IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Multi-Disease Detection and Predictions Based on Machine Learning

Ashwat Gore[1], Intekhab Alam[2], Akshata Bhute[3], Tejasvini Shejawal[4] , Mr. S.B. Kothari[5]

[1234] ***Student, Department of Information Technology,***
SKN Sinhgad Institute of Technology and Science, Lonavala, Pune , Maharashtra, India.
[5]***Assistant Professor, Department of Information Technology,***
SKN Sinhgad Institute of Technology and Science, Lonavala, Pune , Maharashtra, India.

ABSTRACT

Chronic diseases such as skin disease, cancer, diabetes, stroke, and arthritis are the leading causes of disability and death in India and throughout the world. As compared to other diseases these diseases having high rate of deaths, so there is need of optimistic solution over chronic diseases. Medical data growth is in healthcare communities, accurate analysis of medical data benefit early disease detection, patient care and community services. However, the analysis of patients depends on accuracy of diagnosis and then treatment as well. The wrongly diagnosed patients die in chronic type diseases. So in the high risk of mortality there is need of accurate diagnosis aid for chronic diseases. So we are proposing diagnostic system based on machine-learning for providing promising solution with high accuracy. The proposed system consists of many diseases such as lung cancer, brain tumour, skin disease detections and stages predictions. High rate of deaths due to chronic diseases such as skin disease, lung cancer, brain tumour need to develop proper diagnosis system which helps doctors. The wrong diagnosis leads to human deaths so we need to progress on accurate diagnosis of multiple diseases. Many works have been already carried out for different diseases but there are not any promising solution found that gives accurate diagnosis for all in one. The proposed system consists of many diseases such as lung cancer, brain tumour, skin disease detections and stages predictions. We are trying to develop a system for multi-disease detection and stage predictions that gives early detection and saves lots of lives by reducing death rate by chronic diseases.

Keywords: *Convolutional Neural Network, Computed Tomography, Support Vector Machine.*

I. INTRODUCTION

Now a days skin diseases are growing rapidly by busy and stress full life. All type of age groups are under skin diseases so need of early detection of skin disease by using features or reports. Due to large amount of smoking and air pollution around the world, lung cancer has become one of the most common and deadly disease in recent decades. It is one the most dangerous disease among men and women and early identification and treatment is the best available option for the infected people. Main objective behind to develop a system helps the doctors to cross verify their diagnosed results which
gives promising solution over existing death rates.
By utilizing our proposed work try to invent unique platform and most promising solution for early diagnosis of multiple diseases. The existing work analysis accuracy declines when the quality of medical data is incomplete. Moreover, different places exhibit unique characteristics of certain regional diseases, which may weaken the predictions of disease wrong. So we are giving more accurate solution by using machine learning and Convolutional neural network to detect diseases and make predictions. The proposed system consists of many diseases such as lung cancer, brain tumour, skin disease detections and stages predictions.

