Deep Learning Feature Extraction for Brain Tumor Characterization and Detection

Otman Basir[1], Kalifa Shantta[2#]
[1,2]Electrical and Computer Department, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada.

ABSTRACT

Deep Learning is a growing field of artificial intelligence that has become an operative research topic in a wide range of disciplines. Today we are witnessing the tangible successes of Deep Learning in our daily lives in various applications, including education, manufacturing, transportation, healthcare, military, and automotive, etc. Deep Learning is a subfield of Machine Learning that stems from Artificial Neural Networks, where a cascade of layers is employed to progressively extract higher-level features from the raw input and make predictive guesses about new data. This paper will discuss the effect of attribute extraction profoundly inherent in training approaches such as Convolutional Neural Networks (CNN). Furthermore, the paper aims to offer a study on Deep Learning techniques and attribute extraction methods that have appeared in the last few years. As the demand increases, considerable research in the attribute extraction assignment has become even more instrumental. Brain tumor characterization and detection will be used as a case study to demonstrate Deep Learning CNN's ability to achieve effective representational learning and tumor characterization.

INTRODUCTION

We can define feature extraction as the translation of some input data into a collection of attributes. Feature extraction commences with an opening series of reliable machine learning data, subsequently developing the adopted standards (aspects). It is anticipated that such features are non-redundant and descriptive. It simplifies the experimental stages and meaningful learning when the data input into an algorithm is too large to be managed and thought to be redundant. (a significant volume of data, but little information). The information will be improved into a lesser representation group of factors, known as the features vector.

Consequently, the extraction of discriminatory features in the signal improves the reduction of the data vector length by disregarding redundancy in the data and by concentrating the appropriate information into a feature vector of a considerably smaller dimension. Itis anticipated that the preferred features will define feature extraction as the translation of some input data into acquiring the relevant information from the provided data set. The resulting task is well executed by recognizing the reduced volume of data rather than being supplied the big data. Feature extraction corresponds to reducing the number of advantages required to describe a massive sequence of detail. In conducting an analysis inquiry of complex data, the convoluted variable's sum is the principal difficulty. It is usual that analysis with several variables requires considerable computational power and a large volume of memory and access the categorization algorithm to surfeit the training method to compute the new one. The expression "feature extraction" is a frequently applied term for a type of planning set of parameters that emerges from these difficulties while still defining the data with sufficient accuracy. The process of performing an organic image is a reduced paradigm for advocating decision-making; for example, recognition, classification, and detection of the pattern. A technique that reduces the provided data volume by removing the comprehensive characteristics presupposes various attributes from the earlier supplied features. This aims to reduce the feature analysis cost, allow greater categorization efficiency, and develop classifier precision. Another particular facet is achieved from such obtainable input data through this procedure. Classification is performed by applying another set of aspects used to attain categorization, such tasks being utilized to achieve unique properties. Such a mechanism is to reliable abstract features from the image. This is a dimensionality reduction method that recognizes a reduced group of attributes, being the series of the primary ones and the procedure of acquiring suitable characteristics that are enclosed inside the supplied input data.

Eventually, the magnitude of such data will be reduced to retain only the principal information. The main stage in numerous PC vision and item classification assignments is the extraction of useful features; consequently, several scientists have concentrated on designing robust features for multiple image grouping jobs. Considerable attention is now paid to convolutional neural networks (CNN) and feature-

learning algorithms. Every image is fed straight into the convolutional neural networks. Subsequently, the procedure abstracts its most significant features [1]. In one of the next sections, there are more details for CNN. This paper's layout is divided into the methodology, deep-learning history, Deep Learning background, a Literature review, feature extraction, deep learning analysis, a convolution neural network, followed by experiment results, and finally, the conclusion.

METHODOLOGY

The data sets contain T2-weighted 256×256 in-plane resolution. The first benchmark data set comprises 66 (18 benign and 48 malign) brain MRI images. The second dataset is 160 brain MRI pictures, where 20 are benign and 140 malign. 38benign and 188 malign images. The third data set download contains 37 normal brain MR images. The total dataset is 85 and 188 brain MRIs, normal and abnormal images, respectively. Since this is a small dataset, there were not enough MRI images to train the convolutional neural network. Therefore, data augmentation was useful to overcome the data imbalance issue in the data set. After data augmentation, the dataset now consists of 850 benign and 1128 malign MRI images, an overall 1978 MRI images. The dataset is divided into three groups: the data for training is 70%, the data for validation is 15%, and the data for testing is15%. The MRI resized to 227*227. The Alexnet CNN was constructed to distinguish the MRI brain scans as benign or malign, which involved 5 convolutional layers and 3 pooling layers, and finally, fully connected layers followed by the classification layer. There are 4096 neurons at the fully connected layer. There are two classes labeled as 0 and 1, representing the normal and abnormal brain image. The 3.6 Python programming language under Anaconda platform software was used to implement the CNN algorithm. We will compare and evaluate the CNN algorithm results with the results obtained in previous work "Brain Tumor Diagnosis Support System: A Decision Fusion Framework" [2]. In previous work, the applied classifiers were K-nearest neighbor KNN, support vector machine SVM, and artificial neuron network ANN to detect the brain tumor employing MRI images. The number of the classic classifiers' features was 13 features after applying discrete wavelet transform DWT and principal component analyses of PCA for each MRI image. The 3.8 Weka platform software uses to implement the traditional classifiers.

