



AK Tech Training and Placements

Transform Dreams into Reality

BREAST CANCER DETECTION

Cancer is one of the menacing and unpredictable disease. If it is not detected in its first stage then it could endanger the person life. Similarly according to Breast Cancer Institute (BCI), Breast Cancer is one of the most dangerous type of diseases that is very effective for women in the world. For detecting breast cancer mostly machine learning techniques are used. In this project we propose an adaptive ensemble voting technique for diagnosed breast cancer using Wisconsin Breast cancer database. The main objective of this work is to explain how CNN and logistic regression , Support vector machine(SVM), K-nearest neighbor(KNN) algorithm provides better solution when it works with ensemble machine learning algorithms for predicting the breast cancer. When compared to related work from the literature. It is shown that the CNN approach achieves 94.12% accuracy from another machine learning algorithm.

EXISTING SYSTEM

Generally there are two type of tumors. One is benign and other is malignant tumor in which benign Tumor is non-cancerous and malignant is a cancer Tumor. There are various methods and algorithms that are available for detecting the breast cancer such Support vector machine (SVM), Naïve Bayes, KNN and ANN etc. ANN is a deep learning technique which is generally used to predict the continuous as well as non-continuous data. Before the Artificial neural network (ANN) is applied some preprocessing on the data is required in order to get a good accuracy. On the dataset firstly feature selection process using recursive feature elimination is carried out and then top 16 features are selected out and then the ANN is applied on it.

DISADVANTAGES

- **Hardware dependence:** Artificial neural networks require processors with parallel processing power, in accordance with their structure. For this reason, the realization of the equipment is dependent.
- **Unexplained behavior of the network:** This is the most important problem of ANN. When ANN produces a probing solution, it does not give a clue as to why and how. This reduces trust in the network.

- **Determination of proper network structure:** There is no specific rule for determining the structure of artificial neural networks. Appropriate network structure is achieved through experience and trial and error.
- **Difficulty of showing the problem to the network:** ANNs can work with numerical information. Problems have to be translated into numerical values before being introduced to ANN. The display mechanism to be determined here will directly influence the performance of the network. This depends on the user's ability.
- **The duration of the network is unknown:** The network is reduced to a certain value of the error on the sample means that the training has been completed. This value does not give us optimum results.

PROPOSED SYSTEM

The proposed system is working on various algorithms and based on that it propose that which algorithm is best to predict the breast cancer. The proposed system is using support vector machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression and Convolutional Neural Network (CNN). CNN is a deep learning technique which process on the images and finds the best features of the images and can be used to predict a categorical data. This a powerful technique which can be used in various domains. Generally neural networks consist of individual units called neurons. Neurons are located in a series of groups—layers. Neurons in each layer are connected to neurons of the next layer. Data comes from the input layer to the output layer along these compounds. Each individual node performs a simple mathematical calculation. Then it transmits its data to all the nodes it is connected to. Convolutional neural networks (CNN) is a special architecture of artificial neural networks. CNN uses some features of the visual cortex. One of the most popular uses of this architecture is image classification. For example Facebook uses CNN for automatic tagging algorithms. Computer sees the image as an array of pixels. For example, if image size is 300 x 300. In this case, the size of the array will be 300x300x3. Where 300 is width, next 300 is height and 3 is RGB channel values. The computer is assigned a value from 0 to 255 to each of these numbers. This value describes the intensity of the pixel at each point. To solve this problem the computer looks for the characteristics of the base level. In human understanding such characteristics are for example the trunk or large ears. For the computer, these characteristics are boundaries or curvatures. And then through the groups of convolutional layers the computer constructs more abstract concepts. The Convolution layer is always the first. The image (matrix with pixel values) is entered into it. Imagine that the reading of the input matrix begins at the top left of image. Next the software selects a smaller matrix there, which is called a filter (or neuron, or core). Then the filter produces convolution, i.e. moves along the input image. The filter's task is to multiply its values by the original pixel values. All these multiplications are summed up. One number is obtained in the end.

