International Journal of Pure and Applied Mathematics
Volume 120 No. 6 2018, 305-312
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue
http://www.acadpubl.eu/hub/

Feature Extraction In Medical Images by Using Deep Learning Approach

Suresh Dara and Priyanka Tumma
Department of CSE
B V Raju Institute of Technology
(UGC Autonomous)
Narsapur, Telangana, India-502313.
darasuresh@live.in, priyanka.tumma65@gmail.com

Nageswara Rao Eluri
Department of Computer Science
King Khalid University
Abha, Kingdom Of Saudi arabia
eluri76@gmail.com

Gangadhara Rao Kancharla
Department of CSE
Acharya Nagarujuna University
Guntur, India-522510
kancherla123@gmail.com

***Abstract*—Deep learning is presently an effective research area in machine learning technique and pattern classification association. This has achieved big success in the areas of application namely computer vision, speech recognition, and NLP. This paper gives the impact of feature extraction that used in a deep learning technique such as Convolutional Neural Network (CNN). In this paper, feature extraction method is proposed and performed on medical images which CT scan Cancer datasetss. The experimental results have presented with proposed approach.**

***Index Terms*—Machine Learning, Deep Learning, Feature Extraction, Convolutional Neural Network(CNN), Multi-Layer Perceptron (MLP)**

I. Introduction

Conversion of given input data in to set of features are known as Feature Extraction. In machine learning, Feature Extraction begins with the initial set of consistent data and develops the borrowed values also called as features, expected for being descriptive and non-redundant, simplifies the consequent learning and observed steps. In few cases it points to be improved man-kind analysis. It is mainly associated with Dimensionality Reduction [1].

When the given input in the algorithm is very huge for handling and it is suspicious for being redundant, then that can convert in to decreased set of features. Certainly small group of the beginning features are known as feature selection. The preferred features are expected to get the applicable information from the given set of data, so that the task can be well executed by accepting the decreased data instead of the full given data.

Feature Extraction associates the decreasing the amount of assets needed to define a huge set of information. When performing analysis inquiry of complicated data, the main problem comes out from the sum of variables convoluted. Analysis with the big number of variables generally needs a big amount of memory and computation power, and also it access the algorithm of classification to overfill to training pattern to calculate to the new pattern. Feature Extraction is a common name for form of designing sequence of variables to come out of these problems, while still defining the data with enough efficiency.

The procedure to perform organic image to decreased model for promoting decision making such as detection of pattern, classification, recognition. An approach that decreases the amount of given data by extracting the detailed attributes is a procedure of assuming different features from the previously given features in order to decrease the cost of feature analysis, develop classifier accuracy and permit bigger classification efficiency.

The process by which a different selective feature is attained from those accessible input data. Classification is done by using the different group of features. This is the task which is used to achieve attributes which are peculiar. It is a mechanism for extracting consistent data from the image. After detecting a face, some important information is gathered from the images which are used in next step for identifying the image. The process of identifying and description of global or local properties of objects present in image.

A dimensionality contraction approach that identifies a decreased set of features that are the sequence of the initial ones. The process of obtaining relevant attributes that encloses with-in the given input data. The size of the given data will be finally decreased to preserve main information only. When the input is very huge to be processed, that information will be transformed into a decreased group of features.

A.Mueen et al. [2], proposed classification on new image method using multi-level image features and Machine learning method and state-of-the-art and support vector machine(svm). Here the author has extracted three levels of features global, local and pixel.Those will be combined together as one big feature vector.

K.P.Philip et al. [3], presented an algorithm which detects features in based on image on approximate geometrical models. This algorithm has based on the traditional and generalised Hough transforms that include notions from fuzzy set theory.By using this new algorithm it can be deeply estimated the exact location of boundaries surrounded by the organ and to determine the region around the organ.

Computer Aided Diagnosis(CAD) has been connected to the cerebrum CT image handling. Three classical kinds of features i.e., grey scale, shape and texture. Human cerebrum CT image is removed on the symmentric feature [4].

The remaining part of this article organised as follows:Section II discusses the background of Deep Learning Section III discusses the proposed methodology of Feature Extraction. The experimental results and discussion have discussed in Section IV. Finally, Section V concludes the article.