DEEP LEARNING HISTORY

For many centuries, wise people have dreamt of constructing a machine with the power to imitate the human brain. Rina Dechter introduced the expression "deep learning," and its origin dates from 1986. The term "neocognitron," which was introduced in 1980, inspired the convolutional neural network [3]; however, recurrent Neural Networks (RNNs) were suggested in 1986 [4]. Following this, in the 1990s, LeNet enabled the deep neural networks (DNNs) to apply in practice, although this was not generally accepted [5]. The hardware restriction causes the LeNet structure to be rather simplistic; therefore, it cannot be used with large datasets. Deep belief networks (DBNs), together with a layer-wise pre-training structure, were introduced in about 2006 [6]. Their principal objectives were to learn a basic two-layer non-learning paradigm, such as limited Boltzmann machines (RBMs), to freeze every parameter, place a new layer at the highest level, and train only these variables for the new layer. This method enabled researchers to train more profound neural networks than was previously possible, thereby motivating a rebranding of neural networks to deep learning. After several decades of advancement, deep learning that originated from artificial ANNs became one of the most effective instruments compared with other well-performing machine-learning algorithms. It has been observed that some deep-learning techniques developed from the first ANNs, which included RNNs, RBMs, DBNs, and convolutional neural networks (CNNs) [7]. Although graphics processing units (GPUs) are famous for their action in the computation of large-scale frameworks in organizing structures on one mechanism, there has been the development of numerous distributed deep-learning networks to accelerate the training of deep-learning paradigms [8]. Large volumes of data have labels or even noisy labels; therefore, specific research attaches greater importance to enhancing training modules' noise strength by applying unmonitored or semi-supervised deep-learning approaches. The max of the existing deep-learning paradigms concentrates exclusively on one modality, resulting in a restricted representation of real-world data. At present, researchers are giving

greater concentration to a cross-modality framework, which can be a massive advancement in deep learning. Google's AlphaGo, an inspiring deep-learning application, totally surprised the world when it was launched at the beginning of 2017 [9]. It was propelled under the pseudonym "Master" and won sixty successive internet games with specialist human Go players, including three wins over Ke Jie from 29 December 2016 to 4 January 2017. AlphaGo utilizes sufficient hardware resources and modern deep-learning algorithms. It can win against Go players who are global champions.

BACKGROUND

We apply machine learning to analyze an increasing volume of data, which, over time, has become considerably more complicated. Indeed, the emergence of deep learning during the past decade has helped generate learning paradigms that are becoming more effective. Several machine-learning jobs seek to classify difficulties, such systems working similarly to what is depicted in Figure 1.
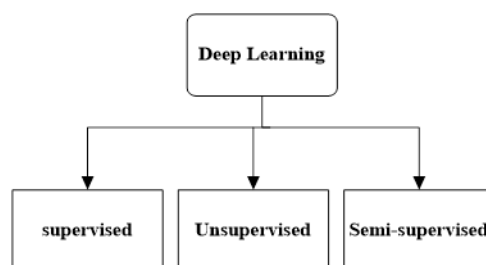

Fig.1. Machine Learning a Hierarchy

Firstly, an attribute extracted from the provided input data may generate a new representation of the data specific to the current job. Subsequently, a categorization method is learned and the features mentioned above to complete the task. When training is complete, this technique ought to be applied to data not seen in the course of the training stage, and should forecast its response precisely, and in this case, should also predict the class label. The input's extracted features were frequently handcrafted, particularly until recent times, meaning that such features are planned explicitly for the input data and the current task. Usually, these are not exclusively tied to the data type, such as handwritten words' images, but rather to a particular subset, such as images of handwritten English words in ink written on parchment. Usually, most such features cannot manage change well; however, machine learning is another method of extracting features from the data to learn a feature extractor.

Nevertheless, a learning system is constructed to extract attributes from the input rather than designing a method to categorize specific images. This indicates that the network is learning directly from the input pixels, features of a more significant level concerning images. Therefore, for numerous reasons, we believe that this is a better method of utilizing handcrafted features. It is possible to adjust the trained paradigm to various input types by applying such training to each data set; however, it may be necessary for each data set to be hand-tuned. Furthermore, a professional understanding of the images under analysis should not be required by this method.

LITERATURE REVIEW

The following section provides a literature review in the last few years in the deep learning feature extraction:

Muhammad et al. [10], suggest utilizing CNN for attribute extraction from food photos. Using a three-fold cross-validation plan on food image dataset for trained LSVM. The classification stage is done by used three attributes distinct fully connect layer of CNN. Two jobs clarify classification. The first job images are classified into sixty-one categories and the second into seven categories. The conclusion

shows that applied four thousand ninety-six attributes to obtain 70% and 90% accuracy for both classes, respectively.

Xiaojun Lu et al. [11], in this study, suggest a technique that utilizes attribute fusion to represent images better for face recognition by attribute extraction by a deep convolutional neural network (DCNN). They used PCA to reduce the capacity of the fused attribute. The SVM machine classifier is implemented for two classes. Experimental outcomes display that this technique can distinguish faces with severe occlusion and significant confusion and scale differences. The conclusion mentions that this technique reaches an 89% recall rate on FDDB and 97% average precision.

Er-Yang Huan et al. [12] suggest a body constitution recognition procedure grounded on CNN that can recognize person constitution kinds based to face pictures. The determined model first utilizes CNN to extract the face picture attributes and then fused the abstracted features with the hue attributes. The combined details are driven to the Soft-max classifier to obtain the grouping outcome. They state that the method suggested in this work can reach an accuracy of 65.3%

Bhandari et al. [13], Introduced a process to abstract brain tumors from 2D MRI brain using a Fuzzy C-Means grouping procedure followed by classical algorithms and CNN. The empirical research was conducted on an actual time dataset with different cancer dimensions, places, patterns, and various image strengths. The classical algorithm section applied six traditional classifiers, particularly Support vector machine, k-nearest neighbors, Multilayer perceptron, Logistic Regression, Naïve Bayes, and RF, whatever were applied in Scikit-learn. Then, applied CNN using Keras and Tensorflow as it produces more reliable achievements than regular ones. CNN obtained a precision of 97%.

Hong Liang et al. [14] research outlines the popular procedures utilized in paragraph attribute extraction first and then extend repeatedly used DL process in paragraph attribute extracted and its implementation and forecasts the application of deep learning in attribute abstraction. They conclude that related to other machine learning approaches. DL can distinguish complex interactions from the attribute and train lower-level attributes from almost unprocessed primary information.