Since the filter has read the image only in the upper left corner, it moves further and further right by 1 unit performing a similar operation. After passing the filter across all positions, a matrix is obtained, but smaller than an input matrix. The network will consist of several convolutional networks mixed with nonlinear and pooling layers. When the image passes through one convolution layer, the output of the first layer becomes the input for the second layer. And this happens with every further convolutional layer. The nonlinear layer is added after each convolution operation. It has an activation function, which brings nonlinear property. Without this property a network would not be sufficiently intense and will not be able to model the response variable (as a class label). The pooling layer follows the nonlinear layer. It works with width and height of the image and performs a down sampling operation on them. As a result the image volume is reduced. This means that if some features (as for example boundaries) have already been identified in the previous convolution operation, than a detailed image is no longer needed for further processing, and it is compressed to less detailed pictures. After completion of series of convolutional, nonlinear and pooling layers, it is necessary to attach a fully connected layer. This layer takes the output information from convolutional networks. Attaching a fully connected layer to the end of the network results in an N dimensional vector, where N is the amount of classes from which the model selects the desired class.

ADVANTAGES

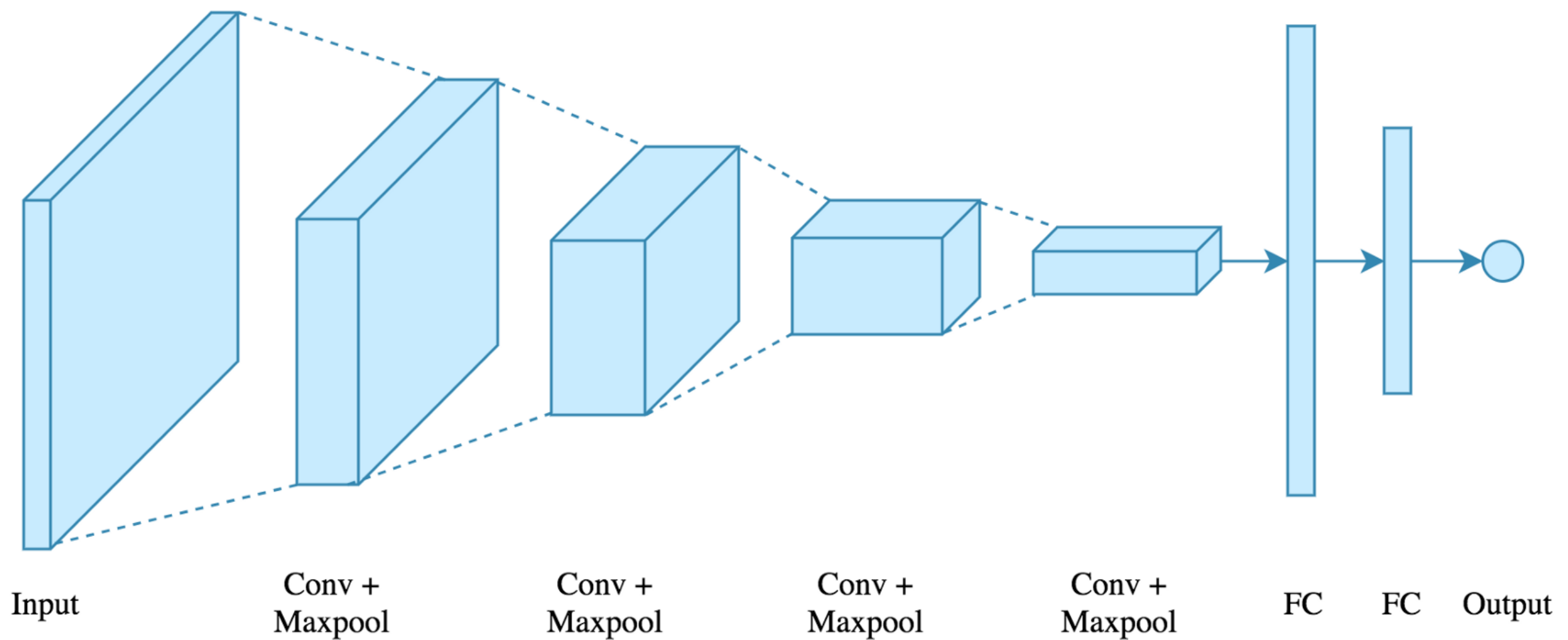
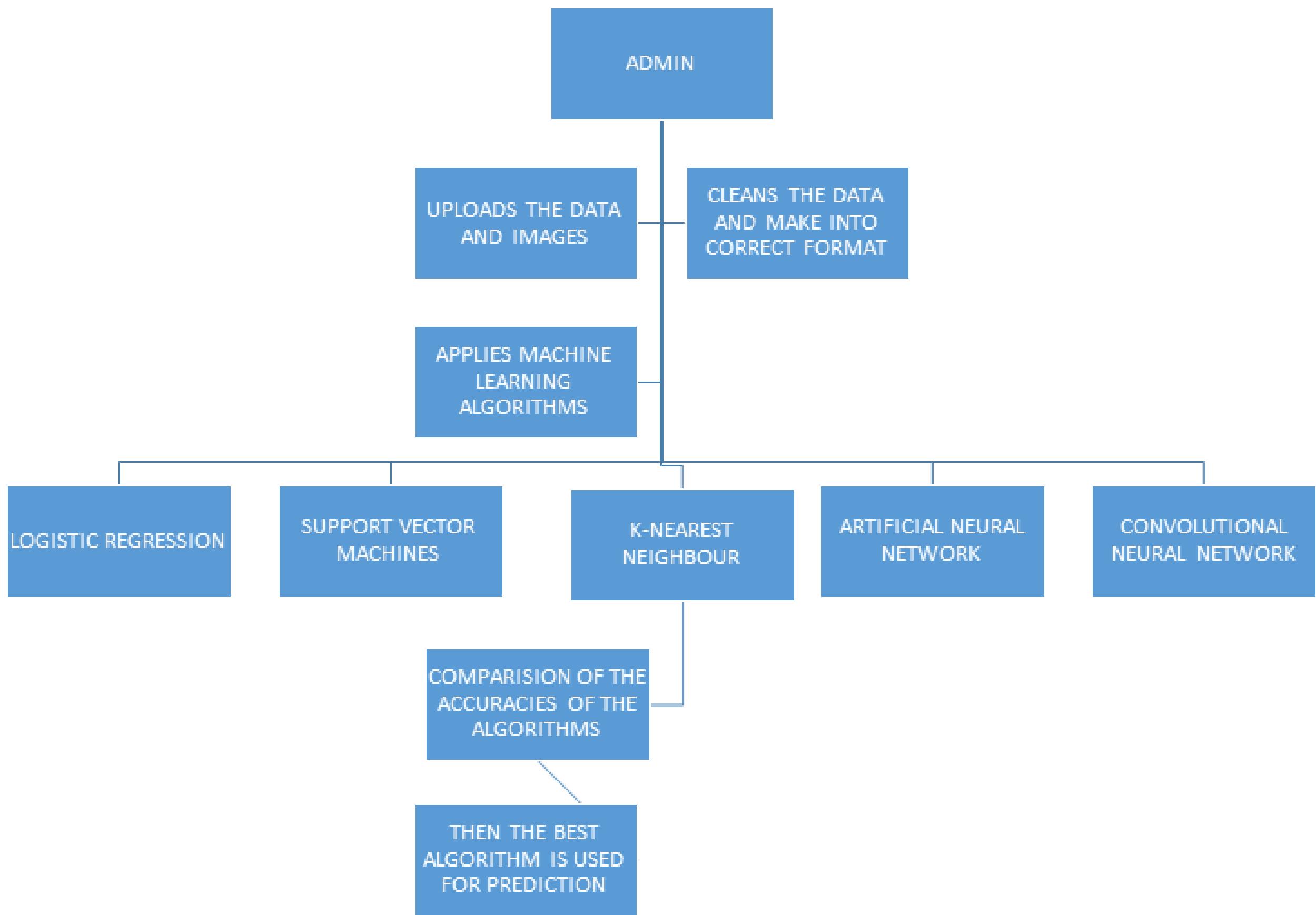
- The usage of CNNs are motivated by the fact that they can capture / are able to learn relevant features from an image /video at different levels similar to a human brain. This is feature learning.

