Deep Learning Applications in Medical Image Analysis

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

SREELAKSHMI PRASANTH

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

Deep Learning Applications in Medical Image Analysis

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

SREELAKSHMI PRASANTH

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MAR ATHANASIUS COLLEGE OF ENGINEERING
KOTHAMANGALAM

CERTIFICATE

This is to certify that the report entitled **Deep Learning Applications in Medical Image Analysis** *submitted by Ms.***SREELAKSHMI PRASANTH**,Reg.No. **MAC15CS057** *towards partial fulfillment of the requirement for the award of Degree of Bachelor of Technology in Computer science and Engineering from APJ Abdul Kalam Technological University for December 2018 is a bonafied record of the seminar carried out by her under our supervision and guidance.*

....	....	....
Prof. Joby George	**Prof. Neethu Subash**	**Dr. Surekha Mariam Varghese**
Faculty Guide	***Faculty Guide***	***Head of the Department***

Date: Dept. Seal

ACKNOWLEDGEMENT

First and foremost, I sincerely thank the God Almighty for his grace for the successful and timely completion of the seminar.

I express my sincere gratitude and thanks to Dr. Soosan George.T, Principal and Dr. Surekha Mariam Varghese, Head Of the Department for providing the necessary facilities and their encouragement and support.

I owe special thanks to the staff-in-charge Mrs.Neethu subhash , Mr.Joby George and Mrs.Joby Anu Mathew for their corrections, suggestions and sincere efforts to co-ordinate the seminar under a tight schedule.

I express my sincere thanks to staff members in the department of Computer Science and Engineering who have taken sincere efforts in helping me to conduct this seminar.

Finally, I would like to acknowledge the heartfelt efforts, comments,criticisms, co-operation and tremendous support given to me by my dear friends during the preparation of the seminar and also during the presentation without whose support this work would have been all the more difficult to accomplish.

ABSTRACT

Healthcare sector is totally different from other industry. People expect highest level of care and service regardless of cost. Medical images are an integral part of a patients electronic health record. The interpretations of medical data are done by medical experts. A delayed or erroneous diagnosis cause harm to the patient.It is ideal for medical image analysis to be carried out by an automated and efcient machine learning algorithm. Deep learning is emerged as a technology to enhance the performance of existing machine learning techniques.Supervised and unsupervised learning are the techniques used for analysis of images.Convolutional Neural Networks are the most popular machine learning algorithm in image classification. Machine learning algorithms provide exciting solutions in many elds. It is seen as a key method for various future applications in health sector.

Contents

List of Figures

Figure No. Name of Figures Page No.

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
CT	Computed Tomography
DL	Deep Learning
DNA	Deoxyribonucleic acid
DNN	Deep Neural Network
EHR	Electronic Health Record
EMR	Electronic Medical Record
GAN	General Adversarial Networks
ML	Machine Learning
MLP	Multilayer Perceptron

MRI	Magnetic Resonance imaging
RBM	Restricted Boltzman Machine
RELU	Rectified Linear Unit Layer
RNN	Recurrent Neural Networks

Introduction

Deep learning algorithms, in particular convolutional networks, have rapidly become a methodology of choice for analyzing medical images.Techniques of Machine Learning (ML) and Artificial Intelligence (AI) have played important role in medical field like medical image processing, computer-aided diagnosis, image interpretation, image fusion, image registration, image segmentation.Techniques of ML extract information from the images and represents information eectively and eciently. The ML and AI facilitate and assist doctors that they can diagnose and predict accurate and faster the risk of diseases and prevent them in time. These techniques enhance the abilities of doctors and researchers to understand that how to analyze the generic variations which will lead to disease. Machine learning is defined as a set of methods that automatically detect patterns in data, and then utilize the uncovered patterns to predict future data or enable decision making under uncertain conditions. Deep learning is a part of ML and a special type of artificial neural network (ANN) that resembles the multilayered human cognition system. Deep learning is currently gaining a lot of attention for its utilization with big healthcare data.

The success of machine learning algorithms at computer vision tasks in recent years comes at an opportune time when medical records are increasingly digitalized. The use of electronic health records (EHR) quadrupled from 11.8 to 39.6 percentage amongst office-based physicians in the US from 2007 to 2012. Medical images are an integral part of a patients EHR.

Modalities of digital medical images include ultrasound (US), X-ray,computed tomography (CT) scans and magnetic- resonance imaging (MRI) scans, positron emissiontomography(PET)scans,retinal photography,histology slides, and dermoscopy images. Some of these modalities examine multiple organs (such as CT, MRI) while others are organ specic (retinal photography, dermoscopy). The amount of data generated from each study also varies. A histology slide is an image le of a few megabytes while a single MRI may be a few hundred megabytes. This has technical implications on the way the data is pre-processed, and on the design of an algorithms architecture, in the context of processor and memory limitations.

Medical image processing refers to a set of procedures to obtain clinically meaningful information from various imaging modalities, mostly for diagnosis or prognosis. The extracted information could be used to enhance diagnosis and prognosis according to the patients needs. Distinct medical imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography, could provide distinct information for the patient being imaged. Structural and functional information could be extracted as necessary, and these are used as quantitative features for future diagnosis and prognosis. Research in medical image processing typically aims to extract features that might be difficult to assess with the naked eye. There are two types of features. The first is the well-known semantic feature defined by human experts, and the other is the agonistic feature defined by mathematical equations. The agonistic features suffer from less operator bias than the semantic features. Still, semantic features are well recognized in radiology research, which is an accumulation of years of human expertise. However, many semantic features are time-consuming to compute and sometimes there are inconsistencies among experts. The extracted agonistic features might be used as imaging biomarkers to explain various states of the patient. A recent research approach known as the radiomics approach employs hundreds or thousands of agnostic features to obtain clinically relevant information for diagnosis and prognosis.

Machine learning approaches are applied to associate imaging features obtained from medical image processing with relevant clinical information. Machine learning started as a field in computer science to endow algorithms to solve problems without being explicitly programmed. It typically learns representations from training data, which are generalized in separate test data. The technology has been applied to computer-aided diagnosis and computer-aided detection in medical imaging . The CADx system can identify a disease-related region and quantify the properties of that region, which could be used as a guidance for surgeons and radiologists. Despite their usefulness in analyzing medical imaging, machine learning approaches have several limitations. They show excellent performance when applied to training data but typically suffer losses in performances when applied to independent validation data. This is partly due to the overfitting of the training data. Performance of machine learning techniques must be evaluated with both training and independent validation data. Many machine learning studies have demonstrated great technical potential but only a few have shown actual clinical efficacy including gains in survival. Machine learning techniques also have issues related to feature definition. For example, they rely on a pre-defined set of features. Additionally, sometimes the features are difficult to define for a given problem. Researchers need to choose from different combinations of features, algorithms, and degrees of complexity to sufficiently solve a given problem, and many studies depend on trial-and-error to find the right combination. A major challenge in particular is choosing the right features to correctly model a given problem.