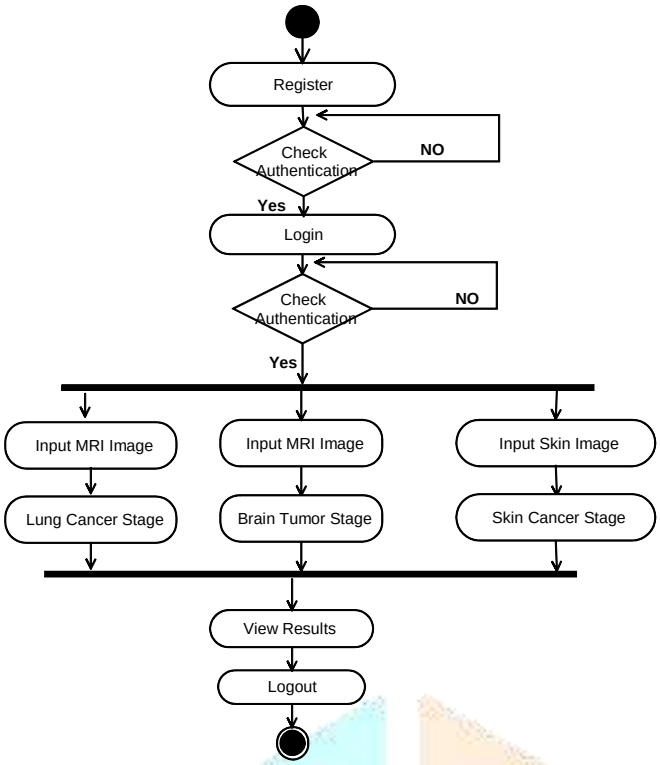


Fig.1 Flow of proposed work

Above figure shows internal working of multiple diseases detection and stages prediction system. In which we have used NSCLC Radio-genomics lung cancer CT image dataset, skin disease dataset and brain tumour MRI images. After getting this DICOM medical formatted images those converted by DICOM converter to PNG format. Image acquisition contains image reading by using opencv-python for process it. After getting image of lung cancer it transfer into noise reduction for removing noise from it. Then image processing techniques applied on it like binary image conversion and gray image conversion followed by segmentation. The image features get collected and drafted into machine model in training phase of machine learning for future prediction of multiple diseases and stages evaluations. We have worked on supervised dataset due to which it become easier for detecting and give predictions.

II. OBJECTIVE & GOALS

- To reduce death rate by chronic diseases in the world.
- To give unique solution for multiple diseases.
- To provide higher accuracy over previous research.
- To give most promising tool that can acceptable by all the doctors.
- Detection of multiple types of Diseases.
- Provide diagnosis for:

1. Lung Cancer
2. Brain Tumour
3. Skin Disease

- By improving proper diagnosis increase rate of saving life's.

III. LITERATURE SURVEY

Animesh Hazra et al. [1] proposed Skin diseases when irritated winding path out of hand. Skin diseases are convoluted and remove loads of lives each year .When the early side effects of skin diseases are overlooked, the patient may wind up with uncommon outcomes in a short length of time. Stationary way of life and over the top worry in this day and age has exacerbated the circumstance. On the off chance that the diseases are distinguished early, at that point it very well may be monitored. Be that as it may, it is constantly prudent to practice every day and dispose of undesirable propensities at the soonest. Tobacco utilization and undesirable weight control plans increment the odds of stroke and skin illnesses. Eating in any event 5 helpings of products of the soil multi day is a decent practice. For coronary illness patients, it is prudent to confine the admission of salt to one teaspoon every day.

Anuradha S. Deshpande et.al[2] states among every other sort of malignant growth, Lung Cancer is one of the most widely recognized reasons for death all through the world. It is important to get legitimate treatment on schedule, to diminish demise pace of individuals on the planet.. In the Lung malignant growth discovery framework, we have recognized different phases of growth by utilizing Support vector machine classifier (SVM). The watershed division is used in this framework which is the best division procedure. Utilizing MATLAB programming, we have User Interface UI (GUI) and it is utilized to play out every one of the procedures of the framework. We will probably get increasingly exact and exact consequences of the various phases of disease by utilizing different techniques. We have played out the combination of CT and MRI examining. This method improves the nature of the information. In this manner, we get the suitable stage.

Abbas Khosrav et.al [3] proposed Characterization of lung disease utilizing a low populace, high dimensional dataset is trying because of lacking tests to become familiar with an exact mapping among highlights and class marks. Current writing more often than not handles this errand through high quality component creation and determination. Lately profound learning is observed to have the option to recognize the fundamental structure of information using auto encoders and different procedures. In this work, a profound auto encoder grouping system is proposed which initially adapts profound highlights and after that prepares a fake neural system with these educated highlights. Exploratory outcomes demonstrate the profound educated classifier outflanks every single other classifier at the point when prepared with all properties and same preparing tests. It is likewise exhibited that the presentation improvement is factually huge.