Hawaii et al. [15], suggested a fully automatic brain tumor classification technique grounded on Deep Neural Networks. The planned algorithm was applied in glioblastoma illness images in different grades. A novel building of a convolutional network is offered. The flowing planning is offered in which the output of a core convolutional network is unapplied as an extra basis of data for the following convolutional network.

Matilda Lorentzon [16], in her master theses, "feature extraction for image selection using machine learning," the feature extraction methods used are histograms of oriented gradients, features from the discrete cosine transform domain, and attribute extracted from a pre-trained CNN. She concludes that attributes removed from CNN for the content classification provided the best results for the content classification.

Ronjian Li et al. [17], in their paper, DL grounded imaging data accomplishment for enhanced brain illness analysis. They suggested DL based frame for evaluating multi-modality imaging information. Alzheimer's illness images from two modalities PET &MRI were used to evaluate this technique. They conclude that results showed their method significantly outperformed prior methods.

Pingping Zhu [18], in this study, the convolutional neural network algorithm is applied to abstract target attributes from the sonar photos. The SVM is implemented in the recognition stage, which was trained based on manually labeled data. The result demonstrates that deep learning attribute extraction provides

good achievement compared to applying different feature extraction methods similar to the histogram of oriented gradients and local binary patterns.

Heba Mohsen [19], in their paper, brain tumors classification using learning neural networks. They applied the Deep Neural Network to categorize the MRI data set of 66 brain tumor images. They conclude that using the DNN algorithm demonstrates great accuracy related to standard classifiers.

Pereira et al. [20], auto distinguish technique is applied to classify cancer utilizing a convolutional network within three kernels. The technique achieved the full core's initial location concurrently and improved areas in dice likeness and quantity metric 0.87, 0.82, and 0.76.

Siar et al. [21], the CNN algorithm is applied to identify multiple sclerosis and normal tumors concurrently. The convolutional network was capable of categorizing 96% of the images into three groups correctly.
Szilagyi et al. [22] suggested Fuzzy C-Means to division brain cancers from MRI scans. They suggested procedure usages previous data at two points of the implementation. The groups of voxels formed by fuzzy C-means are categorized as maybe cancers and non-cancers grounded on data abstracted from train sizes; two is the selection of fuzzy-means factors (amount of groups, fuzzy exponent). They conclude that introductory assessment outcomes showed the procedure's capability to exact abstract information regarding cancer's existence and location.

Xu et al. [23], an effective and operative way utilizes a convolutional network applied for grouping and segmentation. The suggested approach applied Image-Net for abstract attributes. The outcomes achieved 97% and 84% precision for grouping and segmentation, respectively.

Pan et al. [24], DL buildings, and base neural systems for cancer categorizing by MRI images have been considered and evaluated. .The outcomes demonstrate that the system routine grounded on the Sensitivity and Specificity of the convolutional network better by 19% associated with ANN.

 Basheera et al.[25], In this work, They suggest a new method that utilizes CNN for classifying brain tumors into benign and three various kinds. The tumor is primarily segmented from the MRI images applying an improved ICA composite model. After the segmented image, deep attributes are extracted and organized. The outcomes have been validated by measuring the classifier's achievement on a data set free with Harvard Medical School.

Menegola et al. [26], complete several trials that differentiate learning from beginning to adjust-setting pre-trained networks and display that adjust-tuning function preferable presented a short information collection of approximately a thousand pictures of skin damages. Nevertheless, their workouts are minor rules to be capable of outcome for any typical result tasks.

Bhandari et al. [27], examine the role of CNN's in segment brain tumors by initially taking an informative look at CNN's and conducting dissertation research to discover an instance pipeline for segmentation. Also, Examine the future effectiveness of CNNs by searching for a new field—radionics. This explores the quantitative features of brain tumors such as shape, texture, and signal intensity to predict clinical results such as existence and treatment response.

Naima Otberdout et al. [28], the correlation matrix is employed to encrypt the DCNN attribute discrimination for facial appearance in this paper. The space geometry of the correlation matrix of Symmetric Positive. By performing the classification of the facial expressions, employ the Gaussian kernel on the SPD manifold. They show that attributes computed by DCNN are powerful and more efficient compared to standard classification. They implemented this approach using the VGG-face and

ExpNet architectures with large practices on the Oulu-CASIA and SFEW datasets. They conclude that the proposed technique reaches achievement for face appearance identification.

Hashemzehi et al. [29], a Novel hybrid model that contains neural autoregressive distribution estimation (NADE) and CNN is suggested for brain tumor detection. They tested the system with a total of 3064 T1 spin-lattice" relaxation time MRI by three kinds of brain tumors. The result shows that the mixture CNN-NADE has a significant classification showing that clinical images' obtainability is restricted.

Seetha et al. [30], suggest brain tumor detection by applying Convolutional Neural Networks (CNN) classification. Small kernels designed, and neuron weight to obtain more in-depth architecture. The weight of the neuron is specified as small. Investigational outcomes display that the CNN records 97% accuracy with little complication and the recent approaches.

Khan et al. [31], present the convolutional neural network (CNN) algorithm and Data Augmentation and image processing to categorize brain MRI scan images into malignant and non-malignant. They compare the achievement of the scratched CNN algorithm with pre-trained VGG-16, ResNet-50, and Inception-v3 models. The conclusion that model accuracy achieving 100%, while VGG-16 achieved 96%ResNet-50 achieved 89% and Inception-V3 achieved 75% accuracy.

