Classification and Development of Tool for Heart Diseases (MRI Images) Using Machine Learning

Lalaantika Sharma M.Tech (*pursuing*)
Faculty of Engineering and Technology
Shoolini University
Solan, India
sharmal0707@gmail.com

Gaurav Gupta* (*Assistant professor*)
Faculty of Engineering and Technology
Shoolini University
Solan, India
solan.gaurav@gmail.com

Varun Jaiswal (*Assistant professor*)
Faculty of Engineering and Technology
Shoolini University
Solan, India
computationalvarun@gmail.com

***Abstract*— Heart diseases are one of the major killers worldwide. Early detection of heart disease such as Global Hypokinesia can reduce this global burden. Computational method has potential to predict disease in early stages automatically and especially helpful in resources limited countries. Computational method to predict global hypokinesia based on confirms cases of global hypokinesia through MRI was developed. Almost all feature extraction method was used on MRI images and model was generated on merged and different images separately. High accuracy of model independent test set justified our approaches and reliability of model. The newly developed was implemented in python and available for open use.**

Keywords—cp-charm; cellprofiler; Global hypokenesia; feature extraction; MRI; prediction.

I. INTRODUCTION

Nearly half of all African-American adults have some form of cardiovascular disease, 48 percent of women and 46 percent of men. Heart disease is the No. 1 cause of death in the worldwide and the leading cause of death in the United States, killing over 375,000 Americans a year [5]. So automate detection for such a disease is important. Global Hypokinesia is a disease in which heart is weak which serves blockage of coronary arteries. The various portions of the heart such as ventricles, walls, membranes and arteries etc. are deteriorated and work abnormally in Global hypokinesia. Global hypokinesia can be a reason of heart failure due to low output of heart. Symptoms of global hypokinesia are water and fluid retention, irregular or fast heartbeat, pulmonary congestion, fatigue, dizziness and weakness [2]. The test used to detect Global Hypokinesia is MRI test. The image is described as homogeneous areas according to one or several a priori attributes [7].

MRI is popular and reliable method for testing Global hypokinesia so computational method developed by using MRI images Magnetic resonance imaging (MRI) is noninvasive test that used to recognize medical conditions. MRI procedures a commanding magnetic field, radio frequency pulses and a computer system to yield thorough images of organs, soft tissues, bone and almost all other internal body parts. MRI does not work with harmful ionizing radiation which means x-rays. MRI makes physicians to diagnose various parts of body and define occurrence of definite diseases. The images output of machine can be used to inspected on a printed form or in a CD and transmitted electronically. To produce whole images of the structures inside the heart in cardiac MRI uses a powerful magnetic field, radio waves and a computer system. MRI is used to diagnose cardiac disease and to estimate the heart's structure and role in patients through congenital heart disease.

The threat of having heart disease is worldwide so there is need of a tool which can predict disease on the bases of MRI images. So the tool is developed to predict global hypokinesia on bases of MRI images. The data is collected from IGMC Shimla. The method used in tool was segmentation free. The output can be taken directly by software and will predict disease automatically. Pre-processing of MRI pictures is a step in image analysis which achieve image enhancement and noise reduction methods which are taken to increase image quality, then some morphological operations are applied to detect the Global hypokinesia in the patients through images [4] [1]. The MRI image is shown in Fig 1

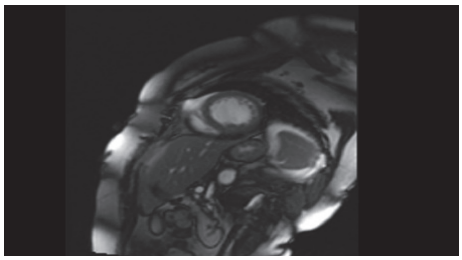

Fig 1: MRI test image

The morphological operations are basically applied on some assumptions about the size and shape of the heart and in the end the heart is mapped onto the actual gray scale image with 255 intensities to create noticeable heart in the image [3].

The purpose of this work is to adapt a new procedure for image processing which automate the process (segmentation free). This software gives good accuracies without any expert physician can predict disease mainly known to be a very powerful feature extraction tool.

The reason that merely a slight number of persons reported being identified as a result of screening indicates that are missing opportunities to prevent heart disease. There are different imaging tests present for heart. Heart Diseases are recognized by radiologist diagnosis tests and forward results to doctors. The research work done here make it easy and simple to predict results and help radiologist. Image processing has been an active field of research topic due to its diverse applicable environment. In this work, there is special focus on making process automate. This tool is used to predict Global Hypokinesia cardiac disease.

Machine learning is used by the software we used for the prediction there is option of two approaches for prediction one is PCA-LDA and another is WND which is for classifier type. The approach used in research is PCA-LDA algorithm of machine learning. In PCA-LDA both techniques are projecting the data onto a smaller feature subspace: with PCA, there would find the directions (components) that maximizes variance in dataset (without considering class labels), and with LDA there would be the components that maximize the between-class separation.