II. Background

A. Analysis of Deep Learning

Deep Learning further known as deep structured learning, and this, a chunk of a large group of Machine Learning approach depends on learning portrayal of an input, as disputed for a task of particular data. Learning can be broadly divided into three types: supervised, unsupervised or sometimes even semi-supervised [5].

Deep Learning is the group of Machine Learning algorithms [6]:

- It uses various layers of non linear that deals with the units of Feature Extraction and conversion. All consecutive layers considers the output from the before layer as the input.
- Learning to the various layers of representations will compares to various layers of abstraction;

Deep Learning algorithms seperate high-level,complex reflections as information portrayals through a progressive learning process.Complex dilebarations are learnt at a given inlight of generally more straight forward reflections figured in the first level in the chain of command. A key advantage of Deep Learning is the investigation and learning of gigantic measures of unsupervised information, making it a valuable tool for Big Data Analytics where crude inforamtion is to a great extent unlabeled and un-sorted. The investigation has done on how Deep Learning can be utilized for tending to some vital issues in Big Data Analytics, including extracting complex examples from huge volumes of information, semantic ordering, information labelling, quick data recovery, and simplifying discriminative undertakings.

In deep learning, Layers that are also used including have hidden layers of an ANN and arranged in formulae [7]. It also included suppressed variables created layer wise in profound models such as hubs in DBN and DBM.

There are two important reasons for deep learning that has recently become useful: i) It desires bulk of labeled data. ii) It desires generous computing power. High Performance GPUs will have lateral construction, will capable of Deep Learning. Although grouped with clusters or Cloud Computing, this set-up develops teams to decrease guidance time for a Deep Learning network from months to hours or even less.

It specifies for the class of Machine Learning approaches that will determine different layers of representations in deep-architectures. In this area, we will have an detailed analysis of strongly developed deep learning architectures, they are: 1)Deep-Belief Networks(DBNs) 2)Convolutional Neural Networks(CNNs) and 3)Multi-Layer Perceptrons(MLPs) [8].

Fig. 1 shows an overview of Deep-Belief Network (DBN). It is an abundant distinct model, otherwise the group of Deep Neural Network(DNN), suppressed with different labels of latent variables (”hidden units”), within the relation among the layers while no through the entities among the every layer [9].

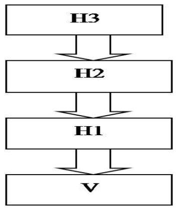


Fig. 1. Deep Belief Network

DBNs is seen as an architecture of smooth Un-supervised Networks called as Restricted Boltzmann Machines (RBMs) points every hidden layer of sub network's will treats as the Visible Layer(V) for the later on networks. A Restricted Boltzmann Machines is the random, abundant model with a hidden layer, the given input is a visible layer and connections are done in between the layers but not in between inner layers. This architecture tends for the quick, layer-by-layer Un-Supervised guidance method, where Contrastive Divergence(CD) was tested with every alternate network in round, beginning with the minimum match of layers.

Fig.2 shows a classic stack of RBM and single or multiple layers are added for discrimination tasks. Restricted Boltzmann Machines that determines the collective training information beyond applying the data labels because it is a probabilistic generative model [8]. They can definitely handle bulk of un-labeled input for applying convoluted data structures. When the arrangement of a Restricted Boltzmann Machines is discovered, then the aim isto know that the Weights in among the layers. It is managed mostly by an Un-Supervised Learning of RBM. A normal Restricted Boltzmann Machines contains within two layers: i) Nodes in single layer are totally associated to nodes in the another layer and ii)The nodes in same layers are not associated. Therefore, every node is not dependent of another node in the one single layer; Entire nodes are given in another layer. This characteristics grants for checking the abundant weights W of individual Restricted Boltzmann Machines by applying Gibbs sampling;

Before checking, a layer-by-layer of Restricted-Boltzmann Machines are observed: Output of the one layer (1stRBM) is treated as Input to the later layer (2ndRBM) and the procedure replays till the remaining RBMs are trained.