FEATURE EXTRACTION
We apply machine learning to analyze a growing volume of data, which is becoming increasingly complicated. The emergence of deep learning during the past decade has undoubtedly helped generate learning paradigms that are becoming more effective. Numerous machine-learning jobs seek to categorize difficulties. Since features are extracted from the input data, it can be considered as a novel representation of the data, particularly for this task. Subsequently, in addition to these features, which complete the job, a categorization method is learned. This method should be applied to unobserved data in the training stage period upon completion of the training. It ought to give a precise forecast of its response precisely and in this situation the class label. Frequently, and particularly until recent times, the features extracted from the input were handcrafted, implying that they are specially designed for the input data and the present task. It is standard practice for these not to be exclusively tied to the data type; for instance, handwritten images of words' pictures, but rather to a specific subset, such as English words handwritten in ink on parchment. Usually, most such features cannot manage change well; nevertheless, machine learning is a different method of extracting features from the data to learn a feature extractor. However, a learning system is built to extract attributes from the input instead of planning a classifying image. This shows that the network is learning, from the input pixels directly, a greater level of features concerning ideas. Consequently, for many reasons, we regard this as an improved method of utilizing handcrafted features. It can adjust the trained paradigm to several input types by applying this training to each data set; nevertheless, it may need to hand-tune each data set for handcrafted features. However, a specialist understanding of the images being analyzed is not necessary for this approach. Concerning information retrieval and image analysis, feature extraction is a significant and fundamental issue. Although a considerable amount of time is needed to hand-design a useful feature, deep learning allows the acquisition of such attributes whose objective is new applications. Deep learning has attained much as a new feature extraction technique. Traditional systems and deep learning principally differ in that deep learning automatically acquires aspects from huge information rather than handcrafted attributes. This is generally dependent on the previous knowledge of designers, and it is certainly not possible to obtain an advantage from big data. It is possible for deep learning to automatically get attribute representation from big data, including millions of variables. Deep learning's principal benefit is that it is unnecessary to extract attributes from the image manually. During training, the network learns to pull elements by feeding the image to the system pixel values [32].

Convolution neural network construction is typically a series of feedforward layers that apply and pooling layers and convolutional filters. After the final pooling layer, CNN employs numerous completely associated layers that convert the previous layers' 2D feature maps of 1D vector for categorization [33]. Although one advantage of CNN architecture is that it does not need a feature extraction procedure before training, a CNN from first principles occupies much time. Moreover, it is not easy because it requires a significantly sizeable categorized dataset for training and constructing earlier the paradigm being prepared for grouping, which is not possible in every instance. Furthermore, the hardware needed to process numerous filters for greater-sized images is 256×256. Another extensively applied DL architecture used for regression or categorization in the deep neural network (DNN) has been successful in several fields. This is a standard feedforward network where the information moves from input to the output layers through numerous invisible layers, being a minimum of two [34].

DEEP LEARNING ANALYSIS

Deep structured learning is another name for deep learning, and it is a group from machine learning. Deep learning requires two substantial factors; one is a massive amount of classified data and powerful computing. It appoints the group of ML methods, which define several layers of performance in deep-designs. Next, a detailed examination of developed deep learning designs: Deep-Belief Networks (DBNs), Multi-Layer Perceptron's (MLPs), and Convolutional Neural Networks. The central design block is a bipartite undirected graphical model named the Restricted Boltzmann Machine (RBM). The (RBM) was original proposed by Hinton [35]. More technically, RBM is a stochastic neural model. Neural network purpose: We have neuron-like units with zero or one activation function based on the neighbors they're linked to; stochastic, meaning these activations have a like hood element. Figure 2 presented a graphical representation of the RBM.
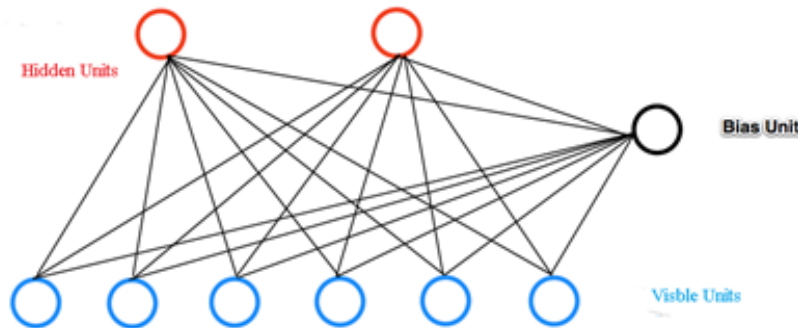

Fig.2. Graphical description of Restricted-Boltzmann Machine

The network is restricted because the no-hidden unit is connected to any other visible unit, and no invisible unit is connected to any other invisible unit. DBNs can be observed as a structure of single, unsupervised systems, i.e., RBMs + Sigmoid Belief Networks. DBNs' most significant success is their ability of "learning attributes," which is obtained by a 'layer-by-layer training procedure where the higher-level attributes are learned from the preceding layers. Figure.3 displays a stack of restricted Boltzmann machines, and single or multiple layers are summed for recognition jobs.
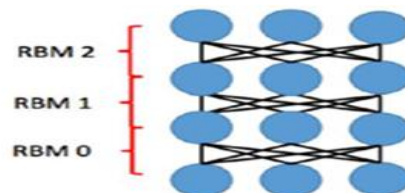
Fig. 3. Restricted-Boltzmann Machine.

1. DBNs can be formed by "stacking" RBMs.
2. Each layer is trained as a Restricted Boltzman Machine.
3. Train layers sequentially starting from the bottom (observed data) layer. (Greedy layer-wise).
4. Every layer learns a higher-level representation of the layer below.
5. The training criterion does not depend on the labels. (Unsupervised)[36].

It contains visual units, any 0 or 1 or real-number, and a collection of binary invisible junction. A structure through visual path v and unseen path h is given with energy specified:

$$E(s, h) = -\sum_{i \in visible} \sigma_i s_i - \sum_{j \in hidden} b_j h_j - \sum_{ij} s_i h_i w_{ij} \qquad (1)$$

Where α_i, b_j, w_{ij} are the system factors. Assumed this energy, the system assumes to each couple *s, h* likelihood:

$$P(s, h) = \frac{1}{Y} e^{-E(s,h)} \qquad (2)$$

All likely pairs of visual and invisible paths are summed to produce *Y* as a separation function.

