International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

Recent Advancements in Machine Learning and Artificial Intelligence Techniques for Cancer Diagnosis

Swathi C Shekhar[1],Sushma P[1],Suhas Bharadwaj B[1], Sampath Kumar B S[1],

S Balaji[2], Roopa Reddy[2]

[1]Department of Computer Science and Engineering, Jyothy Institute of Technology Tataguni, Off Kanakapura Road, Bangalore-560082, India

[2] Centre for Incubation, Innovation, Research and Consultancy, Jyothy Institute of Technology Tataguni, Off Kanakapura Road, Bangalore-560082, India

---***---

Abstract -*In the current epoch of information age, the medical field is being upgraded at an alarming rate especially in the field of cancer diagnosis and rightly so as cancer research is one of the most complex and mind-bending challenges. The effects of cancer can only be subdued if it is diagnosed in its early stages and in turn will increase the longevity of the patients. Recent advancements in storage, computation, processing etc. have enabled researchers to effectively implement techniques of machine learning and digital image processing for medical applications. After surveying different techniques of digital image processing and machine learning, various methods were listed which were used for processing the images based on the type of images acquired, noise/disturbances prevalent in those images, area of the biopsy sample captured (histologic or cytologic) and other similar parameters. The classification of these images collected after pre-processing is usually done through a machine learning algorithm. Many implementations of similar techniques were studied to familiarize with the accuracy of different machine learning algorithms, for different scenarios. Exploration of deep learning and convolutional neural network techniques was done through this exercise. Numerous APIs and tools available for easy implementation of deep learning techniques are also briefly discussed. Deep learning is now the frontier of computer vision and it has given very good results across the domains. The reasons for choosing deep learning methods in place of conventional methods being followed so far are also discussed. This paper also gives a bird's eye view of the various techniques that can be used while handling medical image processing and classification.*

Key Words:**Digital Image Processing, Whole-Slide Imaging, Machine Learning, Deep Learning, Convolutional Neural Networks, Cancer Diagnosis.**

1. INTRODUCTION

Digital image processing and machine learning have been extensively used in medical applications and especially for cancer diagnosis and classification. Deep learning is spreading its roots in the field of biomedical image analysis and these systems are being used as a machine aid for human experts. These technologies give access to huge amount of data thereby increasing the possibility of achieving high accuracy classification systems. Given different tumor types and the many categories within the tumor type, it is essential to have substantial computational resources including processor power, memory space for reducing the time consumed to a manageable limit once we have designed automated system and the machine learning algorithms with statistical analysis for classifying the tumor stage/grade [1][2].

Deep Learning with complex structures helps in recognizing and classifying biomedical images with high accuracy, as there are numerous algorithms with variations available[3]. This paper focuses on available conventional methods of digital image processing, machine learning algorithms, deep learning techniques, WSI handling methods, APIs and the Python packages and modules that enable designing the software tools for cancer diagnosis.

2. CONVENTIONAL METHODS

Steps involved in diagnosing the presence of cancer using digital image processing and machine learning techniques is shown in Fig. 1. In conventional methods, image preprocessing is the initial step for removal of noise in the images followed by image segmentation which takes the preprocessed images for segmenting to get the region of interest. The third step, namely, feature extraction extracts the features from the segmented region which are fed into a machine learning algorithm for further classification.

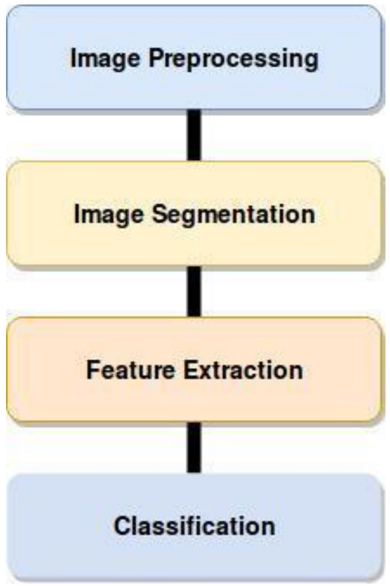

Fig. 1:General Purpose Machine Learning System for ImageProcessing

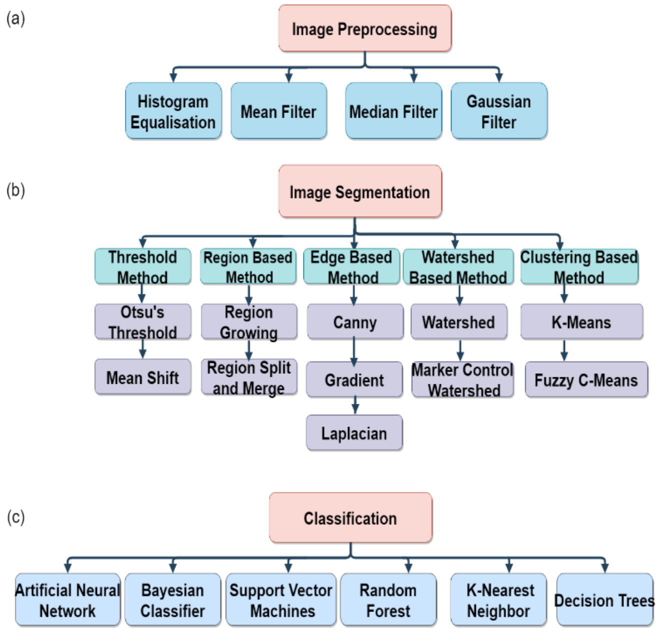

Fig. 2: Image Processing Pipeline

Fig. 2(a) depicts the image pre-processing methods. These methods are used to remove the unwanted disturbances and/or noise in the images and to manipulate to get better images than the original images. Fig. 2(b)gives an overview of image segmentation methods. These methods divide the images into segments based on the region of interest for further processes.Fig. 2(c)shows various classification methods. These methods are commonly used for classification of the features extracted using the feature extraction methods.

2.1 Image Pre-processing

Image pre-processing is a crucial and elementary step of image processing. It transforms the images into more amiable form for both systems and individuals. Representing and performing various operationsin downstream steps such as segmentation and feature extraction are much easier for pre-processed images[4][5]. One of the important aspects of the image pre-processing is to differentiate foreground subject from background noise. This distinction of the noise is important for the image processing and computer vision techniques as it reduces the computational resources utilized and neglect the background noise if the foreground is the only required aspect of interest [6].

There are several methods for carrying out the image pre-processing, including color illumination methods, smoothing and sharpening filters, denoising methods, etc. A brief description about some of the popular methods used for image pre-processing follows.

Histogram Equalization

It allows the contrast of the image to be changed by modifying the intensity of the entire spectral distribution according to the histogram. The main objective of this pre-processing technique is to provide a linear trend pertaining to cumulative distribution probability of function (CDF) associated with the image. It is the major component on which histogram equalization is based on.The main advantage of this pre-processing step is its simplicity and effectiveness in its applications. It gives the resulting image which is linear and also enhances the quality of input image with reference to contrast [5][7][8][12].

Mean Filter

Mean filtering is a simple and intuitive step adopted for soothing the images to decrease the intensity variation that exists between one pixel and its adjoining ones. It is implemented mainly to reduce the noise and average of all the pixels of the image is calculated to assign the value for each pixel under consideration. This process screens and eliminates the odd pixel intensities and is considered as one of the convolution filters[7] [9].

Median Filter

Median filter is similar to mean filter in preserving the details of the image and also considers every pixel in the image along with its neighboring pixels for analysis. Unlike mean filter, this method allows median to be calculated by sorted assortment of all pixel intensity values of all the surrounding images and assigning a numerical value. The median pixels calculated will be used to represent the pixel value to be considered. Implementing median filter reduces the noise and also retains the sharpness of the image and is considered as one of the best methods for speckle noise reduction[9].