The development of an artificial neural network (ANN) largely circumvents this problem by learning feature representation directly from the raw input data, skipping the feature extraction procedure [2]. ANN attempts to mimic human brain processes using a network composed of interconnected nodes. The initial ANN model was a simple feed-forward network known as perceptron that could perform linear classification tasks. The early perceptron models required complex computation power beyond what was typically available at that time. Multilayer perceptron was proposed to improve the simple perceptron model by adding hidden layers and developing learning techniques, such as back-propagation. MLP formed the basis for the modern DL approaches. The deep portion of DL refers to having many layers whose structures are suitable to model big data. On a practical side, the DL approach requires a high computational load. With the recent developments of computational infrastructure, such as graphical processing units and cloud computing systems, DL has become practical and has achieved groundbreaking results in many fields.

DL has been successfully applied to many research fields and has become ubiquitous. As such, many studies have already adopted DL to improve medical imaging research, and increasingly more studies will adopt DL for medical imaging research in the future.

Existing System

Medical images are an integral part of a patients EHR. An electronic health record (EHR) contains patient health information, such as:Administrative and billing data,Patient demographics,Progress notes,Vital signs,Medical histories,Diagnoses Medications,Immunization dates,Allergies,Radiology images,Lab and test results.These are currently analyzed by human radiologists, who are limited by speed, fatigue,and experience.It takes years and great nancial cost to train a qualied radiologist, and some health-care systems. outsource radiology reporting to lower-cost countries such as India via tele-radiology. A delayed or erroneous diagnosis causes harm to the patient. Therefore, it is ideal for medical image analysis to be carried out by an automated, accurate and efcient machine learning algorithm.

Medical image analysis is an active eld of research for machine learning, partly because the data is relatively structured and labelled, and it is likely that this will be the area where patients rst interact with functioning, practical articial in telligence systems. This is signicant for two reasons. Firstly, in- terms of actual patient metrics, medical image analysis is a litmus test as to whether articial intelligence systems will actually improve patient outcomes and survival. Secondly, it provides a testbed for human-AI interaction, of how receptive patients will be towards health altering choices being made, or assisted by a non-human actor.

The increased portability and accessibility of electronic medical records may increase the ease with which they can be accessed and stolen by unauthorized persons or unscrupulous users versus paper medical records, as acknowledged by the increased security requirements for electronic medical records included in the Health Information and Accessibility Act and by large-scale breaches in confidential records reported by EMR users.

Handwritten paper medical records may be poorly legible, which can contribute to medical errors. Pre-printed forms, standardization of abbreviations and standards for penmanship were encouraged to improve reliability of paper medical records. Electronic records may help with the standardization of forms, terminology and data input. Digitization of forms facilitates the collection of data for epidemiology and clinical studies.

EMRs can be continuously updated (within certain legal limitations). If the ability to exchange records between different EMR systems were perfected(”interoperability”) it would facilitate the co-ordination of health care delivery in non-affiliated health care facilities. In addition, data from an electronic system can be used anonymously for statistical reporting in matters such as quality improvement, resource management and public health communicable disease surveillance.

Medical imaging is the technique and process of creating visual representations of the interior of a body for clinical analysis and medical intervention, as well as visual representation of the function of some organs or tissues (physiology). Medical imaging seeks to reveal internal structures hidden by the skin and bones, as well as to diagnose and treat disease. Medical imaging also establishes a database of normal anatomy and physiology to make it possible to identify abnormalities. Although imaging of removed organs and tissues can be performed for medical reasons, such procedures are usually considered part of pathology instead of medical imaging.As a discipline and in its widest sense, it is part of biological imaging and incorporates radiology which uses the imaging technologies of X-ray radiography, magnetic resonance imaging, medical ultrasonography or ultrasound, endoscopy, elastography, tactile imaging, thermography, medical photography and nuclear medicine functional imaging techniques as positron emission tomography (PET) and Single-photon emission computed tomography (SPECT). Measurement and recording techniques which are not primarily designed to produce images, such as electroencephalography (EEG), magneto encephalography (MEG), electrocardiography (ECG), and others represent other technologies which produce data susceptible to representation as a parameter graph vs. time or maps which contain data about the measurement locations. In a limited comparison, these technologies can be considered as forms of medical imaging in another discipline.

Medical imaging is often perceived to designate the set of techniques that noninvasively produce images of the internal aspect of the body. In this restricted sense, medical imaging can be seen as the solution of mathematical inverse problems. This means that cause (the properties of living tissue) is inferred from effect (the observed signal). In the case of medical ultrasonography, the probe consists of ultrasonic pressure waves and echoes that go inside the tissue to show the internal structure. In the case of projectional radiography, the probe uses X-ray radiation, which is absorbed at different rates by different tissue types such as bone, muscle, and fat.

The results of a CT scan of the head are shown as successive transverse sections.An MRI machine generates a magnetic field around a patient.PET scans use radio pharmaceuticals to create images of active blood flow and physiologic activity of the organ or organs being targeted.Ultrasound technology is used to monitor pregnancies because it is the least invasive of imaging techniques and uses no electromagnetic radiation.In the clinical context, ”invisible light” medical imaging is generally equated to radiology or ”clinical imaging” and the medical practitioner responsible for interpreting (and sometimes acquiring) the images is a radiologist. ”Visible light” medical imaging involves digital video or still pictures that can be seen without special equipment. Dermatology and wound care are two modalities that use visible light imagery. Diagnostic radiography designates the technical aspects of medical imaging and in particular the acquisition of medical images. The radiographer or radiologic technologist is usually responsible for acquiring medical images of diagnostic quality, although some radiological interventions are performed by radiologists.

As a field of scientific investigation, medical imaging constitutes a sub-discipline of biomedical engineering, medical physics or medicine depending on the context: Research and development in the area of instrumentation, image acquisition , modeling and quantification are usually the preserve of biomedical engineering, medical physics, and computer science; Research into the application and interpretation of medical images is usually the preserve of radiology and the medical sub-discipline relevant to medical condition or area of medical science (neuroscience, cardiology, psychiatry, etc.) under investigation. Many of the techniques developed for medical imaging also have scientific and industrial applications.

Proposed System

Deep Learning for Medical Imaging Automatic medical imaging analysis is crucial to modern medicine.Diagnosis based on the interpretation of images can be highly subjective.Computer aided diagnosis(CAD) can provide an objective assessment of the underlying disease processes. Modeling of disease progression, common in several neurological conditions, such as Alzheimers, multiple sclerosis, and stroke, requires analysis of brain scans based on multimodal data and detailed maps of brain regions.

In recent years,CNNs have been adapted rapidly by the medical imaging research community because of their outstanding performance demonstrated in computer vision and their ability to be parallelized with GPUs. The fact that CNNs in medical imaging have yielded promising results have also been highlighted in are cent survey of CNN approaches in brain pathology segmentation and in an editorial of deep learning techniques in computer aided detection, segmentation, and shape analysis.