$$P(h_i = 1|v; X) = (\sum_{j=1}^{J} X_{ij} v_j + a_i) \quad (1)$$

And

$$P(v_j = 1|h; X) = (\sum_{i=1}^{I} X_{ij} h_i + b_j) \quad (2)$$

Where v means a $I \times 1$ unit vector of visible layer and h means a $Jx1$ unit vector of hidden layer, appropriately; W means the

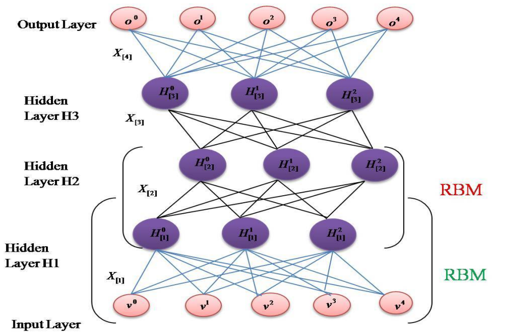


Fig. 2. A typical stack of Restricted-Boltzmann Machine

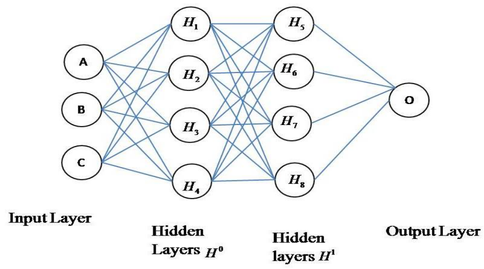


Fig. 3. Convolutional Neural Network

weights for matrix X_{ij} combining both the layers; ai and bj are bias items; and $(.)$ is the sigma function. When the visible units are real value, then the conditional probability distributions are quiet different: consistently, a Gaussian-Bernoulli distribution was assumed i.e., $P(v_j|h;X)$. Weights X_{ij} are modernized depends on a arrangement known as contrastive divergence (CD) approximation. For an example, the $(m+1)^{th}$ weight for X_{ij} been given below:

$$\Delta X_{ij}(m+1) = cw_{ij}(m) + [(v_j h_i)_{data} - (v_j h_i)_{model}] \quad (3)$$

Where refers to the learning rate, and c refers to the momentum factor; $(v_j h_i)_{data}, (v_j h_i)_{model}$were the assumptions below the probability distribution that are explained for the data and the model,appropriately. So the assumptions are observed by using Gibbs sampling boundlessly, practically, single- method CD is rarely used, because it functions good [10]. Another representative parameters are observed correspondingly. With the generative mode, the RBM training builds a Gibbs sample method to make simple hidden units depends on the visible units and conversely (1) and (2). The weights in between the visible layer and hidden layers are then observed applying the CD rule . The procedure will replays till convergence. The Restricted Boltzmann model data distribution applying convoluted units beyond handling labeled data.

The data regarding the given information is stored within the weights among each layer-by layer after pre-training. The Deep Belief Network will computes the last layer performs the effective outcome and the total network is tweaked by applying marked information and back-propagation approach for quick segregation.

Here are other alterations for pre-training: rather than applying RBMs, Additionally, modern conclusions showed that when a huge number of data that is to be trained is available, a complete Supervised training applying irregular original weights rather than the pre-trained weights will essentially performs good. For an example, a discriminative model with one individual hidden layer begins with a network (i.e., a shallow NN), that has been trained by back-propagation approach. Above convergence, a different hidden layer is added into the shallow Neural Network (among the starting hidden layer and the effective outcome layer) and the complete network is discriminatively trained again. These procedures will replays till all the hidden layers are trained.

To sum things up, Deep Belief Networks applies an anxious and dynamic layer-by layer strategy to decide the idle factors in each shrouded layer and a back-propagation approach for adjusting. This productive preparing approach builds up the generative execution and the segregative energy of this system. The following Fig.3, shows that the lower-level of this figure is treated as input layer with two-dimensional nxn models as proposal.