Gaussian Filter
Gaussian filter is used to blur images and remove noise from the images. As it is a non-linear filter, it has advantages over linear filters [7].

2.2 Image Segmentation

Image segmentation represents the decisive part of the image processing technique and largely contributes in determining the outcome of the application. This step allows the region of interest to be focused as an area to give definitive results for the system by separating the other regions of image as background image. Discontinuity and similarity of intensity are the major classification factors considered for segmentation process. The first step considers the abrupt peaks or dips in the adjoining pixel intensity and the second step separates the foreground and background images by predefined parameters [4]. One of the important applications of image segmentation is clinical sample analysis for the diagnosis of the disease where the region of interest is usually the cells, tissues or images of the whole area with negligible difference from its surroundings [10].Image segmentation plays a very big role in image processing and the field is very vast with various algorithms and techniques. Fig. 2(b)has shown some of the important ones.

Thresholding Method
Thresholding is the simplest segmentation process which considers the intensity level distribution of each pixel, based on which the image is divided accordingly. This method is applicable for the images with lower intensity value of pixels of the region of interest than the background. Selecting this method is highly dependent on the availability of the image information. There are many types of thresholding methods for image processing including global, variable and multi-thresholding methods.

Global thresholding is carried out by approximating constant T as the threshold value and the whole image and output image is obtained with respect to this thresholding constant T. In variable thresholding, constant may vary in different parts of the image and multi-thresholding is a subtype of variable thresholding which consists of different threshold values based on which the output image is obtained [11].

Region-Based Method
This method relies on similar characteristics of images to divide into sub-regions. Further, two primary techniques are used in this method: (i) region growing methods and (ii) region splitting and merging methods. In region growing methods, the initial pixels of the image are considered to be seeds, based on which region of the image spreads. The selection of these seeds can be done through sophisticated applications or manually. The propagation of region may be controlled by specifying the connective pixels which are in between the growing seeds with the prior information about the image pixel connectivity thus, reducing the uncontrolled growth of the image region [11] [12]. In region splitting and merging methods, the image with same features or characteristics are recursively divided into sub-regions, and these iteratively divided portions of the images are consolidated into a single subregion, considering that to have similar characteristics in the merging techniques [12].

Edge-Based Method
This technique mainly relies upon the instant change in the intensity level along the edge. Single pixel does not contribute in providing adequate information about intensity change. Hence, accumulative rate of change in the intensities are calculated by determining the first and second order derivatives of the pixels. A threshold value is set in consideration of first order derivative, whereas zero level crossings are checked in second order derivative. This method employs technique of connecting the edges of the desired region that has to be segmented. To deploy detection of the edges, many operators such as Canny, Sobel and Robert can be implemented. The output of these techniques is mostly in the form of binary images. This method is one of the types where in discontinuity detection is carried out in segmenting the region of interest [11] [12] [13].

Watershed-Based Method
The watershed transformation has gained wide importance in image processing and biomedical fields as it possesses wide range of advantages. The various advantages of this technique are that it is simple to implement and efficient in determining regions in images with poor contrast level.The watershed algorithm produces a unique solution for each sub-division of the image. This method can be implemented for various structured images with "n" dimensions or images with varied grid sizes. Though there are many advantages in implementing this method, there is the possibility of over segmentation and images with low smoothness level adds to the difficulty in efficient segmentation [4][14].

Clustering Based Method
Segmentation using this technique is carried based on similarity-based regions. Similar elements are considered to be clusters and these clusters are formed by dividing the data into smaller ones in data clustering. Clustering based methods are of two types: hierarchical and partitioned. Hierarchical representation is tree representation of the data where root node represents data and leaves represent the clusters. In partition-based methods the objective function of the system is minimized recursively.

Hard clustering is a simple technique where the image is divided in the form of clusters that consists of several pixels. Each pixel in the cluster exclusively belong to that cluster and hence should not be present in any other cluster. Thus, membership of the pixel related to the cluster should be either one or zero.Soft clustering is a natural type of clustering and is applicable in situations where dividing images in the form of clusters is not efficient given the huge noise in the data. By selecting appropriate segmentation algorithm, the efficiency of the model can be improved [11].

2.3 Feature Extraction

Feature extractionidentifies the attributes or the informationof the data that enhances certain characteristics of the data and also contributes inreducing the "curse of dimensionality" of data by eliminating the redundant information[15].For example, in biopsy images of a single cell, the shape, morphology, texture, histogram of oriented gradients, and wavelet features are extracted. Some important shape and morphological based features including nucleus area, brightness of nucleus, nuclear distance (longest, shortest perimeter, roundness solidity and compactness) are used for computing feature vector of microscopic image features such as texture features, morphology and shape features, histogram of the oriented gradient, wavelet features, color features and others [16][17].

2.4 Classification

After completing the feature extraction, the output generated is used as a precursor to train machine learning algorithms. By using several datasets, the algorithms are trained to develop classification model that analyses the features extracted from training data, and data that need to be classified and predicted for the unknown test sample [16][18]. Fig. 2(c)shows major types of machine learning algorithms. One of them will be chosen based on the application under construction. Some of the commonly used algorithms are briefly discussed below.

Artificial Neural Networks (ANNs)

Target functions such as real valued, discrete valued and vector valued approximations are possible through neural network analysis. ANN is highly relied up on for solving problems related to learning tasks. It provides a robust approach for handling real world data generated from sensors. ANNs consist of many nodes that are called neurons which are divided in the form of layers. A typical ANN has three layers - input layer, hidden layer, output layer - where each layer is connected to the next layer and the weights of the particular features are calculated in these neurons and propagated forward. Weight adjustments are done automatically by the network by feeding the result backwards. Thus, ANNs are gaining more prominence in learning tasks as they deploy corrections of weights. Pattern recognition problems can be effectively solved using ANNs [18][19]. Fig.3(a) depicts thebasic diagram of a generic ANN which contains three layers, namely, input layer, hidden layer and output layer. Input layer takes the input with its corresponding weights and processed in each of the hidden layer and passed on to the output layer for classifying the output.

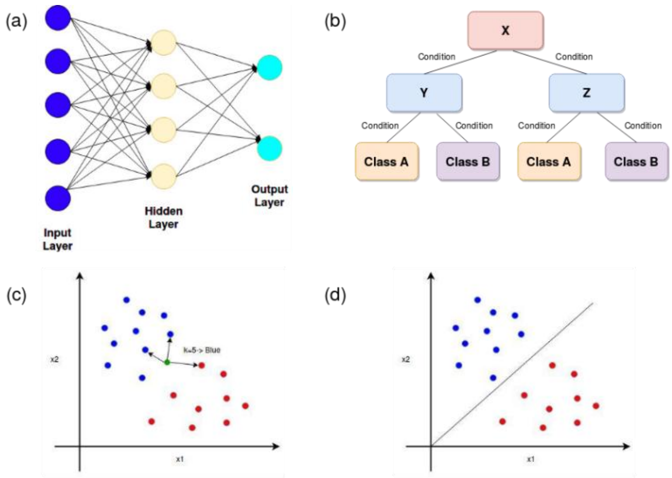

Fig. 3: Representation of Some Classification Models

Bayesian Classifier

Bayesian classifier techniques take a probabilistic approach for classifying the classes in the form of nodes of direct acyclic graph based on the probability of the testing dataset belonging to each class. The highest probability valued class is considered for assigning its label to the testing dataset. Bayesian classifiers have gained importance in learning and responsive tasks, and classifying tasks [19].