Among the biggest challenges in CAD are the differences in shape and intensity of tumors/lesions and the variations in imaging protocol even within the same imaging modality. In several cases, the intensity range of pathological tissue may overlap with that of healthy samples. Furthermore, Rician noise, nonisotropic resolution,and bias eld effects in magnetic resonance images (MRI) cannot be handled automatically using simpler machine learning approaches.To deal with this data complexity, hand-designed features are extracted and conventional machine learning approaches are trained to classify them in a completely separate step.

Deeplearning provides the possibility to automate and merge the extraction of relevant features with the classication procedure. CNNs inherently learn a hierarchy of increasingly more complex features and, thus, they can operate directly on a patch of images centered on the abnormal tissue. Example applications of CNNs in medical imaging include the classication of interstitial lung diseases based on computed tomography (CT) images, the classication of tuberculosis manifestation based on X-ray images, the classication of neural progenitor cells from somatic cell source, the detection of haemorrhages in color fundus images and the organ or body-part-specic anatomical classication of CT images. A body-part recognition system is also presented in Yan et al.A multistage deep learning framework based on CNNs extracts both the patches with the most as well as least discriminative local patches in the pretraining stage. Subsequently, a boosting stage exploits this local information to improve performance. The authors point out that training based on discriminative local appearances are more accurate compared to the usage of global image context.CNNs have also been proposed for the segmentation of intense stage brain tissues and brain extraction from multimodality MR images.

Hybrid approaches that combine CNNs with other architectures are also proposed. In [5], a deep learning algorithm is employed to encode the parameters of a deformable model and thus facilitate the segmentation of the left ventricle (LV) from short-axis cardiac MRI. CNNs are employed to automatically detect the LV, whereas deep Autoencoders are utilized to infer its shape. Yu et al designed a wireless capsule endoscopy classication system based on a hybrid CNN with extreme learning machine (ELM). The CNN constitutes a data-driven feature extractor, whereas the cascaded ELM acts as a strong classier.

ANN

Artificial neural networks structurally and conceptually inspired by human biological nervous system . Preceptron is one of the earliest neural network that was based on human brain system. It consist of input layer that is directly connect to output layer and was good to classify linearly separable patterns. To solve more complex pattern, neural network was introduced that has a layered architecture. Neural network consist of interconnected neurons that takes input and perform some processing on the input data, and finally forward the current layer output to the coming layer. Each neuron in the network sums up the input data and apply the activation function to the summed data and finally provides the output that might be propagated to the next layer. Thus adding more hidden layer allows to deal with complex as hidden layer capture nonlinear relationship. These neural networks are knows as Deep Neural network. Extra layers in DNN enable composition of features from lower layers to the upper layer by giving the potential of modeling complex data. Deep learning is the growing trend to develop automated applications and has been termed in 10 breakthrough technologies of 2013. Today, several deep learning based computer vision applications are performing even better than human i.e. identifying indicators for cancer in blood and tumors in MRI scans. It is improvement of artificial neural network that consist of more hidden layer that permits higher level of abstraction and improved image analysis. It becomes extensively applied method due to its recent unparallelled result for several applications i.e. object detection, speech recognition, face recognition and medical imaging.

ANN is a statistical machine learning method inspired by brain mechanism from neuroscience. Researchers designed a learning algorithm that resembles how the brain handles information. A neuron is the basic unit of the brain mechanism. The neuron is an electrically excitable cell that receives signals from other neurons, processes the received information, and transmits electrical and chemical signals to other neurons. The input signal to a given neuron needs to exceed a certain threshold for it be activated and further transmit a signal. The neurons are interconnected and form a network that collectively steers the brain mechanism. ANN is an abstraction of an interconnected network of neurons with layers of nodes, and it consists of an input layer aggregating the input signal from other connected neurons, a hidden layer responsible for training, and an output layer. Each node takes the input from nodes from the previous layer using various weights and computes the activation function, which is relayed onto the next layer of nodes. The activation function approximates the complex process of a physical neuron, which regulates the strength of the neuronal output in a non-linear manner.

3.1 Supervised Learning Model

Supervised learning is the Data mining task of inferring a function from labeled training data.The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value.

3.1.1 Convolutional Neural Networks

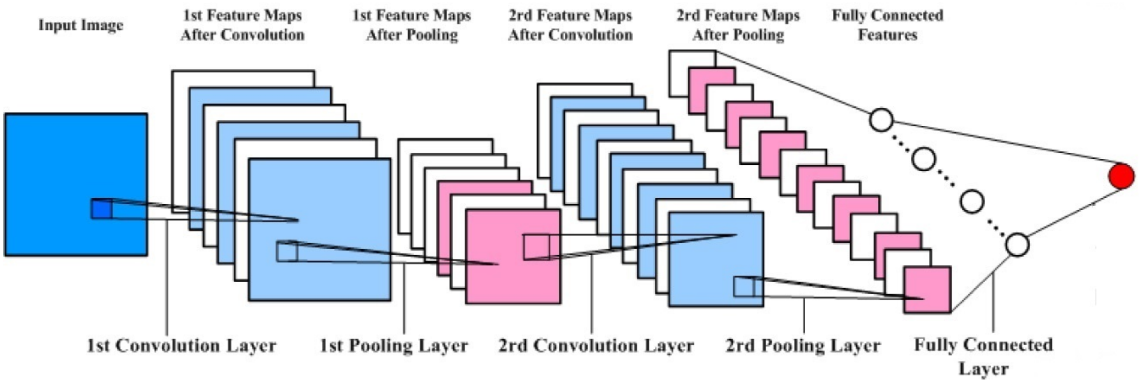

Fig. 3.1: Convolutional Neural Network

Both the 2-dimensional and 3-dimensional structures of an organ being studied are crucial in order to identify what is normal versus abnormal. By maintaining these local spatial relationships, CNNs are well-suited to perform image recognition tasks. CNNs have been put to work in many ways, including image classication, localization, detection, segmentation and registration. CNNs are the most popular machine learning algorithm in image recognition and visual learning tasks, due to its unique characteristic of preserving local image relations, while performing dimensionality reduction.This captures important feature relationships in an image (such as how pixels on an edge join to form a line), and reduces the number of parameters the algorithm has to compute,increasing computational efciency.CNNs are able to take as inputs and process both 2-dimensional images, as well as 3-dimensional images with minor modications. This is a useful advantage in designing a system for hospital use, as some modalities like X-rays are 2-dimensional while others like CT or MRI scans are 3-dimensional volumes. CNNs and Recurrent Neural Networks (RNNs) are examples of supervised machine learning algorithms, which require signicant amounts of training data. Unsupervised learning algorithms have also been studied for use in medical image analysis. These include Autoencoders, Restricted Boltzmann Machines (RBMs), Deep Belief Networks (DBNs), and Generative Adversarial Networks (GANs).