By regional approachable area, top-most neuron layers enlarge few elemental and composite optic features. Every convolutional layer in Fig.4 contains of different feature-maps, that is created by convolving instruction by various filters. This is pursued with the non-linear activation:

$$Y_j^{(l)} = f(\sum_i k_{ij} W x_i^{(l-1)} + b_j) \quad (4)$$

Where $Y_j^{(l)}$ is the j_{th} outcome of the i_{th} smooth layer $C_i; f(.)$ refers to the non-linear function, highest new applications applied as the extent hyperbolic tan rule like the non-linear activation function;$f(x) = 1.7159 tanh(2x/3).K_{ij}$ refers to the kernel filter in the filter bank that convolute by the feature map $x_i^{(l-1)}$ by the before layer to produces a modern feature maps with the present layer. W represents a various convolution operator and bj refers the bias. Every filter K_{ij} will combine to every chunk of feature maps with the before layer and decreases the dimensional decision of the feature map. Basically, every entity with the sub-sampling layer is composed by equating a 2×2 field in the feature map or by MAXPOOLING above the tiny region. The basic framework to be determined is the weights among layers, they are commonly trained by classic back-propagation method. This approach repeats for the gradient descent algorithm with mean squared-error as the loss function. The plan was closely inputs I with a continuous sequence of few essential and sparse functions [11].

$$S^* = arg||I - MS||_2^2 + \lambda|S|_1 + \alpha||S - D.\tanh(kx)||_2^2 \quad (5)$$

where M refers to the matrix with a linear basis set, S refers to the sparse coefficient matrix, D refers to the diagonal gain matrix and k refers to the filter bank with predictor framework. The target is for checking the excellent support function sets W and the filter bank k that diminishes the re-construction error (1st part of (5)) with a Sparse representation (the 2nd part of (5)), and the code prediction error respectively (the 3rd part in (5)).

A Multi-layer Perceptron (MLP) is the set of an ANN. An Multi-Layer Perceptron contains of minimum 3 layers of nodes. Excluding the given node, every node is the neuron that applies the non-linear activation function. Multi-Layer Perceptron handles a Supervised Learning approach called back-propagation for training the data. [12]

The Multi-Layer Perceptron contains three or more layers i)Input Layer ii)Output Layer iii)Hidden Layer. These are of non linearly-activated nodes that makes it a DNN. As MLPs are totally connected, every node with single layer combines by the particular weight w_{ij} to every node in the next layer. The two activation functions are both sigma functions, and are described by $x(v_i) - \tanh(v_i)$ and $x(v_i) = (1 + e^{-vi})^{-1}$.

The Initial part is a hyperbolic tan function that dimensions from -1 to 1, while the other is the log function, which are identical in configuration but dimensions from 0 to 1. Here vi is the outcome of the i^{th} node and v_i is the weighted sum of the given network. Learning occurred in the perceptron by developing the connected weights back to every bit of data are handled, depends on the load of fault in the outcome related to the original result. It refers to an example of Supervised Learning, and is drifted out with back-propagation. We represent the fault in outcome node j in the pth data point by $e_i(p) = d_i(p) - y_i(p)$, where d refers the target value and x refers to value composed by the perceptron. The node weights are adjusted depends on correlations that reduces the fault in the total outcome, given by $\epsilon(p) = d_i(p) - y_i(p)$. Using gradient descent, the change in each weight is: $\triangle w_{ji}(p) = -\frac{\partial \epsilon(p)}{\partial v_j(p)}$ where y_i is the outcome of the before neuron and refers to learning rate, that are preferred to establish the weights quickly coincide to a response, without oscillations. The derivative to be estimated based on the convinced regional area v_j, that by default varies. This is clearly proven that for the outcome node this derivative can be reduced to

$$-\frac{\partial \epsilon(n)}{\partial v_j(n)} = e_j(p)\phi(v_j(p))$$

Where refers to derivative of the activation function that represented above, that by default will not differ. The search is also crucial for the adjustment in weights to a hidden node, but that can be seen in the relevant derivative:

$$-\frac{\partial \epsilon(n)}{\partial v_j(n)} = \phi(v_j(p)) \sum_k -\frac{\partial \epsilon(p)}{\partial v_k(p)} w_{kj}(p)$$

This builds on the development in weights of the k^{th} nodes, which performs the outcome layer and so this algorithm performs the back-propagation of the activation function. [13]