Ali M. Hasan et.al[4] introducing Mind tumor division in magnetic resonance imaging (MRI) is viewed as a complex methodology in light of the inconstancy of tumor shapes and the intricacy of deciding the tumor area, size, and surface. Brain tumor division is a tedious assignment profoundly inclined to human mistake. Henceforth, this examination proposes a mechanized strategy that can recognize tumor cuts and portion the tumor overall picture cuts in volumetric

MRI mind filters. Initial, a lot of calculations in the pre-preparing stage are utilized to clean and institutionalize the gathered information. An altered dim level co-event lattice and Analysis of Variance (ANOVA) are used for highlight extraction and indulge in choice, individually. A multi-layer perceptron neural system is received and used as a classifier, and a bouncing 3D-box-based hereditary calculation is used to distinguish the zone of obsessive tissues in the MRI cuts. At long last, the 3D dynamic shape without edge is connected to portion the cerebrum tumors in volumetric MRI checks.

Shadab Adam Pattekari and Asma Parveen [5] states Choice Support in Skin Disease Prediction Framework is created utilizing Naive Bayesian Characterization strategy. The framework differentiates concealed learning from an authentic coronary illness database. This is the best model to predict patients with coronary illness. HDPS can be further improved and extended. For, instance it can fuse other therapeutic qualities other than the above rundown. It can likewise fuse other information mining systems. Ceaseless information can be utilized rather than simple straight out information.

S.Florence et.al. [6] introducing The medicinal services condition is an ever increasing number of information advanced, yet the measure of learning getting from those information is less, on the grounds that absence of information investigation apparatuses. We have to get the hidden connections from the information. In the human services framework to foresee the skin assault consummately, there are a few methods which are now being used. There is some absence of accuracy in the accessible methods like Naïve Bayes. Here, this paper proposes the framework which employments neural system and Decision tree (ID3) to foresee the skin assaults. Here the dataset with 6 credits is utilized to analyze the skin assaults. The dataset utilized is part of skin assault dataset given by UCI AI store. The consequences of the expectation give more exact yield than different procedures.

IV. METHODOLOGY USED IN EXISTING SYSTEM

Reliable identification and classification of chronic diseases requires pathological test, namely, needle biopsy specimen and analysis by experienced pathologists as it involves human judgment of several factors and a combination of experiences, a decision support system is desirable in this case. After diagnosis of manual judgments the rate of saving life of patients is not much good. The techniques used in existing work are unable to give Nobel solution over chronic diseases and accuracy is very poor. Algorithm using structured and unstructured data from hospital. To our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Reducing death rate by wrong diagnosis using giving accurate diagnosis. Existing system does not provide diagnosis system which helps the doctors.

V. METHODOLOGY USED IN PROPOSED SYSTEM

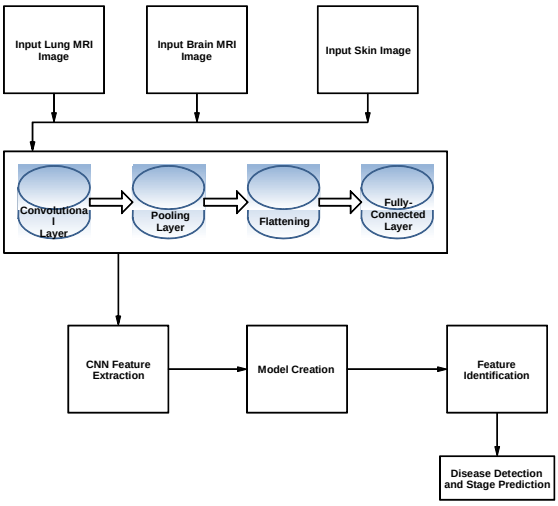


Fig.2 System Architecture

A] Modules:

1) Lung Cancer Detection
2) Skin Disease Identification
3) Brain Tumour Detection

B] Features:

1.Lung cancer type detections according to CT-Scan Images.
2.Skin disease identification according to different features
3.Brain Tumour detections according to MRI Images.
4.Display Precaution list to avoid the various disease.