Currently, CNNs are the most researched machine learning algorithms in medical image analysis [4]. The reason for this is that CNNs preserve spatial relationships when ltering input images. As mentioned, spatial relationships are of crucial importance in radiology, for example, in how the edge of a bone joins with muscle, or where normal lung tissue interfaces with cancerous tissue. As shown in Fig.3.1, a CNN takes an input image of raw pixels, and transforms it via Convolutional Layers, Rectied Linear Unit (RELU) Layers and Pooling Layers. This feeds into a nal Fully Connected Layer which assigns class cores or probabilities, thus classifying the input into the class with the highest probability.

Convolutional Layer

A convolution is dened as an operation on two functions. In image analysis, one function consists of input values (e.g. pixel values) at a position in the image, and the second function is a lter (or kernel); each can be represented as array of numbers. Computing the dot product between the two functions gives an output. The lter is then shifted to the next position in the image as dened by the stride length. The computation is repeated until the entire image is covered, producing a feature (or activation) map (shown in fig.3.2).This is a map of where the lter is strongly activated andseesa feature such as a straight line,a dot,or a curved edge.If a photograph of a face was fed into a CNN, initially low-level features such as lines and edges are discovered by the lters. These build up to progressively higher features in subsequent layers,such as a nose, eye or ear, as the feature maps become inputs for the next layer in the CNN architecture.

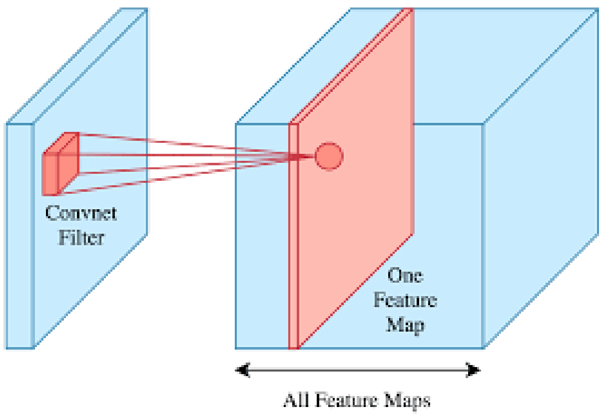

Fig. 3.2: Convolutional Layer

Convolution exploits three ideas intrinsic to perform computationally efcient machine learning: sparse connections, parameter sharing (or weights sharing) and equivariant representation. Unlike some neural networks where every input neuron is connected to every output neuron in the subsequent layer, CNN neurons have sparse connections, meaning that only some inputs are connected to the next layer. By having a small, local receptive eld (i.e., the area covered by the lter per stride), meaningful features can be gradually learnt, and the number of weights to be calculated can be drastically reduced, increasing the algorithms efciency. In using each lter with its xed weights across different positions of the entire image, CNNs reduce memory storage requirements. This is known as parameter sharing. This is in contrast to a fully connected neural network where the weights between layers are more numerous, used once and then discarded. Parameter sharing results in the quality of equivariant representation to arise. This means that input translations result in a corresponding feature map translation.

Rectified Linear Unit Layer

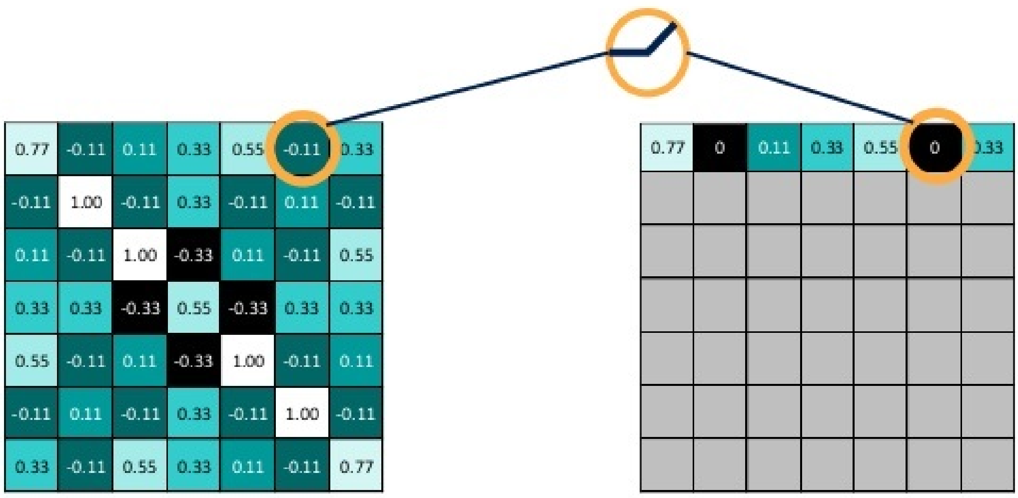

Fig. 3.3: Rectified Linear Unit Layer

The RELU layer is an activation function that sets negative input values to zero (shown in figure 3.3). This simplies and accelerates calculations and training, and helps to avoid the vanishing gradient problem. Mathematically it is dened as: f(x)=max(0,x) where x is the input to the neuron. Other activation functions include the sigmoid, tanh, leaky RELUs, Randomized RELUs and parametric RELUs.

Pooling Layer

The Pooling layer is inserted between the Convolution and RELU layers to reduce the number of parameters to be calculated, as well as the size of the image (width and height, but not depth). Max-pooling is most commonly used; other pooling layers include Average pooling and L2-normalization pooling. Max-pooling simply takes the largest input value within a lter and discards the other values;effectively it summarizes the strongest activations over a neighborhood.Max-Pooling is shown in figure 3.4 The rationale is that the relative location of a strongly activated feature to another is more important than its exact location. Convolutional neural network (CNN) architectures are motivated by the primary visual cortex in which there are layers of alternating simple and complex cells. Simple cells are specific to the stimuli position like a convolutional kernel while complex cells are less specific to the location of the stimuli just like the pooling operation.