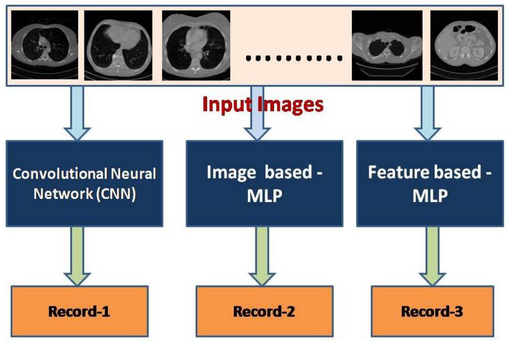


Fig. 4. Proposed methodology classification of image with and without deep learning

III. Proposed Approach for Feature Extraction

The analysis of this methodology is presented in below the Fig.4. In the proposed strategy, an orderly approach is displayed to direct the suitable trials to give solution for the exploration questions. The input image is divided into three unique models (1) a Convolution Neural Network (2) A MLP based picture (i.e., the picture is taken as the contribution of MLP), and (3) A MLP based component (i.e., contribution of the MLP is physically separated Features against the picture). As shown in Fig.4, the technique gathers the input images straightforwardly and after that apply those images to individual neural system show for characterization. Comparable parameter settings are utilized for making and checking the correctnesses, while most astounding precision among various neural system models is utilized to confirm the outcome, e.g., CNN. The individual models are discussed in the following section:

A. Automatic Feature Extraction based CNN Classifier

In this methodology, a comparable design of Le-Cun with a lightly altered version is used to compose the design. In convolution layer, formats of enrolled channels are utilised. Each one channel is limited spatially (traverse along with height and weight), but enlarges with complete deepness of input volume. The images that has, Height H, Depth D and Width W shading channels (i.e., $H \times D, W$), the enrolled channels isolates a image width as $W1 = \frac{(W-F+2p)}{S+1}$ here F speaks to the spatially expands neuron estimate; P is the main part of zero padding, and S is the size of way. Thus, the height is partitioned by $H1 = \frac{(H-F+2p)}{S+1}$ depth $D1$ is the extent of number of channels K. For instance, a image having $28 \times 28 \times 3$ (3 is for the shading channels), if the open field (or channel) has a size of $5 \times 5 \times 3$ (altogether 75neurons + 1bias), a 5x5 window with profundity three moves along the width and height and produces a 2-D activation map.

The Pooling Layer works individually above all the deepness portion for the input and rescales it extensional applying the MAX operation. It obtained the size of volume of HDW, and seperates the image into $W1 = \frac{(W-F)}{S+1}$ as Width and

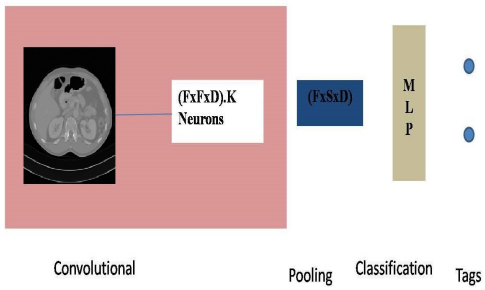


Fig. 5. Automatic Feature Extraction based CNN Classifier

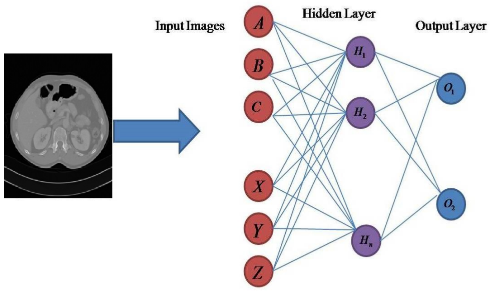


Fig. 6. An MLP based image

$H1 = \frac{(H-F)}{S+1}$ as Stature and profundity D1 is same as the info D. After the calculation against each shading channels the MAX task is finished. In this way, the feature matrix is then diminished in POOLING layer. . In the beneath layer, a completely completely connected network were utilized. Here, a MLP that depends on completely associated organize in CNN has been performed for final arrangement. Fig.5. demonstrates the design of a CNN with automatic feature extraction.

B. MLP based Image

In this proposed strategy, the complete picture is encouraged to a MLP based NN (Neural Network). Begining with an image information is exhibited and later the full image data is executed to the MLP. A conjugate gradient descent based back-propagation is require for preparing. Various neurons that are hidden and the training period of time are adjusted iteratively. The sketch of this approach is given in Fig. 6.