In a proposed system, we are proposing experiment on chronic diseases like lung cancer, brain tumor and skin disease with limited set of supervised data.
We proposed a new Convolutional neural network based multimodal disease risk prediction model for limited chronic diseases with higher accuracy. We are going to solve accuracy issue in diagnosis of lung cancer with accurate stage predictions. We work on brain tumors detections by machine evaluations depends on tumor sizes in mm. Skin disease detection depends on diagnosed dataset such as features of skin images. In proposed system consists of mainly 3 modules Lung Cancer Detection, Skin Disease Identification, Brain Tumor Detection. Admin and users are two modules included in our system. Admin first gather the information about multiple diseases in the form of text as well as in the form of images. After gathering of information like preprocessing on the data, training of the data, model generation according to the features of the data. User insert the MRI image for checking brain tumor or lung cancer and

enter text features for detection of skin disease. Using our proposed system, we predict the disease according to different type of stages. System also provide precaution list to the user for prevent form disease. We are trying to develop a system for multi disease detection and stages predictions gives early detection and saves lots of lives by reducing the death rate by chronic diseases. We propose a novel Lung detection and Stage prediction mechanism is proposed which first learns deep features and then trains an artificial neural network with these learned features. Experimental results depicts the deep learned classifier outperforms all other classifiers when trained with all the attributes and same training samples. It is demonstrated that the performance improvement is statistically impressive. Classification of lung cancer using a low population, high dimensional data-set is difficult due to insufficient samples to learn an accurate mapping among features and class labels. Current literature usually handle the task through handcrafted feature creation and selection. Deep learning is found to be able to identify the underlying structure of data through the use of CNN and other techniques.

Advantages

1. This shows that application of machine learning has the potential to significantly detect and classify with almost high accuracy for the low population..
2. High dimensional lung cancer data-set without requiring any hand-crafted, case specific features.
3. High processing speed enhanced CNN classifier model.

A. **Image Processing**

An image is made up of RGB colours. Pre-processing unit consists of noise removal, grey scale conversion, binary conversion of images followed by feature extraction. In future extraction there are five steps followed in which fingertips searches by eccentricity. Next elongations of images are measured by accounting pixel segmentation as well as rotation of images.

B. **Image Filtering**

Filtering is a technique to modify or enhance the image, i.e. to highlight certain features or remove other features. It includes smoothing, sharpening. and edge enhancement. Image filtering algorithms produce an output pixel by observing the neighbourhood of the input pixel in an image. Image filtering algorithms are used to remove different types of noise from the image.

C. **Feature Extraction**

In feature extraction, algorithmic study is used to find feature vectors of systematic results combined with K curvature and convex hull algorithms together. In present work “K convex hull” algorithm is used to detect fingertip with greater accuracy. In our system, CNN is used for future recognition in which we having the input unit of training data set of images.

D. **Segmentation**

Image segmentation is the way toward apportioning an advanced picture into various portions (sets of pixels). All pixels in an area share a typical property. Least complex property that pixel can share power. The objective is to disentangle and change the portrayal of the picture into something that is increasingly important and less demanding to break down.

E. **Edge Detection**

Edge shows the boundaries between regions in an image which helps in object detection. There are many edge detection operators and algorithms available. Edge Detection Operators and Algorithms used in our research like Convex hull method.

F. **Feature Recognition**

Brain-inspired systems used to replicate how humans learn. Consist of input, hidden and output layers that transform the input into something that the output layer can use. Significantly good for finding patterns which is complex to human for extract and teach the machine to recognize. CNN collects their knowledge by detecting the patterns and relationships in data and learns (or is trained) through experience, not from the programming.

VI. MATHEMATICAL MODEL

$M = \{Q, \sum, q_0, F, \delta\}$.

Q : Total number of states
$\sum$: Set of input symbols
q_0 : Initial state
F : Set of final states
δ : Transition function

Process

Let us consider M as a system for multi disease detection and prediction system.