One of basic Bayesian classification method is Naive Bayes' classification where classification assumes that the attributes are independent of each other, thus simplifying the calculation of probability. It is based on Bayes Theorem where probability of a particular event is calculated while another event has already occurred [20][21].

Decision Trees

Decision Tree represents the hypothesis space in the form of tree where root node represents input variables and leaves correspond to the output or the decisions. For classification problems, decision trees are considered to have more prominence as it is one of the most popular techniques. With its simple architecture, decision trees have proved to be an easy learning technique. While classifying the new dataset, traversal of the appropriate branches corresponding to features in the dataset will lead to leaf of the tree that represents the class to which our dataset belongs to. The decision trees often suffered with overfitting due to inadequate number of datasets while constructing the tree [19]. A decision tree is represented in the Fig. 3(b) where X is the root node, Y and Z are the possible outcomes from the root node. Further, Y and Z lead to other possibilities for the classification result.

Table.1: **Comparison table of conventional method**

	Methods	Tasks	Benefits	Assumptions/Limitations
1	Gabor Filter	Image Preprocessing	Enhance images for texture based feature extraction	Applied on histopathological images
2	Mean Filter	Image Preprocessing	Used for smoothening and reducing noise of images	Kernal representation of size and shape of neighbourhood
3	Median Filter	Image Preprocessing	Gives sharp edges of images, after processing	Average of neighbour pixel values are considered
4	Gaussian Filter	Image Preprocessing	Used for blurring images before noise removal	Convolution is applied for each pixel
5	Histogram Equalisation	Image Preprocessing	Images with uniform intensities are obtained	Applied on histopathological images
6	Distance Transform Algorithm	Image Preprocessing	Used for conversion of binary images to gray scale images	Uses Euclidean distance for transformation
7	Watershed Algorithm	Segmentation	Easily adaptable for n-dimensional images	Smoothness of image is lost leading to over segmentation
8	Marker Controlled Watershed	Segmentation	Accuracy of segmentation is more compared to Watershed Algorithm	Edges of region of interest has to be enhanced before segmentation
9	Sobel Edge Detection	Segmentation	Used for edge detection	Gradient measurement of pixels are considered
10	Canny Edge Detection	Segmentation	Multi-step algorithm for edge detection	Uses Gaussian Filters for noise reduction
11	Otsu's Thresholding	Segmentation	Used fro thersholding	Considers foreground and background intensities separately for segmentation
12	Gray Level Co-occurance Matrix	Feature Extraction	Energy, Entropy, Contrast, Correlation, such features are extracted	SVM can be used for classification
13	Convolutional Neural Network	Feature Extraction	Intake images and output is set of features	Even though can be used for classification, training CNN needs large amount of data
14	Support Vector Machines	Classification	It finds best separation between two identities	Binary Classifier
15	Bayesian Network	Classification	Model with acyclic graph is obtained	Less accurate than other classifying algorithms
16	Decision Tree	Classification	Easily understandable training algorithm	Error in training examples leads to highly complicated tree
17	K-Nearest Neighbour	Classification	Noise Tolerant algorithm	Takes too long for computation
18	Genetic Algorithm Approach	Segmentation/Classification	Color based segmentation used for feature selection and classification	Complication in representing training/ output data
19	Neural Network Based Approach	Segmentation/Classification	Color based segmentation used for feature selection and classification	Complex structure

Entropy of system is calculated to know the homogeneity. The Entropy of each node of the system is calculated by,

$$E(s) = \sum - \mathrm{p}_i log_2 \mathrm{p}_i$$

where E(S) is the entropy of the system and $p_i$is the probability of the corresponding node. Information gain for each attribute is calculated by the equation.

$$Gain(T, X) = E(T) - E(T, X)$$

where *Gain* (T, X) is the gain of node T with an attribute value of X.

K-Nearest Neighbor

It is aninstance-based classification method where classification of the data points is carried out by considering Euclidean distance of the "k" nearest data points to that point. Fig. 3(c)depictsk-nearest neighbor classification method.

Support Vector Machines (SVM)

The classification process in SVM is performed with separate labelled data with a hyperplane resulting in the maximum difference between the distance of the respective hyperplanes. SVM is widely used as a binary classifier [12][22]. SVMs binary classification is shown in Fig. 3(d).

Table. 1 gives the summary of benefits and limitations of conventional methods of image processing and machine learning [18-25].

Studies in the last few years have shown that Deep learning techniques have added advantages over the conventional methods [26].Themajor drawback in the use of conventional methods is in the selection and the choice of implementing a suitable feature extracting algorithm, which is a crucial step in model training process and also a decisive factor for accuracy of the final output [33]. Due to this, deep learning has become a popular choice of machine learning technique, whichhas sidetracked the conventional methods.

3. DEEP LEARNING (DL)

Conventional machine-learning algorithms and techniques struggle to process raw data which is abundantly available. DL techniques are just representation-learning methods but with multiple levels of representation, which are obtained by composing simple but non-linear modules. Each of these modules transform the representation at one level. By using these transformations, extremely complex models can be obtained. With respect to images, if the first layer detects the edges in a few areas from the raw data, that is, arrays of pixel values then the second layer with a collection of these edges will form a more complex feature and so on, until the actual object that is being detected is recognized by the network with acceptable accuracy.

The interesting fact about DL method is that even though layer parameters are defined by the programmers, the final output of the layers are not defined by them and instead these features are learned from the data by the machine itself. Therefore, it is a stand-alone learning system for many similar datasets [27].

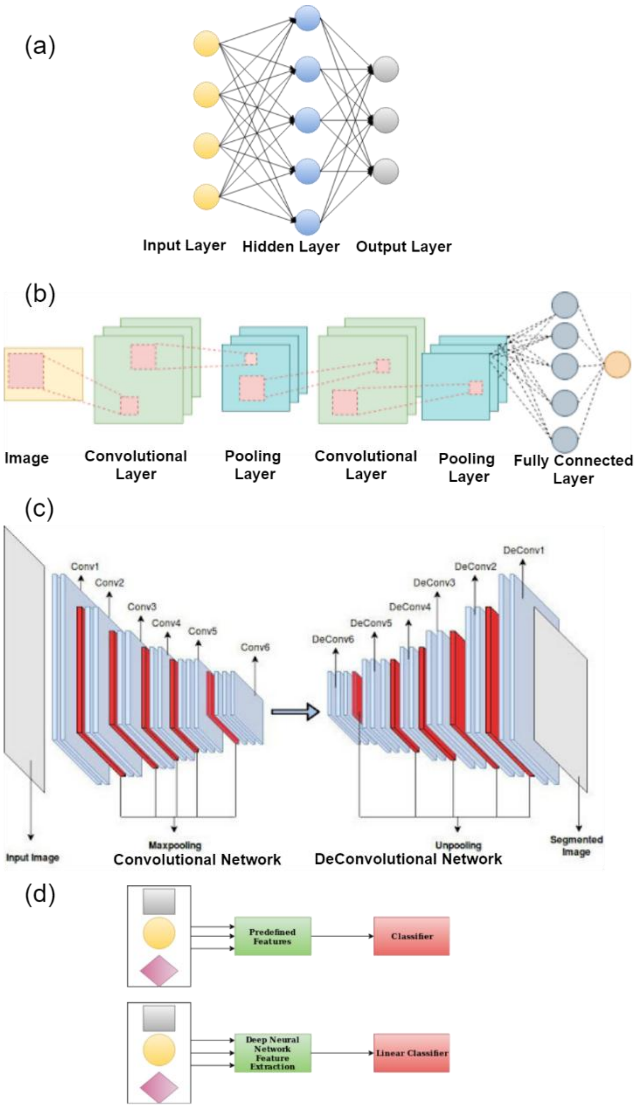

Fig. 4: Different Deep LearningModels

DL methods have made tremendous progress in the fields of image recognition, analyzing particle accelerator data [30], reconstructing brain circuits [31], predicting the activity of potential drug molecules, beating a team of professional game players, beating the world champion in the ancient Chinese game of GO [29], nuclear medicines, molecular imaging [41], robot-assisted surgery [42] etc. Supervised learning is the most widely used type of machine learning irrespective of the algorithm being DL or not. The DL algorithm adjusts and redefines itsweights iteratively to reduce the loss occurred during the training and increases the accuracy of output.