C. MLP based Feature

In this methodology, the vector of feature took out from the image is delivered to Multi-Layer Perceptron base neural network. The MLP based feature takes a shot on the feature matrix. A formula based human produced features/automated feature extraction is done before nourishing the input image to MLP. The conjugate gradient descent based back-propagation algorithm is utilised for the training. By contrasting to the MLP based image, MLP based feature works on relatively

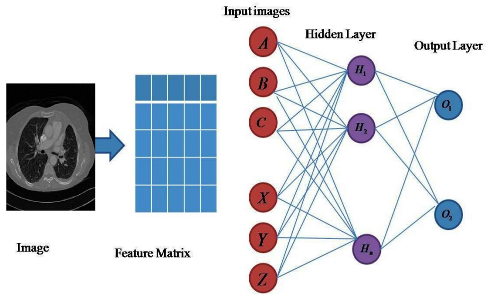


Fig. 7. MLP based Feature

restricted input limit because the features that are seperated from the images are less in number. The Clear information of this method is discussed [14] in Fig. 7.

Gonzalo Farias et al., [15] introduces a special NN known as Sparse Auto-Encoder for two classification problem of TJ-II fusion database. For all the experimental evaluation, the parameters for the SVM classifiers are constant. By using Auto-Encoder all the original features are reduced. The experimental results shows that the robust classifiers with a high successful rate is possible, apart from that the feature space is decreased to less than 0.02

Voshihiro Hayakawa et al., [16] implemented an experiment in which a persistent feature quantity was removed from the handwritten character/image utilizing multi layered NN. The author investigated whether feature space action of sandglass type for NN can be viably used. When analyzing a progression of repeated developments. The experimental result has proven that if a Gaussian Filter is connected to covered learning information as a form of preprocessing then the feature space is increased.

Sue Han Lee et al., [17] concentrated on method or design that augments the utilization of leaf databases for plant predictive modeling. The author introduced a new hybrid models handling the correspondence of various relevant data of leaf features by using MK dataset. The result has proven that the hybrid local global features learned using Deep Learning has improved acknowledgement execution related to previous technique.

Honggui Li et al., [18] gives the importance in between the smart-phone sensor data and personal health by a Deep Learning approach. The data grabed by smart-phone sensors is divided into sections by stacked Auto-Encoder. HAR dataset has been used for global categorizing accuracy for Automatic Feature Extraction and gained 0.15 percent. While a 0.5 percent has gained for manual Feature Extraction.

Heba Mohsen et al., [19] used DNN classifier which is one of the Deep Learning models for breaking down informational collection of 66brainMRIs into 4 classes. The new method architecture but needs less equipment detail and uses a comfortable time of altering for large size images. The experimental result has proved that by using the DNN classifiers appears high exactness related to traditional classifiers.

Bin Jiang et al., [20] proposed a continuous web cross media recovery strategy that relies upon Deep Learning. This method handles for the underlying assessment of picture acknowledgment tests to distinguish the requirements of the acknowledgment calculation. For large database of data, the analysis can be handled to correct the error rate of image, that will be helpful for machine learning. The results showed that the analysis has achieved high accuracy in picture content cross restorative, using less recovery time.

Cristina Nader Vasconcelos et al., [21] investigates in several aspects that how a Deep Convolutional Neural Network based classifier that occupied to arrange with a small and unbalanced biomedical data set. Although the best balance in between specificity and sensitivity is related for classifiers. The maximum sensitivity for this dataset has gained from 57

David Freire-Obregon et al., 2018 [22] proposed a Source Camera Identification (SCI) method for mobile devices that depend on Deep Learning. The author describes a CNN architecture that able to assume the noise pattern of mobile camera sensors to analyze the mobile device to capture the image. The model-level identification and sensor-level camera identification are two datasets used in this approach.

IV. Results

A. Datasets

We used Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD) data collection [23] in CT images, It is part of a larger effort to build a research community focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from The Cancer Genome Atlas (TCGA). Clinical, genetic, and pathological data resides in the Genomic Data Commons (GDC) Data Portal while the radiological data is stored on The Cancer Imaging Archive (TCIA).