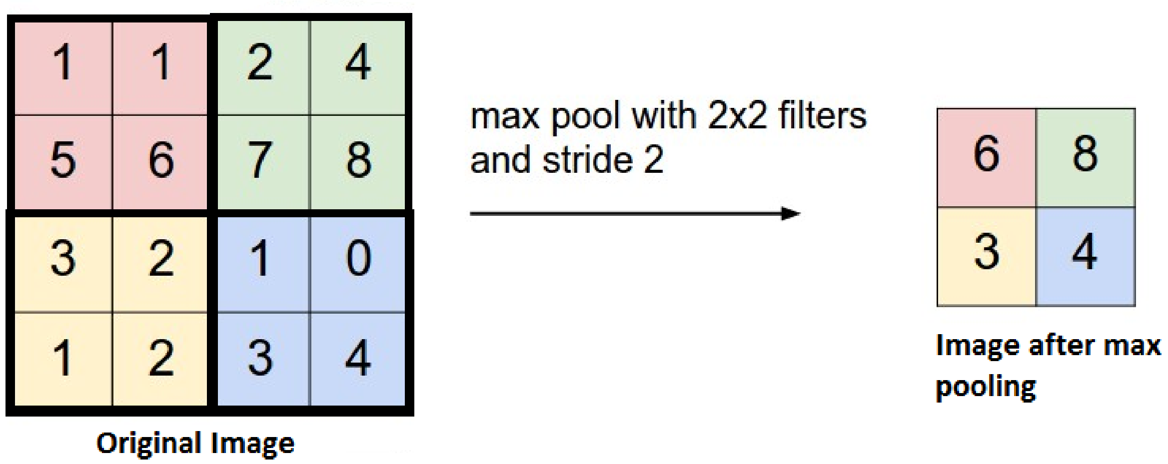

Fig. 3.4: Max-Pooling

Thus pooling was initially designed to help make CNN layers tolerate distortions, just like in the scale invariant feature transform (SIFT) descriptor with 44 sum pooling grid. Pooling allows features to shift relative to each other resulting in robust matching of features even in the presence of small distortions. There are also many other benefits of doing pooling, like:Reduces the spatial dimension of the feature map.And hence also reducing the number of parameters high up the processing hierarchy. This simplifies the overall model complexity.Though sum and max pooling are a bit outdated as mostly, nowadays, strided convolution is used. The purpose of strided convolution is to skip some areas during the convolution operation thereby resulting in:Efficient convolution operation and reduced spatial dimension of the output.

Pooling layers have two big objectives:1) Reduce the computational cost. A pooling layer can reduce the size of the previous layer without any additional parameters, so just shrink your data without adding any cost to the training.1x1 pooling layers are particuarly useful to reduce the size of the previous step specially if you plan to try different convolutions at the same time over the same previous layer, without pooling the number of parameters would be just too large to handle.2) Make the network more generic. Pooling helps the layer generalize because it effectively combines several values into a single one, this decreases the chance of overfitting because something very particular will be lost in the pooling phase.

Fully Connected Layer

The nal layer in a CNN is the Fully Connected Layer, meaning that every neuron in the preceding layer is connected to every neuron in the Fully Connected Layer. Like the convolution, RELU and pooling layers, there can be 1 or more fully connected layers depending on the level of feature abstraction desired. This layer takes the output from the preceding layer (Convolutional,RELU or Pooling) as its input,and computes a probability score for classication into the different available classes.Figure 3.5 is an example of classification of chest X-rays. In essence, this layer looks at the combination of the most strongly activated features that would indicate the image belongs to a particular class. For example, on histology glass slides, cancer cells have a high DNA to cytoplasm ratio compared to normal cells. If features of DNA were strongly detected from the preceding layer, the CNN would be more likely to predict the presence of cancer cells. Standard neural network training methods with backpropagation and stochastic gradient descent help the CNN learn important associations from training images.

The key challenges and limitations are: 1)CNNs are designed for 2-D images whereas

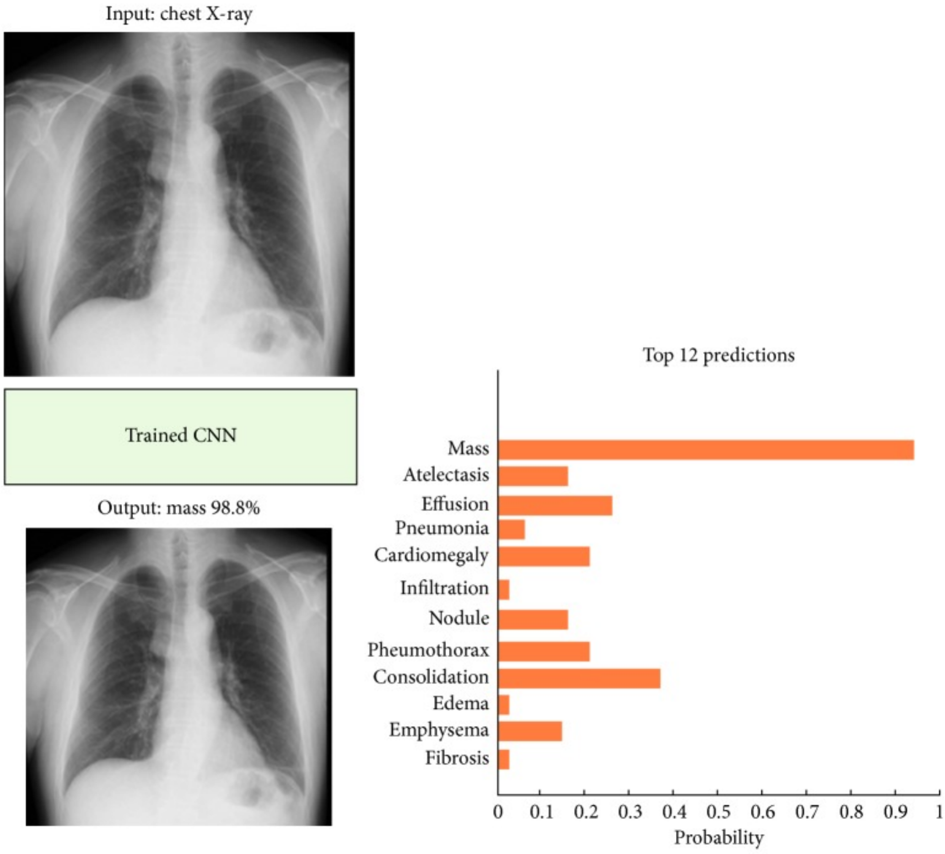

Fig. 3.5: CNN final classification of chest X-rays with classes probabilities

segmentation problems in MRI and CT are inherently 3-D. This problem is further complicated by the anisotropic voxel size. Although the creation of isotropic images by interpolating the data is a possibility, it can result in severely blurred images. Another solution is to train the CNNs on orthogonal patches extracted from axial, sagittal and coronal views. This approach also drastically reduces the time complexity required to process 3-D information and thus alleviates the problem of overtting.2) CNNs do not model spatial dependencies. Therefore, several approaches have incorporated voxel neighboring information either implicitly or by adding a pairwise term in the cost function, which is referred as conditional random field 3) Preprocessing to bring all subjects and imaging modalities to similar distribution is still a crucial step that affects the classication performance. Similarly to conventional machine learning approaches,balancing the datasets with bootstrapping and selecting samples with high entropy is advantageous.