DL techniques represents the stack of multilayer modules that can learn or process non-linear data acquired as input and give output which are nonlinear too. Each module in this contraption plays a vital role in adding weights and selecting the parameters.

A typical ANN, as stated earlier has many identical processing components called as the neurons andobtaining real-valued activation as the output from each of them is one of the popular deep learning techniques [27].The neural networks will have many layers involved and they are popularly divided as input layer, output layer and hidden layers [28].
Representation of “2 Neural Layer Net” or “1 Hidden Layer Neural Net” architecture shown in Fig. 4(a). Each node in the input layer takes an input with the corresponding weights and calculates the features. Forward propagation takes place here. For automatic weights adjustments backward propagation takes place.

Feed forward neural networks are used when fixed sized output is needed from a fixed sized input. Each module in a layer takes in the input from the previous layer and gives the output by assigning or reassigning the weights. This continues till the end of the non-linear structure. Units that are not in the input or output layer are conventionally called hidden units [32][34].

A SoftMax output layer is used in deepneural network orneural network in general to get the class probabilities. The output layer so obtained will convert the logit, that is, ‘z_i’ which is computed for all classes into a probability represented by ‘q_i’ and this is done by comparing the obtained z_i with all the other logitsas shown in the below equation

$$q_i = \frac{exp(z_i/T)}{\sum_j exp(z_j/T)}$$

Studies at Google have recorded the results of “distilling the knowledge in a neural network” where they had trained 10 models by implementing the same architecture and other training procedures. Initially, parameters were all randomly assigned, that is, the bias was random and later they implemented a technique which they refer to as distillation to get a better output than training the model using hard label. This distillation process can also be used when training a number of models simultaneously [72].

Though Deep Neural Network show interesting results in the field of computer vision, it can also be very insensitive to some cases as in “deep neural networks where high confidence predictions are depicted for unrecognizable images” [73].
Researchers from CIFAR (Canadian Institute for Advanced Research) have reviewed deep feed forward networks by exploring the unsupervised DL which showed the researchers worldwide that even unlabeled data can be processed by them. This is also known as pre-training and its first practical application was in the domain of audio

processing. This coupled with the advent of GPUs paved the way to train models very quickly [38]. Further, there was efficient improvements in this field to encourage the researchers to develop more efficient ways of doing the same. One of the most promising fields of study in this domain comes in the form Convolutional Neural Networks [32][34].

3.1 Convolutional Neural Networks

Convolutional neural network (CNN) is a deep feed forward network which takes input in the form of a bunch of arrays and this type of data is readily available in the form of images. Images are multi-dimensional arrays with each unit holding the pixel values and intensities. The convolutional neural networks take advantage of four properties of natural signals: local connections, shared weights, pooling and the use of many layers.

CNNs are multi-layered neural networks which can be further sub-divided into convolutional layers and pooling layers. Fig. 4(b) gives an overview of convolutional neural network architecture where the raw image is taken as input to the network. Further,the process has different levels with a combination of convolutional and pooling layers. Fully connected layer classifies the input image into its corresponding category.

In convolutional layers, the local patches of precedent layers are connected to units organized as feature maps through the weights in the filter bank. The locally weighted sum is passed to activation function such as rectified linear unit (ReLU). Each unit in feature maps corresponds to same filter bank whereas different feature maps in the layers correspond to different filter banks.The pooling layer which comes next will help to put together features that are semantically similar and merges them to make more complex structures or feature rich layer.

The maximum of a local patch of units are computed in one feature map in a typical pooling cycle. More convolutional and fully-connected layers are preceded by pooling, non-linearity and convolution layers. A CNN takes advantage of the fact that most of the features fed are just compositional hierarchies,that is, smaller units combine together to form larger and more complex units.

Irrespective of the domain selected such as images, text, sounds, similar compositional hierarchicalpattern is observed. Development of such a system is based on how biological neurons work in human brain. It is also possible to study, understand and simulate human brain by building and studying these CNNs. Numerous applications like document processing, semantic analysis of documents, and sounds and images have been created using the CNNs. The document processing system uses a convolutional neural network and can also be trained to implement constraints on languages [26-27] [35-36][40].

Usage and the popularity of the CNN grew after the publication of 'ImageNet Classification with Deep Convolutional Network' paper by Alex Krizhevsky where he used his Deep CNN dubbed as AlexNet on the ImageNet Classification challenges [37].

Large scale distributed clusters, GPUs and gigantic public image datasets and repositories are available to a common man today and this has made convolutional neural networks and other computationally intensive technologies easily executable for everyone.

On fixating all other parameters of the convolutional neural network, for instance, and start tweaking the depth of the network over a fixed dataset, we need to understand the appropriate depth of the neural network which will result in the proper customization of standard neural network architectures for a wide array of problems. The above experiment was carried out by "Very deep convolutional networks for large-scale image recognition" [70] and they, in turn, were inspired by Ciresanet al.(2011); Krizhevsky et al. (2012)[71]. The implementation of the above said project was done by Alex Krizhevsky and was concluded that running the same on a system with multiple GPUs will help with data parallelism and would, therefore, reduce the computational resources utilized [70].

There is another variant of the CNNs called the Fully Convolutional Network (FCN) widely being used for some of the above stated cases. As already discussed in CNN to have multiple layers and each layer is a 3D array, where two of the three dimensions are spatial dimensions and the other feature is the feature dimension. If we represent the 3 layers as x * y * z, the first layer, that is, x * y is also the image dimension in pixels.Fig.4(c)represents afully convolutional neural network architecture which accepts the input images irrespective of its size. This provides the segmented image according to the size of the input image where the required region of interest is highlighted.

As stated earlier, a CNN will have three important functions: pooling, convolution and activation. These components of the network operate on local input regions and they do not depend on the absolute but only on the relative spatial coordinates [39].

The most obvious way of increasing the efficiency of a deep neural network is by increasing the depth of the network and also its width (size of the network). This is the easiest way of acquiring models with higher accuracy and this is true especially when availability of data is more. But this method has a major drawback in the form of the amount of input features that has to be considered, which certainly leads to overfitting.Feeding the huge input data is practically difficult as it is a very tedious task to acquire labelled data [43].

3.2 Feature Extraction Using DL

Feature extraction is the process that builds derived data which would provide non-redundant information that, in turn, facilitates the learning and categorizing tasks. Feature extraction is also associated with dimensionality reduction. The data given to the system as input may contain redundant information which can be abandoned while considering features. This procedure of considering a small group of preferred features is called feature selection. Analyzing a complicated data requires efficient resources such as large memory space as well as high computational power and efficient algorithm to classify the data according to the training pattern.

Feature extraction facilitates analysis by providing efficient set of data, that would define proper features that have prominent role in recognizing the pattern. Traditional image recognition methods mainly focus on gaining features which are predefined while training the model and then use that information to classify the image set [18-19][44].

Deep neural network will automatically extract features from images which will, in turn, be served as input to classification tasks [44]. Difference between conventional machine learning algorithm and DL algorithm is shown in Fig. 4(d).Row one represents conventional machine learning approach of extracting predefined features. Row two represent deep neural network of automated feature extraction.