Mapping Diagram

1. $Q : (q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7)$

q_0: Register/ Login
q_1: Lung cancer detection
q_2: Brain tumor detection
q_3: Skin disease detection
q_4: Stage Prediction
q_5: Add precautions
q_6: Result of image feature extraction
q_7: Result of disease detection
q_f: Final result

2. $\sum : (a, b)$

a: MRI image input
b: Text format input

3. q_0 : Register/ Login

4. F : (q_f)

F1, F2, F3 are functions to execute the result.
F1: Image processing applied on lung cancer and brain tumor images.
F2: Feature extraction from images

- Binarization
- Grey Scale
- Edge detection

5. δ : q_f

- **Transition Diagram**

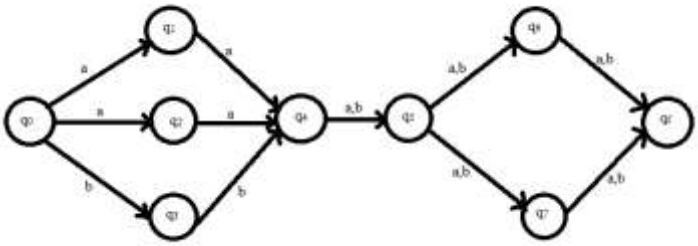


Fig.3 Transition diagram

- **Transition Table**

Table 1 Transition Table

State	Input	
	a	b
q_0	q_1, q_2	q_3
q_1	q_4	-
q_2	q_4	-
q_3	-	q_4
q_4	q_5	q_5
q_5	q_6, q_7	q_6, q_7
q_6	q_f	q_f
q_7	q_f	q_f
q_f^*		

VII. RESULTS AND DISCUSSION

In multi disease detection and stage prediction system we have been implemented highly trained model that can accurately recognize multiple diseases. In this system we used Gaussian blur for gray scale conversion, Otsu's method for binary conversion of images after that convex hull for edge detection has been used.

G. Grey scale conversion

In gray scale conversion colour image is converted into gray by using Gaussian blur. Colour image containing noise and unwanted background which is removed or blurred by using this method.

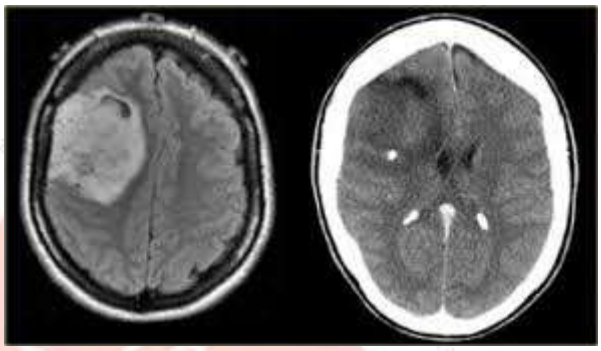


Fig.4 Gray Image

H. Binary conversion

Gray scale image is given to input for Otsu's method for binary conversion. In Binary form of images converted in 0 and 1 form means black and white.

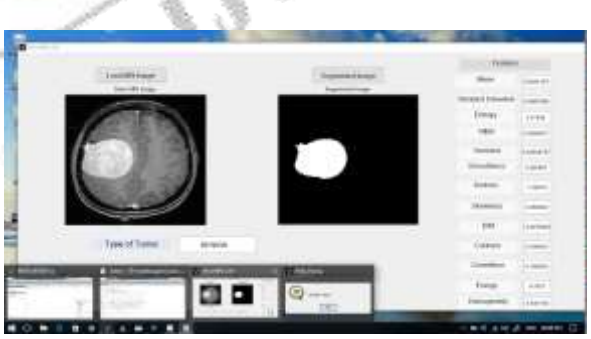


Fig.5 Binary Image

I. Edge Detection

In Edge detection binary image get dimensions by counters using the convex hull algorithm. In which eccentricity finding drawing edges around white portion of binary image.

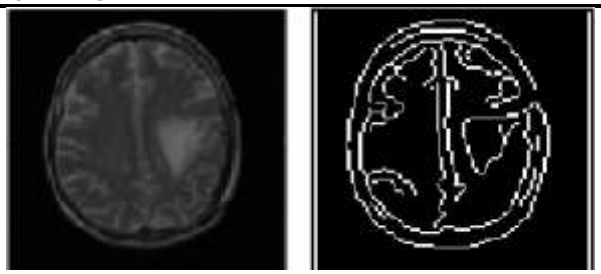


Fig.6 Edge Detection Image

J. **Training Model**

In our system we are using tensor flow for extracting feature's of training dataset. In which 12000 image samples are trained by using training model. Finally plot files generated as an output of our trained model.