It is a fast way to achieve unbiased pair as no direct link between visible and invisible units

$$P(h_j = \frac{1}{s} = \sigma(b_j + \sum_i s_i w_{ij}) \qquad (3)$$

the logistic sigmoid function is σ(·). Likewise, assumed an invisible path *h* the likelihood of a visual junction s_i to be consigned to one is specified:

$$P(s_i = \frac{1}{h} = \sigma(b_j + \sum_j h_j w_{ij}) \qquad (4)$$

To obtain as large as the probability of the visual and invisible vectors team {*s, h*}, the network parameters are tuned using the training data. RBMs are limited in the complication of the information they can appear as it is single two layers deep system. To mitigate this problem, several deeper paradigms constructed on Restricted Boltzman Machine are proposed. The deep belief networks (DBN) are the most ordinary model derived from RBMs [37], the DBM [38], and the DEM[39]. These algorithms are several layers likelihood systems that implement a non-linear conversion to the data. The greedy wise manner is used to train DBNs somewhere every layer learned as its RBM. Except for the top two that remain undirected, the last paradigm holds single the top-down connections of the layers. In all layers, DBMs have undirected weights, unlike DBNs. Similar to DBNs, the weights are learned in a greedy mode. To assessment and maximize the probability straightway is hugely computationally costly, Salah and Laroc [40] suggest an approximate technique that maximizes the lower limit of the log-probability. Lastly, DEM, the utmost modern deep paradigm setup on RBMs, is an entirely joined feedforward network with an RBM on head .it is likely to learn the full paradigm simultaneously as the non-stochastic nature of the unseen layers renders.

This practical learning method increases the obstetric achievement and the characteristic strength of this network.

A multilayer perceptron MLP is a deep, artificial neural network. It contains more than one perceptron. The input layer to collect the signal, the decision produced by the output layer, and the Qualitative

amount of invisible between input and output layers represent the computing tools of MLP. MLP is regularly useful for supervised learning difficulties. They practice collecting input-output sets and training to parade the correspondence (or dependencies) among the inputs and outputs. Figure 4 shows an architecture of MLP containing four hidden units and one output.
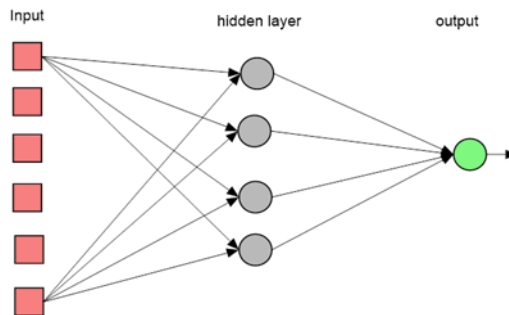

Fig. 4. Architecture of an MLP, containing four hidden units and one output.

The practice comprises tuning the weights and biases of the model to decrease deviation. Tuning both the weight and bias relative to the fault is done by the backpropagation technique, and the error is calculated in different methods. Several artificial neurons constructed the neural networks. Each input into the neuron has its weight related where weight is a real number required to tuning once we finally reach to learn algorithms. No limit input number to a neuron, from one to $i;$ the total input number denoted by i. Figure 5 demonstrates a mathematical model of a discrete perceptron or neuron.
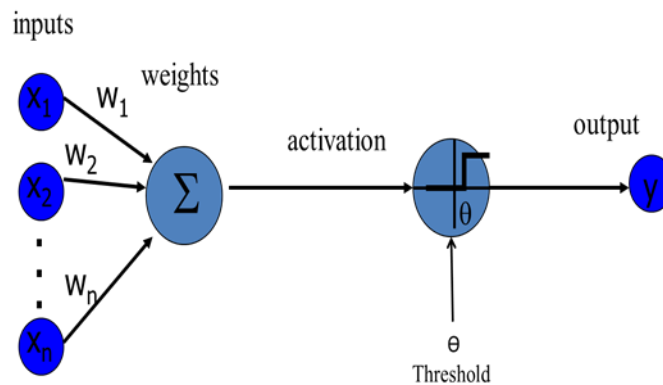

Fig. 5. A mathematical model of a discrete perceptron or neuron.

The inputs may be denoted:$x_1, x_2, ... x_i$. Moreover, the matching weights for the inputs denoted$w_1, w_2, ..$ w_n, and the output a $= x_1w_1 + x_2w_2... +x_nw_i$. In feedforward network: The neurons in every layer feed their output forward to the following layer until we get the last output from the neural network. There can be any amount of invisible layers inside a feedforward network, and the amount of neurons can be random. Other MLP properties of architecture:

1.   There is no direct connection from the input to the output layers.
2.   Completely connection among the layers.
3.   Often more than three layers.
4.   It is not required output and input units are equal.
5.   No connections within a layer.

6. No limit to the invisible units in each layer, and could be greater or less than input or output units.

The data flow starts at the input to the output layer after passing through the invisible layer. The output layer produced the output and compared it with the actual output value. Using the backpropagation method and the series basis of calculus, fractional byproducts of the error function w.r.t. the different scales and biases are backpropagated out of the Multi-layer perceptron.

This action of differential offers a gradient of error, over which the parameters could be tuned and proceed with the MLP one stair nearer to the lowest error. Any gradient-based optimization algorithm can be applied to perform this task, such as stochastic gradient descent. The model keeps doing this optimization until the error reaches the lowest. This case is famous for convergence. MLPs are connected, each junction with a single layer join by the special weight w_{ij} to each junction in the following layer. Two sigma functions represent the activation function, and are defined by $x(s_i)$ - tanh(s_i) and $x(v_i)$ – tanh(s_i) . The primary piece is a hyperbolic tangent role that dimensions from minus one to one. Simultaneously, another is the logarithm role that is alike in shape unless dimensions starting zero to one. The result of the jth junction is s_i, and v_i the weighted calculation of the setnet.