Muhammad Farooq and Edward Sazonov [44] proposed feature extraction using pre-trained CNN AlexNet and classification using SVM for food type recognition. Food images of size 600x800x3 was resampled to size of input of AlexNet 277x277x3 (RGB). AlexNet consists of 23 layers in total with 3 fully connected layers. As SVM is a linear classifier, features extracted from the AlexNet avoids the overfitting of datasets.

Feature extraction in medical images by using deep learning approach is discussed in [45] and the authors have proposed automated feature extraction from three different models, namely,convolutional neural network, multilayer perceptron-based picture and multilayer perceptron-based component. The datasets were divided into 61 samples for training and 8 samples for testing and achieved the classification results of 99%, 93% and 93% accuracy, respectively.

Mahesh et. al., [46] have proposed a system of discriminative feature extraction from X-ray images using deep convolutional neural networks. Convolutional neural network with three different type of layers, namely, convolution, max pooling/sub-sampling and fully connected layers was considered to be helpful in extracting significant features as it retains the spatial correlation of the image. At convolutional layer, features from preceding layers which are mapped, are passed through activation function. Activation function generates output feature map which is passed to subsampling/max pooling layers where down sampling of input maps takes place. The proposed architecture constituted CNN with four convolution layers of size 9x9, 5x5, 5x5, 2x2.Alternatively, sub-sampling/max pooling window of size 2x2 was chosen. To classify the extracted features into 12 classes, fully connected layer was used as the last layer.

Maram et. al., have articulated a system to extract features for iris recognition using CNN [47]. As a CNN automatically extracts best features of the image, it is considered to be efficient in recognizing most unique physical feature, in this case, iris. Pretrained CNN model AlexNet, and SVM are used to extract the features automatically and classify, respectively. After segmenting and normalizing, only iris region is given as input to AlexNet so as to reduce the computational complexity of the model. Imagenet Large-Scale Visual Recognition Challenge (ILSVRC) has trained AlexNet to classify 1.2 million images into 1000 classes and accuracy of the system was observed to be high when segmented images of size 277x277x3 was given as input than normalized images.

In object specific deep learning feature and its application to face detection [48], authors have proposed multi-resolution approach to constructing robust face heatmaps for fast face detection based on Object Specific Channel (OSC) features. Images are fed into CNN and face's heat map is generated from multi-resolution features extracted from face specific channel. Future work of the paper includes identifying cars from photography to cell identification in medical research.

In feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification [49], patch level and image level features are generated, and image classification is done simultaneously during end to end learning process. From histopathological images, nucleus features such as pattern and spatial distribution are extracted by CNN enclosed with guide of nuclei. According to the authors, the proposed framework has achieved best classification results compared to other conventional methods.

Convolutional neural network takes in patches of original image whose weights are taken into account during training. The filters embedded in layers extract low level details from these images. The number of parameters needed for computation is reduced in subsequent pooling and convolutional layers. Fully connected layers attached lately are responsible for classification tasks. Thus, convolutional neural networks are relied more upon for extracting feature information automatically.

3.3 Patch Based CNN

Patches are sub-images derived from the original image. Patch can be uniquely identified by horizontal and vertical location inside image, coordinate of center of patch and its size. Patches can be extracted by calculating pixel location of the particular square when the location and the size are specified. Patches uphold the local features by describing the properties of the certain region of the image, whereas global features provide information about whole image. Global features contribute to extraction of texture information, color distribution or whole image information. Information accessed from the global features often turn out to be inadequate, whereas local features like patches will suit to represent restricted region of complex images.

Extraction of these patches can be done through various methods like grid points specification, where regular grid of desired patch size is projected on the image which provides the points to extract. Gaps might be included between the patches depending on whether they overlap or not. Random point specification is a method that is similar to grid point specification with the only difference being that random point specifications chooses the points in random. Hence, this is distributed over the image. Another method is interest point specification where region of interest is the focus and the points inside the same is considered for generating the patches.

Advantages of patch-based approach are:

- Recognition of the object is location independent. Object that has to be recognized might be present in different locations in different images. As in the case of patch-based approach, those can be identified irrespective of the location.
- Identification of part of the object. Patch-based approach helps to identify the objects presence in the image even if it is partially occluded.
- Irrespective of size scaling in different images, objects can be identified depending on the patch size [50].

Patch-based CNN was specifically used in the music score images [51]. CNN used in the proposed system consists of three convolutional layers which takethe input. Output of these layers are fed into max pooling and LRN layers. Three fully connected layers consists of 512 neurons each. This model also consisted of two dropout layers and they were termed as dense1 and dense2 probability of a 50% drop. Glorotinitialization and ReLU activation are used for initialization and activation, respectively for convolutional fully connected layers. According to the authors patch-based CNN approach has provided promising results in solving writer classification problems.

In patch strategy for deep face recognition [52], the authors propose a system that would take online cropped images as input for face recognition. Multi-branched CNN that learn from each patch and entire face representation is done by considering all the patches that are used. AlexNet and ResNet pre-trained CNN models are used for analyzing the effectiveness of the method. As an end to end training model, usage of both global and local features is done effectively. Six patches of size 136x136 pixels with facial key points from aligned face images were considered. These patches were passed onto pooling and convolutional layers. Feature fusion was accomplished by fully connected layers. This method boosts the performance of face recognition as it enhances the representation of local features.

In the research on “A Multiscale Patch Based Convolutional Network for Brain Tumour Segmentation” [53], the authors propose a system that would take multi-scale version of the patches as input. Batch normalization is performed after each convolutional layer. Convolutional layers of size 3x3x3 are combined by kernels of size 1x1x1. Exponential linear unit function is used as the activation function. Down sampling is carried out by decimating smooth version and up sampling is carried out by nearest neighbor interpolation. The proposed system yields smooth and compact segmentation results.

In the paper on “Patch-based Deep Learning Architectures for Sparse Annotated Very High-Resolution Datasets” [54], comparison of deep learning patch-based frameworks such as ConvNet, AlexNet and VGG models was carried by training and testing these models with publicly available, high resolution datasets. Varied patch dimensions such as 11x11, 21x21, 29x29, 33x33, 45x45 are considered for comparing the accuracy rates and also to choose the appropriate patch size for the model. Small patch size turned out to affect the quality and robustness of features in deep layers.

The paper “Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification” [55] proposes patch-based deep learning approach to explore subtypes of cancer. Even though CNN has acquired prominence in image classification, handling high resolution images implies high computational cost. Training CNN directly with Whole Slide Image (WSI) of size in gigabytes would lead to down sampling and data inefficiency. Hence, patch-based CNN model for lung cancer subtype classification was proposed by Le Hou. Experiment depicts that patch-based classification model is better or similar to image level classifier in classifying the sub-type of cancer with accuracy similar to that of pathologists.

4. WHOLE SLIDE IMAGE ANALYSIS

Conventional glass slides are scanned to create digital slides which is referred to as Whole Slide Images (WSI). Fig. 5(a) shows WSI and its corresponding patches. These images have gained beneficiary results in the field of education, diagnosis and research. WSI has avoided variance of slide quality by reproducing the same image with the exact orientation. Due to its high image resolution, WSI has provided an opportunity of feasible diagnosis for researches [56]. A digital WSI is represented in the form of a pyramid with different magnification levels. Level 0 is of the highest resolution and the following levels are down sampled version of the previous level, that is, the resolution of the images degrades as it goes further [57]. Pyramid representation of WSI image is shown in Fig. 5(b).