K. **Testing Model**

In final phase of data testing in which brain tumour MRI, Skin dataset and lung cancer CT images were matched by our training model with higher percent of accuracy.

After matching all type of disease images respective results of stage and cancer cell detection display on console and stored in text file as well.

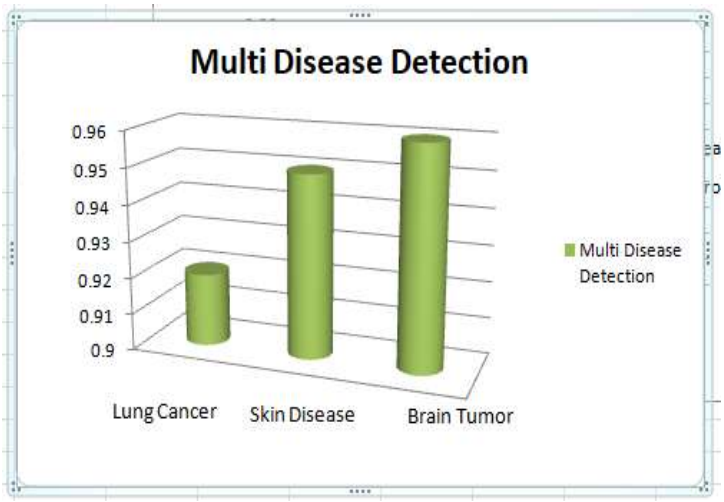
TABLE II

In our experimental setup, in table 1 describe our system modules and respective generated output.

Table 2 Lung cancer stage evaluation criteria

Disease Name	Accuracy	Time Taken(sec)
1 Lung Cancer	87	609
2 Brain Tumor	85	740
3 Skin Disease	89	910

L. **Comparative Study Graph**



Graph 1: Average Accuracy Rate

M. **Accuracy Rate of System**

All lung cancer image samples were trained by our trained model using deep neural network. The same procedure followed by our skin and brain tumor model.

VIII. CONCLUSION

Thus we implemented multi disease detection system over machine learning and CNN techniques which solves existing accuracy problem as well as reduce death rates by chronic type diseases like lung cancer detection, brain tumor detection and skin disease. After detection of disease inform to users that how to prevent from a disease.

For future work, we will implement this technique on some more chronic diseases with rich dataset. Increasing the number of diseases and dataset used for the process, we will improve the accuracy.

IX. REFERENCES

[1] Animesh Hazra, Subrata Kumar Mandal,"Skin Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review". Advances in Computational Sciences and Technology Volume 10, Number 7, 2017

[2] Anuradha S. Deshpande, Dhanesh D. Lokhande, Rahul P. Mundhe, Juilee M.Ghatol, "Lung cancer detection with fusion of CT and MRI image using image processing." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 3, March 2015

[3] Abbas Khosravi ,Syed Moshfeq Salaken, , Amin Khatami, Saeid Nahavandi, Mohammad Anwar Hose "Lung Cancer Classification Using Deep Learned Features on Low Population Dataset" IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE) 2017

[4] Hasan, A., Meziane, F., Aspin, R., & Jalab, H. (2016). "Segmentation of Brain Tumors in MRI Images Using

Three-Dimensional Active Contour without Edge. Symmetry," 8(11), 132. doi:10.3390/sym8110132

[5] Shadab Adam Pattekari and Asma Parveen "PREDICTION SYSTEM FOR SKIN DISEASE USING NAIVE BAYES." International Journal of Advanced Computer and Mathematical Sciences 2015

[6] S.Florence, N.G.Bhuvaneswari Amma , G.Annapoorani , K.Malathi (2015). "Predicting the Risk of Skin Attacks using Neural Network and Decision Tree." International Journal of Innovative Research in Computer and Communication Engineering 2014