All of these limitations result from or are exacerbated by small and incomplete training datasets.Furthermore,there is limited availability of ground-truth/annotated data,since the cost and time to collect and manually annotate medical images is prohibitively large. Manual annotations are subjective and highly variable across medical experts. Although, it is thought that the manual annotation would require highly specialized knowledge in medicine and medical imaging physics, recent studies suggest that nonprofessional users could perform similarly . Therefore, crowdsourcing is suggested as a viable alternative to create low-cost, big ground-truth medical imaging datasets. Moreover, the normal class is often over represented since the healthy tissue usually dominates and forms highly repetitive patterns. These issues result in slow convergence and overtting. To alleviate the lack of training samples, transfer learning via ne tuning have been suggested in medical imaging applications.In transfer learning via ne-tuning,a CNN is pretrained using a database of labeled natural images. The use of natural images to train CNNs in medical imaging is controversial because of the profound difference between natural and medical images. Nevertheless, Tajbakhsh et al showed that ne-tuned CNNs based on natural images are less prone to overtting due to the limited size training medical imaging sets and perform similarly or better than CNNs trained from scratch. Shin et al.has applied transfer learning from natural images in thoraco-abdominal lymph node detection and interstitial lung disease classication. They also reported better results than training the CNNs from scratch with more consistent performances of validation loss and accuracy traces. Chen et al. applied successfully a transfer learning strategy to identify the fetal abdominal standard plane. The lower layers of a CNN are pretrained based on natural images. The approach shows improved capability of the algorithm to encode the complicated appearance of the abdominal plane. Multitask training has also been suggested to handle the class imbalance common in CAD applications. Multitasking refers to the idea of solving different classication problems simultaneously and it results in a drastic reduction of free parameters.

3.2 Unsupervised Learning Model

Majority of deep learning methods focus on supervised deep learning however annotations of medical data especially image data is not always possible i.e. in case when rare disease or unavailability of qualified expert. To overcome, the issue of big data unavailability, the supervised deep learning field is required to shift from supervised to unsupervised or semi-supervised. Thus, how ecient will be unsupervised and semi-supervised approaches in medical and how we can move from supervised to transform learning without eecting the accuracy by keeping in the healthcare systems are very sensitive. Despite current best eorts, deep learning theories have not yet provided complete solutions and many questions are still unanswered, we see unlimited in the opportunity to improve.

3.2.1 Autoencoders

Autoencoders learn feature representations of input data (called codings) in an unsupervised manner without labelled data. It is a model that takes input data, gleans codings from this, and then uses these codings to reconstruct output data (called reconstructions). Autoencoders have several useful features.Figure 3.6 shows an image of autoencoder. Firstly, they are employed as feature detectors that can learn codings in an unsupervised manner,without training labels.Secondly,they reduce the model dimensionality and complexity as codings often exist in a lower dimension.Thirdly,by having to reconstruct outputs,autoencoders generate new data that is similar to the input training data. These features are an advantage in medical image analysis, where labelled training data is scarce. To force models to learn useful representations, constraints need to be added. One example is the Denoising Autoencoder reported by Vincentetal where Gaussian noise is added to the early hidden layers. Applying dropout i.e., randomly turning off some of the neurons in the early hidden layers, accomplishes the same goal, by forcing the model to learn useful codings to generate back the noise free inputs in the output layer. A second example are Sparse Autoencoders, whereby a dened proportion of the neurons in the hidden layers are deactivated or set to zero.

Variational Autoencoders (VAEs) are an emerging and popular unsupervised learning architecture described by Kingma and Welling. VAEs are a generative model, consisting of a Bayesian inference encoder network and a decoder network, that can be trained with stochastic gradient descent. The encoder network projects input data into latent space variables, whose true distribution is approximated using a Gaussian distribution. The decoder network then maps

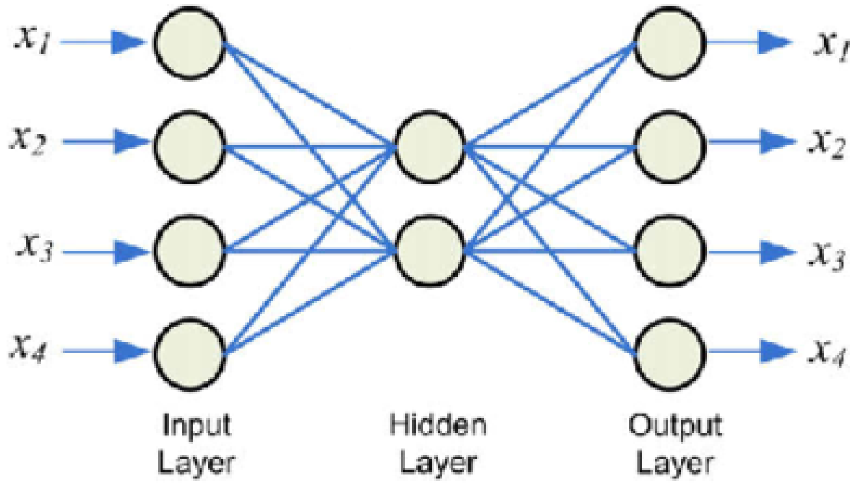

Fig. 3.6: Autoencoder

the latent space back into output data, trained and guided by two cost functions: a reconstruction loss function and the KullbackLeibler divergence.

3.2.2 Restricted Boltzmann Machine

Boltzmann machines were invented by Ackley et al in 1985, and were modied as Restricted Boltzmann Machines (RBMs) a year later by Smolensky. RBMs are generative, stochastic, probabilistic, bidirectional graphical models consisting of visible and hidden layers. These layers are connected to each other but there are no connections within the layers themselves (shown in figure 3.7). RBMs use the backward pass of input data to generate a reconstruction, and estimate the probability distribution of the original input.van Tulder et al modied RBMs into what they described as convolutional RBMs to classify lung tissue into normal, emphysematous, brosed, micronodular, or ground glass tissue. For this task, they used the CT chest scans of 128 patients with interstitial lung disease from the ILD database. Convolutional RBMs were trained with either purely discriminative, purely generative, or mixed discriminative and generative learning objectives to learn lters. These lters were then used to perform feature extraction and create feature activation maps, before classication using a random forest classier. Classication accuracies of between 41 to 68 percent were obtained, depending on the proportion of generative learning and the input patch size. They also found that lters generated from mixed discriminative and generative learning performed the best, concluding that discriminative learning could help unsupervised feature extractors learn lters optimized for classication tasks.