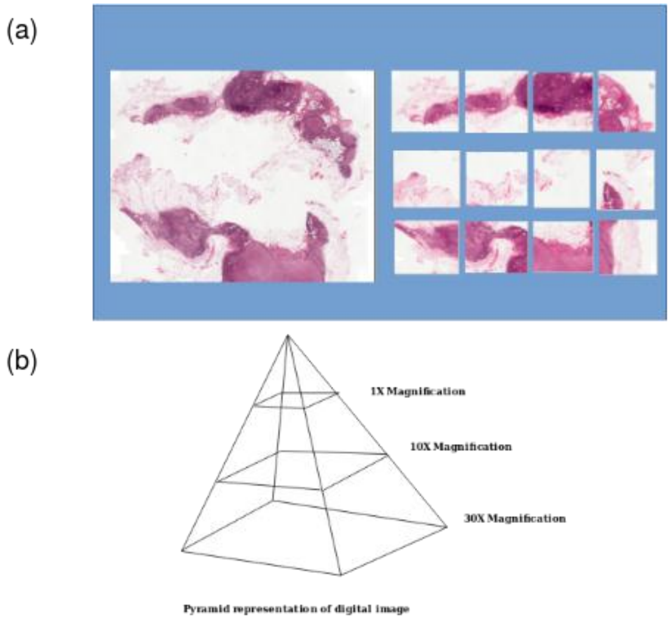

Fig. 5: WSI and its Corresponding Patches and Levels

As computing resources such as processing power, advanced software are easily available, digital images have gained wide variety of applications in pathology.There are many challenges that have to be addressed while utilizing the WSI. This is because each WSI will occupy large storage space due to its high resolution. Size of WSI ranges from a few hundred Megabytes to a few Gigabytes. Hence, storage, transmission and interoperability of WSI are challenging tasks. WSI acquired from different microscopic instruments may have different resolutions and scales of magnification. The format specification of WSI is not universal which leads to a conflict in viewing, analyzing, accessing with a particular software [59]. Even though WSI enables easy processing facilities of pathological images, these are some of the complexities in handling those images [58].To handle WSI many software packages/libraries are available. Few of them are listed below.

OpenSlide

OpenSlide is a C library that is not owned or maintained by any one company and hence it is vendor neutral which supports reading and manipulating digital slides of various formats. OpenSlide is written in C language and can be linked to the application written in any other programming language such as Java, Python. Therefore, with all these features, it is an open source library and has support from Linux, Windows and Mac distributions. Many applications such Diamond, PathFind, Slide Tutor have used OpenSlide for handling digital images [59].

SlideJ

SlideJ is pluginto analyses digital slide images. ImageJ is open source software for analysis of biomedical images. Through its macros and plugins, it allows user not only to analyze and process the images but also to extend customization of the same. ImageJ most importantly consists of a java module that can access various biomedical image formats. It also provides facility to read full digital slides and extract the cropped images at certain specified magnification levels and,thus, help in processing the images [60].

QuPath

QuPath is a software built using C++ and it is extremely user friendly with an interactive front end. QuPath provides panel to identify tumor. It facilitates researchers to batch process images, include scripting, share and develop new algorithms to analyze whole slide images [61].

Other than these, software and libraries such as PySlide, ASAP, Slide Tutor, Diamond and many more are available to handle whole slide images.

5. SUPPORTIVE LIBRARIES AND API'Ss FOR DL

5.1 TensorFlow

TensorFlow is a software which has been made open source by the Google developer's community and it is mainly used for carrying out large numerical computations using flow graphs and it is the most popular DL framework which makes use of programmatic approach to create networks. It is flexible to write code in many languages like Python, JavaScript, C++, etc. using TensorFlow. It is just a tool for assembling and evaluating computational graphs. It provides API that makes easy to build a neural network [36].

TensorFlow can run the graph on multiple hardware platforms, including CPU, GPU and TPU. It accepts the data in the form of multi-dimensional arrays of higher dimensions and ranks that are fed as input to the neural network called Tensors. Execution mechanism is in the form of graph that makes easier to execute the code in the distributed manner across the cluster of computers.

Data is stored in tensors. Once data is stored, computation happens in the form of graph. To work with TensorFlow, graph has to be prepared so that it can be executed. Code is written to prepare a graph and execute that graph by creating the session. Each computation in TensorFlow is represented as a data flow graph where each node in graph represents a mathematical operation and each edge represents multi-dimensional array. TensorFlow is used as backend for Keras [63].

5.2 Keras

Keras is a Python library which have a battalion of DL modules that can either run on TensorFlow or Theano backends. CNN based on Keras and TensorFlow using Python are used for binary classification. Pre-processing takes place using these frameworks. Keras has simplest model known as sequential model. Keras can be installed on different platforms [62].

Some useful dependencies for Keras are NumPy, SciPy, Scikit-learn, Pillow library for image processing, and h5py library for data serialization for saving model. Theano and TensorFlow are used as backend for Keras. It uses Theano and TensorFlow to perform very efficient computations on tensors. Tensors are the basic building blocks for creating neural networks, which is a multi-dimensional array used to fed as an input to neural network.

There are two ways for composing models:sequential and functional. In sequential model, different predefined models are stacked together in a linear pipeline of layers. Functional modules are via the functional API, where it is possible to define complex models, such as directed acyclic graph, models with shared layers, or multiple-output models.

Keras has a number of prebuilt layers for neural network. Some of the layers are:

- Regular dense, fully connected neural network layer.
- RNNs are recurring neural networks that simply make use of the nature of their input when it is sequential
- Convolutional and pooling layers of CNN are a variation of ANNs that use convolutional and pooling layer for learning deeply instead of using complex training models or algorithms and it concentrates on learning step by step iteratively instead of trying to learn everything at once.
- Regularization is a way to prevent overfitting some of the commonly used parameters for dense and convolutional models are kernal_regularizer, bias_regularizer, activity_regularizer. In addition, dropout for regularization can also be used.
- Batch normalization is a way to accelerate learning and generally achieve better accuracy model architectures and model parameters can be saved and loaded easily [63].

5.3 PyTorch

PyTorch is a Python module or package that provides Graphical Processor Unit accelerated tensor computation and also some higher-level functionalities which helps to build DL models or deep neural networks. For instance, it provides mechanisms to change the model at run time, debug easily with any python debugger.

PyTorch has two main features:

- Tensor computation with strong GPU acceleration
- Automatic differentiation for building and training neural network.

Other Python libraries like TensorFlow, the entire computational graph is defined first before running the model, whereas PyTorch allows to define graph dynamically. PyTorch Tensors are very similar to NumPy arrays with the addition that they can run on the GPU. It helps to accelerate numerical computations which can increase the speed of the neural network.

PyTorch uses a technique called automatic differentiation that numerically evaluates the derivative of the function. This automatic differentiation that computes backward passes in neural networks. Backward pass is the process which manually computes the gradients and forward pass computes prediction and loss. The torch autograd is the library that supports automatic differentiation in PyTorch. There are three level of abstraction in PyTorch:

- Tensors is an array or a particular data type to be more specific and it can run on a GPU.
- Variable in a DL structure stores graph or some data for computation.
- A module specifically in the ANN domain is a structure that can store a pre-learnt weight for parameters.

TorchVision library is a part of PyTorch which contains popular datasets, model architectures and image transformations for computer vision. Some of the datasets that it contains are MNIST, CIFAR, Imagenet-12, etc. Models architectures are Alexnet, ResNet, VGG, etc.,some of the transformations are transformation on PIL image, Conversion Transformation, Generic Transformation, etc [64].

