, CUDA and OpenCL are easily available without any cost. Due to the highly parallel computing nature of GPUs, they have much more execution threads than CPUs. They are very fast computing devices, are up to 40 times faster than central processing units (CPUs). Apart from hardware, the availability of open-source packages is also having a key role in the success of deep learning technology. They are easily available, free of cost and have wide ranges. They enable GPUs for the efficient implementation of neural network operations, popular software packages are (in alphabetical order):

a) Caffe: having C++ and Python interfaces and developed UC Berkeley.
b) TensorFlow: it supports both C++ and Python, interfaces and developed by Google.
c) Torch: has Lua interface.
There are some third-party packages written on top of one or more of these frameworks, such as Lasagne (https://github.com/Lasagne/Lasagne) or Keras (https://keras.io/).

Deep learning in 2d Echocardiography

Echocardiography uses ultrasound waves to visualize the heart's internal structure, is an imaging modality. Deep learning applications for echocardiography consist of classification, detection, segmentation, report generation, and tracking. Most of the contributions by using deep learning are related to segmentation and detection. The most popular deep learning model is Unet[10] and its variants, DBNs are also used for it. In [39] the authors proposed a method with a DBN that models the visualization of the left ventricle showing, more robust than previous level sets and deformable techniques. Nascimento et al. [40] proposed a model that divides the data into patches that each patch proposes a segmentation of the left ventricle. The output image is obtained by merging patches by a DBN classifier in which each patch is assigned a weight. In this way, it produces more robust output with limited number of data set. In [41] the authors used a regularized FCN and compared with simple FCN and demonstrated better results. Deep learning has also used for echocardiographic viewpoint classification. Madani et al. [42] used a CNN with six-layer to classify between 15 views (12 videos and 3 still) of TTE, achieved better results than echocardiographers. In [43] the authors proposed a 3D residual CNN network for classification of ejection fraction classification using TTE images. In [44] the authors proposed a network for real-time echo quality scoring for getting feedback to reduce the variability while echocardiography process. They used recurrent layers to utilize the consecutive information in the echo loop. Perrin et al. [45] trained AlexNet with 59151 echo frames to classify congenital heart disease between five pediatric populations. Moradi et al. [46] proposed a method which was based upon VGGnet and doc2vec [47] technique to produce semantic descriptors for echo images. Their model identified 91% of disease instances and 77% of valve disease severity. Moradi et al. [46] proposed a deep learning model for establishing relationship between echocardiography images and medical records. Chen et al. [41] proposed a model capable of doing segmentation of the left ventricle in 5 different 2D views (apical, 2, 3, 4, and 5-chamber view). Carneiro et al. [39] did three studies for left ventricle segmentation, the first study [39] used an ANN model to predict landmarks. The second was extended in Carneiro and Nascimento [40] for tracking of the left ventricle. Girgis et al. combined CNN with optical flow methods for segmentation left ventricle [7]. Behnami [?] proposed a model of RNNs for detection risk of systolic cardiac failure. Costa et al.[50] used Unet for segmentation mitral leaflets by training with 30 videos. Kusunose at al compared various image classification models for classification of regional wall motion abnormalities [51], Kwon [52] used deep learning for prediction of in-hospital mortality based upon echocardiography images. Leclerc et al. [53] used open large dataset for segmentation of left ventricle. Dehghan et al[54] used Unet for detection of anomalies by multi-view regression. Moradi [55] used modified Unet for left ventricle segmentation better than any previous method. Hanif bin et al [56] proposed a method for detection of aortic valve. Omar used CNN for classification of wall motion abnormality. Smistad used convolutional neural network for real-time view classification in TTE. Smistad et al. [57] used modified Unet for LV segmentation. Veni et al[58] combine deformable model with CNN for segmentation of left ventricle. Dong et al. [59] proposed combined traditional techniques with CNN for left ventricle segmentation. Studies are also covered direct disease classification by analyzing echocardiography images.

2 DISCUSSION, CONCLUSION AND FUTURE SCOPE

On the basis of literature, it is evident that the traditional rule-based expert system and machine learning techniques will replace by deep learning methods for image analysis, which are based upon the manual feature extraction method. In [60] the authors claimed that deep learning methods are better in the visualization of complex patterns in high dimensional medical image data. In [61] this paper author claimed that AI techniques are better than human analytical capacities in prediction of medical risk identification. Deep learning may able to reduce processing time and hence improve the quality of patient care. For the achievement of the high-quality result, deep learning requires large datasets for training [3]. This is very difficult when dealing with medical data, due to need of medical expert labor, it is time-consuming and costly second-most of cases are normal this makes dataset unbalance. In [62] the authors argued that additional data with uncontrolled clinical settings and multi-center required for validation of these applications before implementation in clinical use. Literature shows that unsupervised learning is not regularly used. Although Hinton [63] that there is a lack of interpretability in this method, interpretation of nonlinear features infeasible because of dependencies of features from other layers differ from others. The main reason for resist medical experts' use of these models is non-interpretability [64]. Deep learning models are stochastic, means every time a network fits with the same data but result with different initial weights also different features are learned, it is completely “black box” [65], it argues about ethical and legal issues using it in clinical practices. As far as non-interpretability problem is concerned researchers should build simple deep learning models having nonensemble and end-to-end functionality. The usefulness of most of models proposed by their respective authors is unclear, due to lack of external validation and methodological errors. Before building new predictive models for risk prediction in the future, researchers should try to analyze existing models, they can be improved or not. However a deep learning architecture capsule
networks have not been widely used in cardiac imaging which is more interpretable even able to recognize pose, requires less training data. The main problem with this type of networks, they require high computational cost that is not easily accessible.

REFERENCES

[12]. “SPEECH RECOGNITION WITH DEEP RECURRENT NEURAL NETWORKS Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton Department of Computer Science, University of Toronto,” no. 6, pp. 6645–6649, 2013.