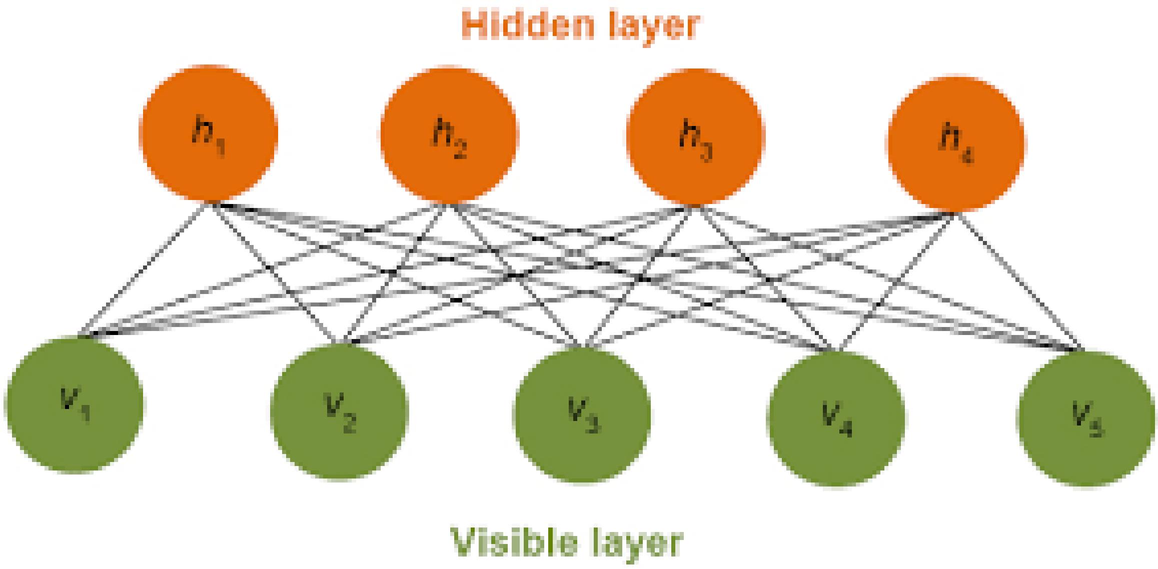

Fig. 3.7: Restricted Boltzmann Machine

Restricted Boltzmann Machines(RBM) can be efciently trained with Contrast-Divergence algorithms and stacked into Deep Belief Networks (DBNs), where the hidden layer output of a RBM becomes the input for the visible layer of a second RBM stacked on it. DBNs were described by Hinton et al. which was largely responsible for the renaissance in deep learning. The insight from Hinton et al was that DBNs could be trained in a greedy, layer by layer fashion , with lower layers learning low level features, and progressively higher layers learning high level features, mirroring real world data hierarchy. DBNs can also be coupled to layers of supervised RBMs to produce a semisupervised deep learning architecture. An application of RBMs was reported by Khatami et al, who used DBNs to classify x-ray images into 5 classes of anatomic areas and orientations.

3.2.3 Generative Adversarial Network

Generative Adversarial Networks (GANs) represent a type of unsupervised learning which holds promise for medical image analysis tasks. As its name suggests, a GAN is a generative model, and is similar to a VAE in that respect. GANs comprise of two simultaneously-trained, competing models,whic hmay be multilayer perceptrons such as CNNs (see figure 3.8). The models may be described as two players competing in a zero-sumgame.One CNN is a generator that generates articial training images.The other CNN is called a discriminator, which classies if images are real training images or articial ones from the generator.That is One neural network, called the generator, generates new data instances, while the other, the discriminator, evaluates them for authenticity; i.e. the discriminator decides whether each instance of data it reviews belongs to the actual training dataset or not.

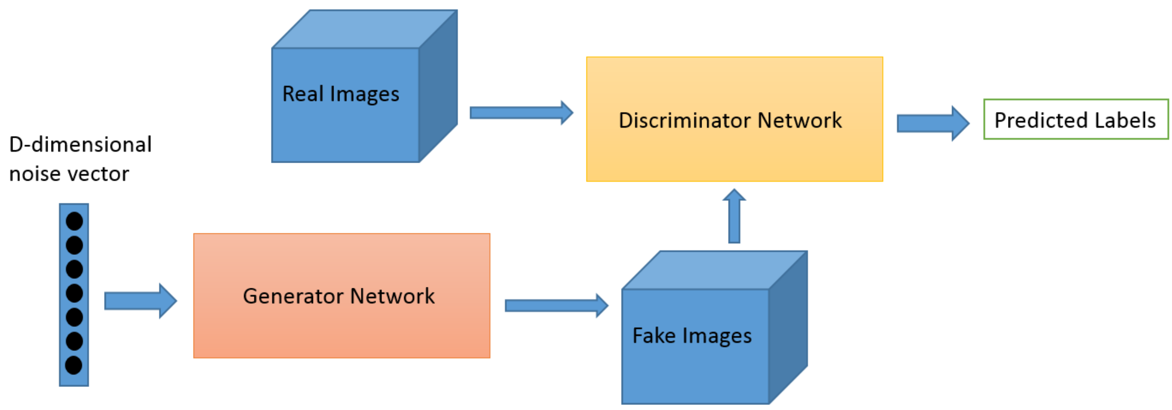

Fig. 3.8: General Adversarial Network

Here are the steps a GAN takes:The generator takes in random numbers and returns an image.This generated image is fed into the discriminator alongside a stream of images taken from the actual dataset.The discriminator takes in both real and fake images and returns probabilities, a number between 0 and 1, with 1 representing a prediction of authenticity and 0 representing fake.So you have a double feedback loop:The discriminator is in a feedback loop with the ground truth of the images, which we know.The generator is in a feedback loop with the discriminator.The desired end-point of this adversarial arrangement is one where the discriminator is unable to tell the difference between a real and a generated image i.e., the probability of assigning an image to either data distribution is 1/2. An advantage is that both generator and discriminator can be trained with back propagation and dropout, without unwieldly inference and Markov chains. GANs are relatively new but some applications in brain MRI segmentation and synthetic medical data generation.

3.3 Applications

3.3.1 Image Detection

Image detection plays an important role in computer aided detection routines. Its main purpose is to nd the tissues of interest, and then measure and analyze whether these tissues produce lesions. Some deep learning methods have been proposed for performing MRI image detection as follows.

To perform organ detection from a given complex dataset with abnormalities, for which it is difcult to identify the labels of the samples in the dataset, Shin et al[5] proposed a deep learning model with a stacked sparse autoencoder. In this study, the stacked sparse autoencoder model was generated by stacking several unsupervised feature learning layers, which were trained by using greedy methods. Subsequently, a pooling operation was applied to compress the features of gradually increased input regions, and to generate a part-based model to perform multiple organ detection in MRI images.

To achieve the automatic detection of lacunes of presumed vascular origin,Ghafoorianetal proposed an automated two-step deep convolutional network method. First, a fully convolutional network was applied to the detection of the initial candidates. Then, a 3D convolutional network was applied to reduce false positives. In this study, Ghafoorian et al suggested that location information plays an important role in the detection of candidate tissue. Therefore, to further improve detection performance, contextual information was generated by using multiple scale analysis and a combination of explicit location features to add into the convolutional network.

To detect cerebral microbleeds (CMBs) from MRI images, Dou et al[5] proposed an automatic 3D convolutional network method. The 3D convolutional network was used to extract high-level features for CMBs via a data driven approach,which can effectively encode the spatial contextual information from MRI images. Since the 3D convolutional network adopted at raditional sliding window strategy,the computational cost of using the method to detect CMBs was relatively high. To further improve the performance of CMBs detection, a two-step cascaded 3D fully convolutional network framework was proposed. The 3D fully convolutional network was rst used to rapidly retrieve potential candidates, and then used these potential candidates to further accurately distinguish CMBs from challenging mimics.