5.4 Scikit-learn

Entire library of both supervised learning methods and algorithms, and an unsupervised learning algorithm with a Python library is called Scikit-learn. This extensive library is basically built on a pre-existing library known as SciPy short for Scientific Python. In order to work with this library some of the dependencies need to be installed like NumPy, SciPy, Matplotlib, Pandas, etc. Module that is built upon SciPy is named as Scikit and provides the learning algorithm and is named as Scikit-learn. This library specifically focuses on the modelling of the data and not on manipulating or summarizing it. Here are some of the most popular functions or options provided by Scikit-learn:

- Clustering is provided to group the unlabeled data
- For cross validating, unseen data availability of separate modules is listed.
- Database can be generated to test a particular model.
- There are also supervised models which includes naïve Bayes, SVMs, generalized linear model and many such things [65].

5.3 Caffe

Caffe is DL framework which is written in C++ and has Python and MATLAB bindings, and it is also open source. Different types of DL architectures which are specified for image classification and also image segmentation are also supported by this. It also supports CNN, Region-Convolutional Neural Network (RCNN) and FCNN designs. NVIDIA cuDNN and IntelMKL which are nothing but CPU and GPU based acceleration kernel modules which find a lot of support in this framework. CaffeOnSpark is a distributed DL framework which is integrated with Apache Spark. Pre-trained weights, optimization settings and the model definitions are also offered by the Apache

Spark.The layers in the Caffe framework should be defined with pre-existing parameters by the model creator and it is also very tough to add new layers in the Caffe framework once a model is created, whereas in the other frameworks like TensorFlow the layers are created dynamically [66].

5.5 Theano

Theano is used for dealing, evaluating, defining multi-dimensional arrays. This extensive library is basically built on a pre-existing library known as SciPy short for Scientific Python. It is also used to analyses, manipulate expressions, matrix-valued expressions, building symbolic graphs automatically and compiling parts of numeric expression into CPU or GPU instructions.

Some of the features of Theano are:

- Theano uses the ND array option from the NumPy library of the Python repository.
- Theano allows to perform intensive computation tasks on GPU instead of CPU, which helps it reduce the computation time.
- It performs differentiation for the user functions having one or many inputs.
- C code is also dynamically generated.
- It does extensive testing and checks for varied types of input [67][68].

5.6 Apache MXNet

It is one of the most popularly used open source frameworks to train the DL neural networks and other networks. This framework supports multiple programming languages and it also helps to reduce the computational and time resources required for the training process. Amazon Web Services and Microsoft Azure are also supporting this framework on their cloud platforms. At its core, it provides a flexible and comprehensive structure. The API of the Python wrapper of the MXNet's Python API primarily has two higher-level modules:The Gluon API package and Module API.

High level interface provided to make use of the MXNet is the Gluon API package. Both symbolic and imperative programming languages are supported by this Gluon package.This makes the implementation of the DL packages easy using Python and later on it can be deployed using a graphic and symbolic language like Scala or C++. A simple API is provided for DL, which is clear and concise. Simple prototypes can also be easily built using the Gluon API and this will help people is rapid development of product prototypes and also its deployment [69].

6. DISCUSSION AND CONCLUSION

As a part of this survey, papers related to various related fields of machine learning, image processing, DL, WSI analysis and computer vision were surveyed. The topics of the papers which were surveyed ranged from similar direct implementation of machine learning and computer vision for cancer detection, to specific fields of the same domains which gave clear understanding of the said topic. Moving further the frontier of computer vision that is CNN which is booming in the field of biomedical analysis was explored.Various high-level APIs and backend infrastructure were studied. Some of the popular APIs were also explored and listed down during the survey. The workings and other information of these technologies were briefed in the latter half of the survey. Selection of these techniques depends upon the distribution of data, factors that contribute to the type of image acquired, biomarker impact on feature extraction, image magnification level, patch size and the amount of data available for training.

REFERENCES

[1] Robertson, S., Aziz pour, H., Smith, K. and Hartman J., 2018. Digital image analysis in breast pathology from image processing techniques to artificial intelligence. Translational Research, 194, pp.19-35.

[2] Wang, C.W., Lee, Y.C., Calista, E., Zhou, F., Zhu, H., Suzuki, R., Komura, D., Ishikawa, S. And Cheng, S.P., 2017. A benchmark for comparing precision medicine methods in thyroid cancer diagnosis using tissue microarrays. Bioinformatics, 34(10), pp.1767-1773.

[3] Maier, A., Syben, C., Lesser, T. and Riess, C., 2019. A gentle introduction to deep learning in medical image processing. ZeitschriftfürMedizinischePhysik.

[4] Anuradha,K.andSankaranarayanan, K., 2012. Detection of oral tumor based on marker-controlled watershed algorithm. International Journal of Computer Applications, 52(2).

[5] Rejintal, A. and Aswini, N., 2016, May. Image processing-based leukemia cancer cell detection. In 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT) (pp. 471-474). IEEE.

[6] Bhattacharjee, S., Mukherjee, J., Nag, S., Maitra, I.K. and Bandyopadhyay,S.K., 2014.Reviewon histopathological slide analysis using digital microscopy. International Journal of Advanced Science and Technology, 62, pp.65-96.

[7] Kumar, N.R. and Kumar, J.U., 2015. A spatial mean and median filter for noise removal in digital images. Int J Adv Res Electr Electron InstrumEng, 4(1).

[8] Al-amri, S.S., Kalyankar, N.V. and Khamitkar, S.D., 2010. Linear and non-linear contrast enhancement image. International Journal of Computer Science and Network Security, 10(2), pp.139-143.

[9] Senthil Kumar, K., Venkatalakshmi, K. and Karthikeyan, K., 2019. Lung CancerDetectionUsingImage Segmentation by means of Various Evolutionary Algorithms. Computational and mathematical methods in medicine, 2019.

[10] Belsare,A.D.andMushrif, M.M., 2012. Histopathological image analysis using image processing techniques: An

overview. Signal & Image Processing, 3(4), p.23.

[11] Kaur, D. and Kaur, Y., 2014. Various image segmentation techniques: a review. International Journal of Computer Science and Mobile Computing, 3(5), pp.809-814.

[12] Irshad, H., Veillard, A., Roux, L. and Racoceanu, D., 2014. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review current status and future potential. IEEE reviews in biomedical engineering, 7, pp.97-114.

[13] Ali, S. and Madabhushi, A., 2012. An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. IEEE transactions on medical imaging, 31(7), pp.1448-1460.

[14] Ali, S. and Madabhushi, A., 2011, March. Active contour for overlap resolution using watershed-basedinitialization(ACOReW):Applicationsto histopathology. In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 614-617). IEEE.

[15] Mitra, R. and Menaka, R., 2016. Characterization of Oral Cancer Lesions Using Texture Features. Internat J Rec InovTrand Comp Comm, 4(6), pp.33-35

[16] Jain, S. and Bhusri, S., CAD System for Non-Small Cell Lung Carcinoma using Laws' Mask Analysis.

[17] Di Cataldo, S. and Ficarra, E., 2017. Mining textural knowledge in biological images: Applications, methods and trends. Computational and structural biotechnology journal, 15, pp.56-67.

[18] Komura, D. and Ishikawa, S., 2018. Machine learning methods for histopathological image analysis. Computational and structural biotechnology journal, 16, pp.34-42.

[19] Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V. and Fotiadis, D.I., 2015. Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, pp.8-17

[20] Nahid, A.A. and Kong, Y., 2017. Involvement of machine learning for breast cancer image classification: a survey. Computational and mathematical methods in medicine, 2017

[21] Hamilton, P.W., Wang, Y. and McCullough, S.J., 2012. Virtual microscopy and digital pathology in training and education. Apmis, 120(4), pp.305-315

[22] Shams, W.K. and Htike, Z.Z., 2017. Oral Cancer Prediction Using Gene Expression Profiling and Machine Learning. International Journal of Applied Engineering Research, 12(15), pp.4893-4898.