The dataset a tiny subset of images from the cancer imaging archive associated with features like contrast and age. These datasets used to identify image textures, statistical patterns and features correlating strongly with these traits and possibly build simple tools for automatically classifying these images when they have been misclassified (or finding outliers which could be suspicious cases, bad measurements, or poorly calibrated machines). This results in 475 series from 69 different samples of patients. It is divided into 61 samples for training purpose and 8 samples for Testing purpose.

B. Discussions

In Fig. 8 and 9 showing experimental results. Here, the X- axis as Epochs and Y-axis as classification accuracy on Train and Test data. Fig. 8 showing that Training and Testing classifications accuracy results. Here, it reports 99% accuracy by using CNN. Fig. 9 shows that, image based results by using image based MLP. The obtained results are 93% accuracy, here we are not reporting feature based MLP results which almost similar to image based MLP.

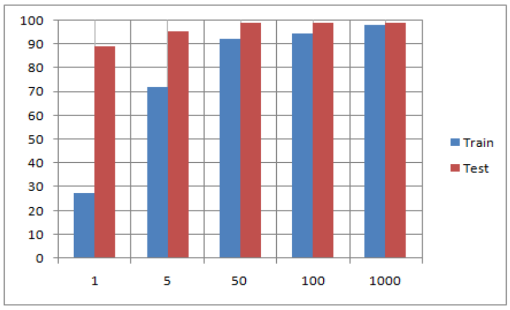


Fig. 8. Train and Test classification accuracy by using CNN on data TCGA-LUAD

V. Conclusion

Feature Extraction is about converting training data and establish it with extra features in order to make machine learning algorithms much adequate. This paper presented a technique, for finding the impact of automatic feature extraction and distribution that uses in Deep Learning such as Convolutional Neural Network (CNN). A path that has been suggested consistently to consider the classification efficiency of Convolutional Neural Network, image, and Feature based traditional Multi-Layer Perceptron (MLPs). Convolutional Neural Network with automatic feature extraction was first classified on a well-established to a traditional Multi-Layer Perceptron with full image, and a standard Feature Extraction were evaluated. In this paper, we proposed and presented results for feature extraction in medical images. The proposed approach results presented in form of classification accuracy.

References

[1] L. Deng, D. Yu *et al.*, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[2] A. Mueen, S. Baba, and R. Zainuddin, "Multilevel feature extraction and x-ray image classification," *Journal of Applied Sciences*, vol. 7, no. 8, pp. 1224–1229, 2007.

[3] K. P. Philip, E. L. Dove, D. D. McPherson, N. L. Gotteiner, W. Stanford, and K. B. Chandran, "The fuzzy hough transform-feature extraction in medical images," *IEEE Transactions on Medical Imaging*, vol. 13, no. 2, pp. 235–240, 1994.

[4] W.-L. Zhang and X.-Z. Wang, "Feature extraction and classification for human brain ct images," in *Machine Learning and Cybernetics, 2007 International Conference on*, vol. 2. IEEE, 2007, pp. 1155–1159.

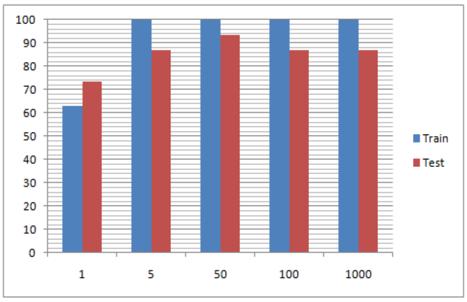
[5] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[6] B. B. Benuwa, Y. Z. Zhan, B. Ghansah, D. K. Wornyo, and F. Banaseka Kataka, "A review of deep machine learning," in *International Journal of Engineering Research in Africa*, vol. 24. Trans Tech Publ, 2016, pp. 124–136.

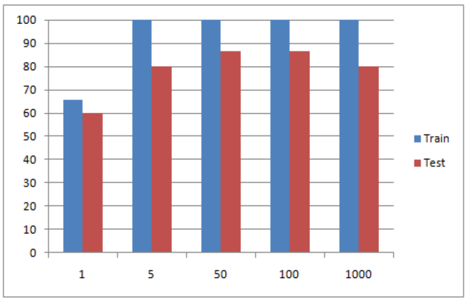
[7] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[8] G. E. Hinton, "A practical guide to training restricted boltzmann machines," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 599–619.