- In terms of performance, CNNs outperform NNs on conventional image recognition tasks and many other tasks.

- For a completely new task / problem CNNs are very good feature extractors. This means that we can extract useful attributes from an already trained CNN with its trained weights by feeding your data on each level and tune the CNN a bit for the specific task.

E.g. : Add a classifier after the last layer with labels specific to the task. This is also called pre-training and CNNs are very efficient in such tasks compared to NNs.

ARCHITECTURE



MODULES

USER REGISTRATION

The patient comes and does the registration at the reception and asks for the appointment. The receptionist fill in the details of the patients in their databases and fixes an appointment with the doctor. After the verification from the doctor the confirmation is given to the patients regarding the appointment of the doctor.

ADMIN

The admin is responsible for the storage and retrieval of the patient's data in the database. Admin Is also responsible for the collection of the charges that has to be pays by the patients for their treatment and is responsible to give the salary to every employee working for that hospital.

PICTORIAL REPRESNTATION

The analyses of proposed systems are calculated based on the User session details. This can be measured with the help of graphical notations such as pie chart, bar chart and line chart. The data can be given in a dynamical data.

ALGORITHMS

LOGISTIC REGRESSION

A popular statistical technique to predict binomial outcomes ($y = 0$ or 1) is Logistic Regression. Logistic regression predicts categorical outcomes (binomial / multinomial values of y). The predictions of Logistic Regression (henceforth, LogR in this article) are in the form of probabilities of an event occurring, i.e. the probability of $y=1$, given certain values of input variables x . Thus, the results of LogR range between 0-1.

LogR models the data points using the standard logistic function, which is an S- shaped curve also called as sigmoid curve and is given by the equation:

$$\frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$$

SUPPORT VECTOR MACHINE (SVM)

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot). The SVM algorithm is implemented in practice using a kernel. The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra, which is out of the scope of this introduction to SVM. A powerful insight is that the linear SVM can be rephrased using the inner product of any two given observations, rather than the observations themselves. The inner product between two vectors is the sum of the multiplication of each pair of input values. For example, the inner product of the vectors [2, 3] and [5, 6] is $2*5 + 3*6$ or 28. The equation for making a prediction for a new input using the dot product between the input (x) and each support vector (xi) is calculated as follows:

$$f(x) = B_0 + \sum(a_i * (x, x_i))$$

K-NEAREST NEIGHBOUR

In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression.[1] In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression: In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors. K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms.

ARTIFICIAL NEURAL NETWORK

Artificial neural networks are one of the main tools used in machine learning. As the “neural” part of their name suggests, they are brain-inspired systems which are intended to replicate the way that humans learn. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something that the output layer can use. They are excellent tools for finding the patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize.

CONVOLUTIONAL NEURAL NETWORK

A CNN consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers. A convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery.

SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

- **System** : Pentium IV 2.4 GHz.
- **Hard Disk** : 40 GB.
- **Floppy Drive** :1.44 Mb.
- **Monitor** : 14' Colour Monitor.
- **Mouse** : Optical Mouse.
- **Ram** :512 Mb.

SOFTWARE REQUIREMENTS:

- **Operating system** : Windows 7 Ultimate.
- **Coding Language** : Python.
- **Front-End** : Python.
- **Designing** : Html,css,javascript.
- **Data Base** : MySQL.