[23] Gonzalez, R.C. and Woods, R.E., 2002. Digital image processing

[24] Kulhalli, K.V., Patil, V.T. and Udupi, V.R., 2012. Image Processing Application for Computer aided Diagnosis of Cancer.InProceedings of the International Conference on Advancesin Computing and Management (pp. 297-301).

[25] Cruz, J.A. and Wishart, D.S., 2006. Applications of machine learning in cancer prediction and prognosis.Cancerinformatics,2, p.117693510600200030.

[26] Maier, A., Syben, C., Lasser, T. and Riess, C., 2019. A gentle introduction to deep learning in medical image processing. ZeitschriftfürMedizinische Physik.

[27] S Tandel, G., Biswas, M., G Kakde, O., Tiwari, A., S Suri, H., Turk, M., Laird, J.R., Asare, C.K., AAnkrah, A., N Khanna, N. and K Madhusudhan, B., 2019. A Review on a Deep Learning Perspective in Brain Cancer Classification. Cancers, 11(1), p.111.

[28] Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural networks, 61, pp.85-117

[29] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. and Dieleman, S., 2016. Mastering the game of Go with deep neural networks and tree search. nature, 529(7587), p.484.

[30] Edelen, A., Mayes, C., Bowring, D., Ratner, D., Adelmann, A., Ischebeck, R., Snuverink, J., Agapov, I., Kammering, R., Edelen, J. and Bazarov, I., 2018. Opportunities in Machine Learning for Particle Accelerators. arXiv preprint arXiv:1811.03172.

[31] De Schutter, E., 2018. Deep Learning and Computational Neuroscience.

[32] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), p.436.

[33] Lee, J.G., Jun, S., Cho, Y.W., Lee, H., Kim, G.B., Seo, J.B. and Kim, N., 2017. Deep learning in medical imaging: general overview. Korean journal of radiology, 18(4), pp.570-584.

[34] Burt, J.R., Torosdagli, N., Khosravan, N., RaviPrakash, H., Mortazi, A., Tissavirasingham, F., Hussein, S. and Bagci, U., 2018. Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. The British journal of radiology, 91(1089), p.20170545.

[35] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

[36] Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O. and Hajirasouliha, I., 2018. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. EBioMedicine, 27, pp.317-328.

[37] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[38] Wu, M., Yan, C., Liu, H., Liu, Q. and Yin, Y., 2018. Automatic classification of cervical cancer from cytological images by using convolutional neural network. Bioscience reports, 38(6), p.BSR20181769.

[39] Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In

Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
[40] Yamashita, R., Nishio, M., Do, R.K.G. and Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9(4), p.611.
[41] Choi, H., 2018. Deep learning in nuclear medicine and molecular imaging: current perspectives and future directions. Nuclear medicine and molecular imaging, 52(2), pp.109-118
[42] Wang, Z. and Fey, A.M., 2018. Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. International journal of computer assisted radiology and surgery, 13(12), pp.1959-1970.
[43] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke,V. and Rabinovich, A., 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
[44] Farooq, M. and Sazonov, E., 2017, April. Feature Extraction Using Deep Learning for Food Type Recognition. In International Conference on Bioinformatics and Biomedical Engineering (pp. 464-472). Springer, Cham.
[45] Dara, S., Tumma, P., Eluri, N.R. and Kancharla, G.R., 2018. Feature Extraction In Medical Images by Using Deep Learning Approach. International Journal of Pure and Applied Mathematics, 120(6), pp.305-312.
[46] Srinivas, M., Roy, D. and Mohan, C.K., 2016, March. Discriminative feature extraction from X-ray images using deep convolutional neural networks. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 917-921). IEEE.
[47] Alaslani, M.G. and Elrefaei, L.A., CONVOLUTIONAL NEURAL NETWORK BASED FEATURE EXTRACTION FOR IRIS RECOGNITION.
[48] Hou, X., Zhu, J., Sun, K., Shen, L. and Qiu, G., 2017, May. Object specific deep feature and its application to face detection. In 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA) (pp. 173-176). IEEE.
[49] Zheng, Y., Jiang, Z., Xie, F., Zhang, H., Ma, Y., Shi, H. and Zhao, Y., 2017. Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification. Pattern Recognition, 71, pp.14-25.
[50] Hegerath, A., Ney, I.H. and Seidl, T., 2006. Patch-based object recognition (Doctoral dissertation, Diploma thesis, Human Language Technology and Pattern Recognition Group, RWTH Aachen University, Aachen, Germany).
[51] Hattori, L.T., Gutoski, M., Aquino, N.M.R. and Lopes, H.S., Patch-Based Convolutional Neural Network for the Writer Classification Problem in Music Score Images.
[52] Zhang, Y., Shang, K., Wang, J., Li, N. and Zhang, M.M., 2018. Patch strategy for deep face recognition. IET Image Processing, 12(5), pp.819- 825
[53] Stawiaski, J., 2017. A Multiscale Patch Based Convolutional Network for Brain Tumor Segmentation. arXiv preprint arXiv:1710.02316.
[54] Papadomanolaki, M., Vakalopoulou, M. and Karantzalos, K., 2017, March. Patch-based deep learning architectures for sparse annotated very high-resolution datasets. In 2017 Joint Urban Remote Sensing Event (JURSE) (pp. 1-4). IEEE.
[55] Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E. and Saltz, J.H., 2016. Patch-based convolutional neural network for whole slide tissue image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2424-2433).
[56] Al-Janabi, S., Huisman, A. and Van Diest, P.J., 2012. Digital pathology: current status and future perspectives. Histopathology, 61(1), pp.1-9
[57] Goode, A., Gilbert, B., Harkes, J., Jukic, D. and Satyanarayanan, M., 2013. OpenSlide: A vendor-neutral software foundation for digital pathology. Journal of pathology informatics, 4.
[58] Boyce, B.F., 2015. Whole slide imaging: uses and limitations for surgical pathology and teaching. Biotechnic & Histochemistry, 90(5), pp.321-330.
[59] Farahani, N., Parwani, A.V. and Pantanowitz, L., 2015. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. Pathol Lab Med Int, 7, pp.23-33.
[60] Della Mea, V., Baroni, G.L., Pilutti, D. and Di Loreto, C., 2017. SlideJ: An ImageJ plugin for automated processing of whole slide images. PloS one, 12(7), p.e0180540.
[61] Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G. and James, J.A., 2017. QuPath: Open source software for digital pathology image analysis. Scientific reports, 7(1), p.16878.
[62] Chollet, F., 2018. Deep Learning mit Python und Keras: Das Praxis-HandbuchvomEntwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co. KG
[63] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. Tensorflow: A system for large-scale machine learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16) (pp. 265-283).
[64] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. and Lerer, A., 2017. Automatic differentiation in pytorch.
[65] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), pp.2825-2830
[66] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean,

J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. Tensorflow: A system for large-scale machine learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16) (pp. 265-283)

[67] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D. and Bengio, Y., 2010, June. Theano: a CPU and GPU math expression compiler. In Proceedings of the Python for scientific computing conference (SciPy) (Vol. 4, No. 3).

[68] Team, T.T.D., Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A. and Belopolsky, A., 2016. Theano: A Python framework for fast computation of mathematical expressions. arXiv preprint arXiv:1605.02688.

[69] Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C. and Zhang, Z., 2015. Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. arXiv preprint arXiv:1512.01274.

[70] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[71] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[72] Hinton, G., Vinyals, O. and Dean, J., 2015. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

[73] Nguyen, A., Yosinski, J. and Clune, J., 2015. Deep neural networksare easilyfooled: High confidence predictions for unrecognizable images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 427-436).