3.3.2 Image Registration

Image registration is the process of matching and superimposing two or more images at different times, different sensors (such as imaging equipment) or different conditions (such as illumination, position, and angle). The general process of image registration is as follows:The features are obtained by the feature extraction of two images; The feature pairs are found by performing a similarity measure; The image space coordinate transformation parameters are obtained by matching feature pairs.The image space coordinate transformation parameters are used to perform image registration. Image registration has been widely applied in medical image processing. Its main purpose is to combine various medical images, which display their information in the same image, and thereby provide multiple information for clinical diagnosis. Therefore, to achieve medical image registration, the building of accurate and effective correspondences between the two images is required. In general, the correspondences between two images can be represented by maximizing the similarity of the feature pairs.

Recently, as neuroimaging techniques developed, various new modalities have been emerging to make the diagnosis and treatment of diseases more accurate. Thus, image registration operations, which combine different modality data, are required. Many learning based image registration methods have been proposed to help select the best related features, which are used to guide the corresponding detection between samples with large changes. However, for most of the existing learning-based image registration methods, there is a great limitation with regard to the fact that they need a lot of known correspondences during the training process. To address this limitation, Wu et al[6] proposed an unsupervised deep learning framework to extract optimal image features for MRI image registration. In this study, rst, a stacked convolutional independent subspace analysis network was developed to learn the hierarchical representations of patches from MRI brain images. The stacked convolutional independent subspace analysis network included two layers: (1) The rst layer was used to extract the low-level feature representations of patches from MRI brain images; (2) The second layer was used to obtain the hierarchical representations. Then, the hierarchical representations were used to perform correspondence detection in the image registration process.

Later, in the same team, Wu et al also proposed an unsupervised deep learning framework to learn the hierarchical representations from MRI images for the purpose of image registration. The unsupervised deep learning framework contained a stacked autoencoder Jin Liu et al.: Applications of Deep Learning to MRI Images WA Survey 7 with a convolutional network. The 3D image patches were used as inputs for training the stacked autoencoder with a convolutional network. In this study, the stacked autoencoder mainly consisted of two networks;namely, the encoder and decoder networks, from which the former was used to learn the low-dimensional features from 3D image patches, and the latter was used to recover the 3D image patches from the learned low dimensional features. However, if the inputs of the stacked autoencoder are very large, the computational cost of directly using the stacked autoencoder will be very high. In this study, a convolutional network was used to learn the translational invariant features, which can reduce the dimensionality of the original features,to reduce the computational cost of the stacked autoencoder.

3.3.3 Image Segmentation

Automatic tissue segmentation in MRI images is of great importance in modern medical research and clinical routines. Many medical image segmentation challenges have been held to encourage the development of automatic segmentation techniques, such as Ischemic Stroke Lesion Segmentation (ISLES), Multimodal Brain Tumor Image Segmentation (BRATS), MR Brain Image Segmentation . Many deep learning methods have also been proposed to perform the segmentation of various tissues in MRI images.

In MRI brain images,one of the most common image segmentations is the segmentation of Gray Matter (GM), White Matter (WM), and Cerebrospinal Fluid (CSF). To segment infant brain tissue images into GM, WM, and CSF, Zhang et al proposed the use of convolutional networks to achieve this goal by combining multi-modal MRI images, which are T1, T2, and Fractional Anisotropy (FA) images.The convolutional network architectures contained a different number of convolutional layers and resulting feature maps. To obtain the nonlinear mappings between the inputs and outputs of each convolutional network, the local response normalization scheme, the networks. Moreover, to segment neonatal brain tissue images into Brain Stem (BS), cortical GM (cGM), myelinated WM (mWM), Basal Ganglia and Thalami (BGT), unmyelinated WM (uWM), ventricular CSF (vCSF), extracerebral CSF (eCSF), and cerebellum (CB), Moeskops et al also proposed a convolutional network to automatically segment these tissues. Similar to the convolutional networks previously proposed by Zhang et al[6], the convolutional network also contained multiple convolutional layers and the resulting feature maps. Additionally, the fully-connected layers were also used in the convolutional network to represent each input patch size,and a single softmax output layer was used to connect these convolutional and fully-connected layers to perform the nal segmentation. Since most brain tumors can affect a patients health, and even shorten their life expectancy, automatic and reliable segmentation techniques for removing brain tumors are required. However, most brain tumors have large spatial and structural variability, which makes them difcult to segment. Thus, automatic and reliable segmentation has become a challenging problem. To address the problem, many deep learning-based brain tumor segmentation methods have been proposed.

For example, Pereira et al used a convolutional network with small convolutional kernels to segment gliomas, which are the most common and aggressive brain tumors in MRI images. They believed that by using smaller kernels more convolutional layers could be stacked, and that the same results with larger kernels could be obtained. Additionally, to further improve the segmentation performance, both intensity normalization and volumetric constrains were used in the convolutional network. Later, Havaei et al[6] also presented a fully automatic brain tumor segmentation method with a convolutional network. Unlike most traditional convolutional networks, the convolutional network in this study included three new components; namely, a two pathway architecture, cascaded architecture, and two-phase training. The two-pathway architecture was used to obtain global contextual and local features, respectively. The cascaded architecture contained input concatenation, local pathway concatenation, and pre-output concatenation, and was used to exploit the output efciency of a convolutional network.

Conclusion

During the recent few years, deep learning has gained a central position toward the automation of our daily life and delivered considerable improvements as compared to traditional machine learning algorithms. Based on the tremendous performance, most researchers believe that within next 15 years, deep learning based applications will take over human and most of the daily activities with be performed by autonomous machine. However, penetration of deep learning in healthcare especially in medical image is quite slow as compare to the other real world problems. Many big research organization are working on deep learning based solution that encourage to use deep learning to apply deep learning on medical images.Looking to the brighter side of machine learning, we are hoping the sooner human will be replaced in most of the medical application especially diagnosis. However, we should not consider it as only solution as there are several challenges that reduces its growth. One of the big barrier is unavailability of annotated dataset. Recent development on other application showed that bigger the data, better the result, however, how big data could be used in healthcare. So far deep learning based application provided positive feedback, however, but due to the sensitivity of healthcare data and challenges, we should look more sophisticated deep learning methods that can deal complex healthcare data efficiently.

REFERENCES

[1] C.-J. Hsiao, E. Hing, and J. Ashman, Trends in electronic health record system use among ofce-based physicians: United states, 20072012, Nat. Health Stat. Rep., vol. 75, pp. 118, May 2014.

[2] R. Smith-Bindman et al., Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems,19962010,JAMA,vol.307,no.22,pp.24002409,2012.

[3] E.H.Shortliffe,Computer-Based Medical Consultations:MYCIN,vol.2. New York, NY, USA: Elsevier, 1976.

[4] W.S.McCullochandW.Pitts,A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biol., vol. 5, nos. 4, pp. 115133, 1943.

[5] G.Litjensetal.(Jun.2017).A survey on deep learning in medical image analysis. [Online]. Available: https://arxiv.org/abs/1702.05747

[6] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, Psychol. Rev., vol. 65, no. 6, pp. 365386, 1958.