CONVOLUTIONALNEURAL NETWORK
Convolutional neural networks (CNN) are constituents of appropriate neural network technique classification. As well as having the capacity to learn image feature representations automatically, CNN has as well as exceed many traditional handcrafted feature methods [41]. A convolutional neural network (CNN) depends on following one another locally connected convolutional layers, each of which has an equal number of filters, being down-sampling layers, and the fully connected layers that serve as classifiers [41]. Figure 6 shows the suggested Convolutional neural network for brain tumor detection employing MRI photos. The following three notions of convolutional neural networks' architecture result in it being efficient: local receptive fields, weight sharing, and down-sampling operations [42]. The local receptive field indicates that every neuron recognizes input from a minor portion of the preceding layer, with a convolution filter of the same magnitude.
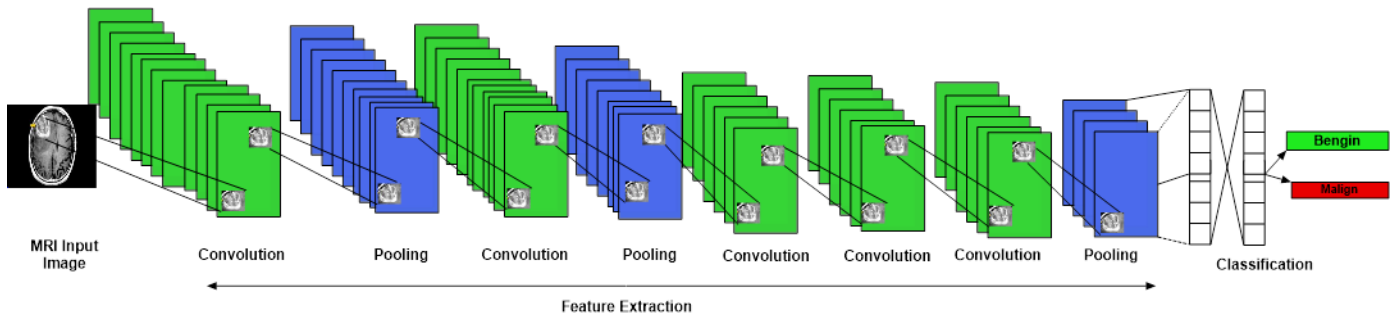

Fig. 6. Proposed CNN architecture to brain tumor detection.

Additionally, local receptive fields are used in convolutional and downsampling layers. The convolutional layer uses weight sharing to regulate the paradigm's capacity and decrease its difficulty. Lastly, the non-linear downsampling is applied to reduce the image's spatial size and decrease the paradigm's free parameters. These concepts help CNN to be strong and effective in recognition tasks. In a different aspect, the Convolutional Neural Networks layers are:

1. The Convolutional Layer: A convolutional layer contains many channels whose parameters should be learned. The stature and weight of the channels are littler than those of the information volume. Each channel is convolved with the info volume to process an actuation map made of

neurons. The channel is slid over the information's width and stature, and the speck items between the info and channel are registered at each spatial position. The first convolutional layer containing six filters produced six feature maps arranged together. Every feature map represents specific features from the image; for example, represented points or defined vertical edges [43]. The convolution operation is described:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_i^l + b_j^l\right) \quad (5)$$

Here j is the specific convolution feature map, M_j is a selection of input maps, k_i^l is the filter, b_j is the feature map bias, l is the layer in the CNN, the activation function is denoted by f [44]. The ReLU is the common activation function, which is utilized to add non-linearity to the system [44].

2. The Pooling Layer: which executes a downsampling operation to decrease the spatial dimension of the convolutional layers. First, the size of the pooling mask and pooling operation type must be determined and after that applied at the pooling layer [45]. The pooling operation is implemented on the pixel values captured by the pooling mask, multiply it by a trainable coefficient, after that added to a trainable bias [42]. The pooling operation is described :

$$x_j^l = f(b_j^l \; pool(x_j^l) + b_j^l \; (6)$$

Where x_j^l is the result of the downsampling operation applied on the jth region in the input, x_j^{l-1} is the jth region of interest captured by the pooling mask in the previous layer, the pool is the specific operation done on the region (max or average), b_j^l is a trainable coefficient, is a trainable bias, and the activation function is f [46].

3. Fully connected layers: which used the extraction attributes in the previous layers to perform the classification job [47]. The result of the last convolutional or pooling layer is fed to the fully connected layers like in an original neural network

EXPERIMENT RESULTS and DISCUSSION

The results obtained in the previous work "Brain Tumor Diagnosis Support System: A Decision Fusion Framework" [2] are listed in Table 1.

Table.1 Confusion Matrix of SVM, ANN, & KNN classifier

ANN				SVM				KNN		
╳	0	1		╳	0	1		╳	0	1
0	27	11		0	28	10		0	30	8
1	18	170		1	20	168		1	18	170

In previous work, the applied classifiers were K-nearest neighbor KNN, support vector machine SVM, and artificial neuron network ANN to detect the brain tumor employing MRI images. The results are specified in the shape of a confusion matrix for the three classifiers. The classification conclusions are arranged for the three classifiers: k- nearest neighbor denoted as KNN, support vector machine shown by SVM. In addition, the result obtained from Convolution Neural Network marked as CNN in this work is illustrated in Table 2.

Table.2 Algorithms evaluation in terms of accuracy and F1 score

Classifier	ANN	SVM	KNN	CNN
Accuracy %	87	86	88	89
F1 Score	0.92	0.91	0.93	0.94

The confusion matrix of every classifier demonstrates in Table 1. Class 0 represents the normal (benign); class 1 represents abnormal (malignant), The KNN algorithm offers the highest accuracy in categorizing classes related to class 0 (benign), where 30 images were correctly classified as benign, and eight images were misclassified as malignant. The artificial neuron network offers the highest performance in the categorization related to class one since 170 images were correctly classified as malignant, while 18 images were misclassified as benign.