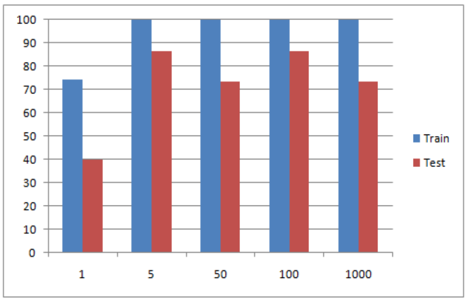
[9] A. G. Ivakhnenko, "Polynomial theory of complex systems," *IEEE transactions on Systems, Man, and Cybernetics*, no. 4, pp. 364–378, 1971.



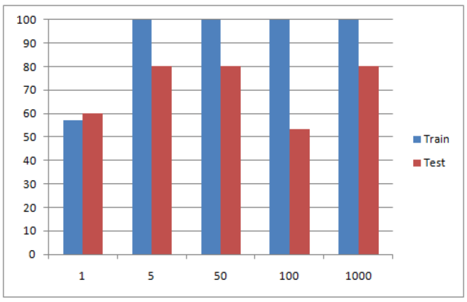
(a) At Hidden Neurons: 12



(b) At Hidden Neurons: 16



(c) At Hidden Neurons: 24



(d) At Hidden Neurons: 120

Fig. 9. Train and Test classification accuracy by feature based MLP on data TCGA-LUAD

[10] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[11] F. Rosenblatt, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms," CORNELL AERONAUTICAL LAB INC BUFFALO NY, Tech. Rep., 1961.

[12] S. Haykin and N. Network, "A comprehensive foundation," *Neural networks*, vol. 2, no. 2004, p. 41, 2004.

[13] E. Tatulli and T. Hueber, "Feature extraction using multimodal convolutional neural networks for visual speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2971–2975.

[14] J. Weng, N. Ahuja, and T. S. Huang, "Cresceptron: a self-organizing neural network which grows adaptively," in *Neural Networks, 1992. IJCNN., International Joint Conference on*, vol. 1. IEEE, 1992, pp. 576–581.

[15] G. Farias, S. Dormido-Canto, J. Vega, G. Rattá, H. Vargas, G. Hermosilla, L. Alfaro, and A. Valencia, "Automatic feature extraction in large fusion databases by using deep learning approach," *Fusion Engineering and Design*, vol. 112, pp. 979–983, 2016.

[16] Y. Hayakawa, T. Oonuma, H. Kobayashi, A. Takahashi, S. Chiba, and N. M. Fujiki, "Feature extraction of video using deep neural network," in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2016 IEEE 15th International Conference on*. IEEE, 2016, pp. 465–470.

[17] S. H. Lee, C. S. Chan, S. J. Mayo, and P. Remagnino, "How deep learning extracts and learns leaf features for plant classification," *Pattern Recognition*, vol. 71, pp. 1–13, 2017.

[18] M. N. K. Boulos, S. Wheeler, C. Tavares, and R. Jones, "How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from ecaalyx," *Biomedical engineering online*, vol. 10, no. 1, p. 24, 2011.

[19] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, 2017.

[20] B. Jiang, J. Yang, Z. Lv, K. Tian, Q. Meng, and Y. Yan, "Internet cross-media retrieval based on deep learning," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 356–366, 2017.

[21] C. N. Vasconcelos and B. N. Vasconcelos, "Experiments using deep learning for dermoscopy image analysis," *Pattern Recognition Letters*, 2017.

[22] D. Freire-Obregón, F. Narducci, S. Barra, and M. Castrillón-Santana, "Deep learning for source camera identification on mobile devices," *Pattern Recognition Letters*, 2018.

[23] B. Albertina, M. Watson, C. Holback, R. Jarosz, S. Kirk, Y. Lee, and J. Lemmerman, "Radiology data from the cancer genome atlas lung adenocarcinoma [tcga-luad] collection," *The Cancer Imaging Archive*, 2016.