The convolution neural network achieved high accuracy in the MRI brain tumor classification as normal or abnormal scans with a precision of 89%. The Softmax fully Connected layer classifier in the CNN building has been applied to assess the suggested method's effectiveness. The classifier's performance is validated by multi measuring factors such as accuracy and F1 score. The F1 score combines the model's precision and recall and is required when you want to investigate a balance between Precision and Recall. Also, it is used to evaluate binary classification systems. The CNN accuracy obtained by softmax and F1 score is 89% and 0.94, respectively. In addition, the Softmax classifier has the most incredible precision in the CNN algorithm. However, The Convolutional neural network algorithm has a high computational cost and required the right volume of data for the training phase. Table 2 illustrates the accuracy and F1 score of each algorithm.

CONCLUSION

This paper has offered a survey of locating the effect of auto attribute abstraction and dissemination, which applied in deep learning; for example, CNN algorithm. The objective of this is to demonstrate the emanating study of existing configuration on attribute extraction techniques during the last few years. As the demand for the application grew, significant analysis and research of the feature extraction platform changed into being particularly active. It would help conduct more experiments to validate the multi-modal architecture's possible advantages against applying a specific convolution neural network. Besides, the CNN algorithm was applied for the brain tumor detection task. In previous work, three classical classifiers, KNN, SVM, and the ANN classifier, were used to classify the same datasets and evaluate the CNN algorithm's performance against these classifiers' results. The CNN is precisely suitable for choosing an auto-feature in medical images such as magnetic resonance imaging. The datasets contain two classes, normal and abnormal. Physicians classified these scans. CNN with softmax achieved the highest scores in terms of F1 score and accuracy.

References

[1]. Minaee, S., Abdolrashidi, A., & Wang, Y. (2015, August). Iris recognition using scattering transform and textural features. In 2015 IEEE signal processing and signal processing education workshop (SP/SPE) (pp. 37-42). IEEE. DOI: https://doi.org/10.1109/DSP-SPE.2015.7369524

[2]. Shantta, K, Basir, O. (2020). Brain Tumor Diagnosis Support System: A Decision Fusion Framework. IRA International Journal of Applied Sciences (ISSN 2455-4499), 15(3), 30-47.DOI: https://doi.org/10.21013/jas.v15.n3.p1

[3]. Wang, H., & Raj, B. (2017). On the origin of deep learning. arXiv preprint arXiv:1702.07800.

[4]. Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.

[5]. Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295-2329. DOI: https://doi.org/10.1109/JPROC.2017.2761740

[6]. Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). Greedy layer-wise training of deep networks. In Advances in neural information processing systems (pp. 153-160).

[7]. Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An Introductory Review of Deep Learning for Prediction Models With Big Data. Front. Artif. Intell, 3(4). DOI: https://doi.org/10.3389/frai.2020.00004

[8]. Deng, L. (2018). Artificial intelligence in the rising wave of deep learning: The historical path and future outlook [perspectives]. IEEE Signal Processing Magazine, 35(1), 180-177. DOI: https://doi.org/10.1109/MSP.2017.2762725

[9]. Strogatz, S. (2018). One giant step for a chess-playing machine. New York Times.

[10]. Farooq, M., & 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. DOI: https://doi.org/10.1007/978-3-319-56148-6_41

[11]. Lu, X., Duan, X., Mao, X., Li, Y., & Zhang, X. (2017). Feature extraction and fusion using deep convolutional neural networks for face detection. Mathematical Problems in Engineering, 2017. DOI: https://doi.org/10.1155/2017/1376726

[12]. Huan, E. Y., Wen, G. H., Zhang, S. J., Li, D. Y., Hu, Y., Chang, T. Y., ... & Huang, B. L. (2017). Deep convolutional neural networks for classifying body constitution based on face image. Computational and Mathematical Methods in Medicine, 2017. DOI: https://doi.org/10.1155/2017/9846707

[13]. Hossain, T., Shishir, F. S., Ashraf, M., Al Nasim, M. A., & Shah, F. M. (2019, May). Brain Tumor Detection Using Convolutional Neural Network. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (pp. 1-6). IEEE. DOI: https://doi.org/10.1109/ICASERT.2019.8934561

[14]. Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: a review. EURASIP journal on wireless communications and networking, 2017(1), 1-12. DOI: https://doi.org/10.1186/s13638-017-0993-1

[15]. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., ...& Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. Medical image analysis, 35, 18-31. DOI: https://doi.org/10.1016/j.media.2016.05.004

[16]. Lorentzon, M. (2017). Feature extraction for image selection using machine learning.

[17]. Li, R., Zhang, W., Suk, H. I., Wang, L., Li, J., Shen, D., & Ji, S. (2014, September). Deep learning based imaging data completion for improved brain disease diagnosis. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 305-312). Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-10443-0_39

[18]. Zhu, P., Isaacs, J., Fu, B., & Ferrari, S. (2017, December). Deep learning feature extraction for target recognition and classification in underwater sonar images. In 2017 IEEE 56th Annual Conference on Decision and Control (CDC) (pp. 2724-2731). IEEE. DOI: https://doi.org/10.1109/CDC.2017.8264055

[19]. Mohsen, H., El-Dahshan, E. A., El-Horbaty, E. M., & Salem, A. M. (2017). Brain tumor type classification based on support vector machine in magnetic resonance images. Annals Of "Dunarea De Jos" University Of Galati, Mathematics, Physics, Theoretical mechanics, Fascicle II, Year IX (XL), (1).

[20]. Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. IEEE transactions on medical imaging, 35(5), 1240-1251. DOI: https://doi.org/10.1109/TMI.2016.2538465

[21]. Siar, H., & Teshnehlab, M. (2019, January). Diagnosing and classification tumors and MS simultaneous of magnetic resonance images using convolution neural network. In 2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS) (pp. 1-4). IEEE. DOI: https://doi.org/10.1109/CFIS.2019.8692148

[22]. Szilagyi, L., Lefkovits, L., & Benyo, B. (2015, August). Automatic brain tumor segmentation in multispectral MRI volumes using a fuzzy c-means cascade algorithm. In 2015 12th international conference on fuzzy systems and knowledge discovery (FSKD) (pp. 285-291). IEEE. DOI: https://doi.org/10.1109/FSKD.2015.7381955

[23]. Xu, Y., Jia, Z., Ai, Y., Zhang, F., Lai, M., Eric, I., & Chang, C. (2015, April). Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation. In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 947-951). IEEE. DOI: https://doi.org/10.1109/ICASSP.2015.7178109

[24]. Pan, Y., Huang, W., Lin, Z., Zhu, W., Zhou, J., Wong, J., & Ding, Z. (2015, August). Brain tumor grading based on neural networks and convolutional neural networks. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 699-702). IEEE. DOI: https://doi.org/10.1109/EMBC.2015.7318458

[25]. Basheera, S., & Ram, M. S. S. (2019, March). Classification of brain tumors using deep features extracted using CNN. In Journal of Physics: Conference Series (Vol. 1172, No. 1, p. 012016). IOP Publishing.DOI: https://doi.org/10.1088/1742-6596/1172/1/012016

[26]. Menegola, A., Fornaciali, M., Pires, R., Avila, S., & Valle, E. (2016). Towards automated melanoma screening: Exploring transfer learning schemes. arXiv preprint arXiv:1609.01228.

[27]. Bhandari, A., Koppen, J., & Agzarian, M. (2020). Convolutional neural networks for brain tumour segmentation. Insights into Imaging, 11(1), 1-9. DOI: https://doi.org/10.1186/s13244-020-00869-4

[28]. Otberdout, N., Kacem, A., Daoudi, M., Ballihi, L., & Berretti, S. (2018). Deep covariance descriptors for facial expression recognition. arXiv preprint arXiv:1805.03869.

[29]. Hashemzehi, R., Mahdavi, S. J. S., Kheirabadi, M., & Kamel, S. R. (2020). Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE. Biocybernetics and Biomedical Engineering. DOI: https://doi.org/10.1016/j.bbe.2020.06.001

[30]. Seetha, J., & Raja, S. S. (2018). Brain tumor classification using convolutional neural networks. Biomedical & Pharmacology Journal, 11(3), 1457. DOI: https://doi.org/10.13005/bpj/1511

[31]. Khan, H. A., Jue, W., Mushtaq, M., & Mushtaq, M. U. (2020). Brain tumor classification in MRI image using convolutional neural network. Mathematical Biosciences and Engineering, 17(5), 6203-6216. DOI: https://doi.org/10.3934/mbe.2020328

[32]. Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. Sensors, 17(4), 818. DOI: https://doi.org/10.3390/s17040818

[33]. Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 54(10), 6232-6251. DOI: https://doi.org/10.1109/TGRS.2016.2584107

[34]. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. Electronics, 8(3), 292. DOI:https://doi.org/10.3390/electronics8030292

[35]. Le Roux, N., & Bengio, Y. (2008). Representational power of restricted Boltzmann machines and deep belief networks. Neural computation, 20(6), 1631-1649. DOI: https://doi.org/10.1162/neco.2008.04-07-510

[36]. O'Connor, P., Neil, D., Liu, S. C., Delbruck, T., & Pfeiffer, M. (2013). Real-time classification and sensor fusion with a spiking deep belief network. Frontiers in neuroscience, 7, 178. DOI: https://doi.org/10.3389/fnins.2013.00178

[37]. Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554. DOI: https://doi.org/10.1162/neco.2006.18.7.1527

[38]. Salakhutdinov, R., & Larochelle, H. (2010, March). Efficient learning of deep Boltzmann machines. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 693-700).

[39]. Ngiam, J., Chen, Z., Koh, P. W., & Ng, A. Y. (2011). Learning deep energy models. In Proceedings of the 28th international conference on machine learning (ICML-11) (pp. 1105-1112).

[40]. Salakhutdinov, R., & Larochelle, H. (2010, March). Efficient learning of deep Boltzmann machines. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 693-700).

[41]. Mohsen, H., El-Dahshan, E. S. A., El-Horbaty, E. S. M., & Salem, A. B. M. (2018). Classification using deep learning neural networks for brain tumors. Future Computing and Informatics Journal, 3(1), 68-71. DOI: https://doi.org/10.1016/j.fcij.2017.12.001

[42]. Al-Waisy, A. S., Qahwaji, R., Ipson, S., Al-Fahdawi, S., & Nagem, T. A. (2018). A multi-biometric iris recognition system based on a deep learning approach. Pattern Analysis and Applications, 21(3), 783-802. DOI: https://doi.org/10.1007/s10044-017-0656-1

[43]. Nagi, J., Ducatelle, F., Di Caro, G. A., Cireşan, D., Meier, U., Giusti, A., ... & Gambardella, L. M. (2011, November). Max-pooling convolutional neural networks for vision-based hand gesture recognition. In 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) (pp. 342-347). IEEE. DOI: https://doi.org/10.1109/ICSIPA.2011.6144164

[44]. Oyedotun, O., & Khashman, A. (2017). Iris nevus diagnosis: convolutional neural network and deep belief network. Turkish Journal of Electrical Engineering & Computer Sciences, 25(2), 1106-1115. DOI: https://doi.org/10.3906/elk-1507-190

[45]. Scherer, D., Müller, A., & Behnke, S. (2010, September). Evaluation of pooling operations in convolutional architectures for object recognition. In International conference on artificial neural networks (pp. 92-101). Springer, Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-642-15825-4_10

[46]. Van Doorn, J. (2014). Analysis of deep convolutional neural network architectures. In 21st Twente Student Conf. on IT.

[47]. Syafeeza, A. R., Khalil-Hani, M., Liew, S. S., & Bakhteri, R. (2014). Convolutional neural network for face recognition with pose and illumination variation. International Journal of Engineering & Technology, 6(1), 0975-4024.