The set of data used in research is taken from IGMC Shimla. Total data set used in current study was 25 images of each patient 30 normal and 30 with having Global hypokinesia patients. Each patient having 900-1200 images MRI. With recommendation of doctors there is use of only 25 image set of each patient. Then further processing is done on these sets. As images cannot be directly used in machine learning, so need is to extract the features from the images. The features are further used to train and test data. These images having accuracies individually and with merge both. The merged images have less loss of information which goes to better accuracies.

II. Material and Method Used

Computational method to predict Global Hypokinesia was developed through following steps and each step uses different techniques. Implementation was done by using softwares and all software's are open source. The research further give a survey of technology used with regard to establishing a system which predict global hypokinesia heart disease on the bases of MRI datasets. In first steps it describes how the MRI images were collected from hospitals and how images were processed. Next it shows how features are extracted from the datasets. Finally, it describes the pipeline which does testing and training part. Main softwares are CellProfiler [9] and Cp-charm [5]. Methods used in present research are as follows: In present research work various steps were followed to reach the conclusion. For each step, a research plan was created, that customized to our needs and these steps involved MRI data collection, processing of collected data, extract features from images, training of datasets and finally testing development. The Fig 2 shows steps used for computation.

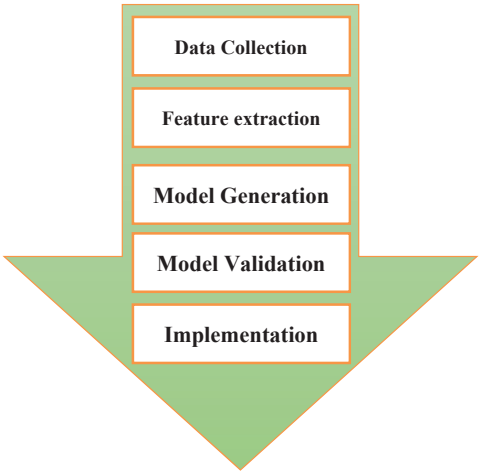

Fig 2 Work Flow

A. Data collection

The data related to cardiac MRI images was collected from IGMC Shimla. There were 30 patients of each category having set of 900-1200 images depending upon cases. The images are in Dicom format and there are various folders of images for categorization of them. The data is available for testing on https://github.com/lalaantika/HyperkinesiaML/tree/master/dataset.

In this method images used for feature extraction were taken directly without doing any change to images. There are 25 image set of every patient. The group of all patients was made on bases of their categorization of MRI. The 30 patients having disease with First image compare with the first image of patient who is healthy. Then further processing on image is done.

In this method images used for feature extraction were firstly processed by the software. Amide software was used to merge Dicom images. This software merge 25 images to one single image so that there was as less as possible loss of information when features are extracted from images. The images were taken from 25 image set of of every patient. The group of all patients was made on the bases of their categorization of MRI. The 30 patients having disease with their merged image was compared with the single merged image of patient who is healthy. Then further processing on image is done.

B. Feature extraction

The image cannot be used directly in machine leaning so feature extraction method was used. Feature extraction in image processing is a method of processing images by using mathematical operations using any method of signal processing for which the input can be an image, a series of images, or a video, like a photograph or video frame; the outcomes of image processing may be whichever, an image or a set of characteristics associated to the image [6]. Now feature extraction in this research is accomplished with CellProfiler. CellProfiler extract features from image set with the help of ImageJ plugins and CP-Charm pipeline. There are 947 features from images. CellProfiler create three files as an output after taking images and processed them. Two files from these three files are CHARM-like_training_data and CHARM-

like_training_labels. These are csv files having data files having feature values of 947 features of all images and labels file contains Meta_keys and categories. Third file created by CellProfiler is Matlab file to save the project.

CP-Charm pipeline used for testing and training part of image. The data set and label file which were outcomes of CellProfiler are taken as input by CP-Charm. Then CP-Charm GUItraintest file was run which input file of features and labels in csv form. There are various preferences provided by CP-Charm for classifier type, Validation type, Cross-validation fold type and number of training/testing rounds etc.

C. Model generation

CP-Charm pipeline used for training and testing part of image. The data set and label file which were outcomes of CellProfiler are taken as input by CP-Charm. Then CP-Charm GUItraintest file was run which input file of features and labels in csv form. There are many preferences provided by CP-Charm for classifier type, Validation type, Cross-validation fold type and number of training/testing rounds etc. These were the values which are used by present research work which are PCA-LDA classifier type, K-FOLD Cross validation where K=10 including no hold-out information having minimum 5 training/testing rounds. The output of this process was two files. One file is pcalda_classifier which is used to further classify and one is text file which shows the percentage of accuracies with mean value, standard deviation and median performances.

D. Model validation

Model is validated after the generation of pcalda_classifier with the classify mode which further generate predict label files which shows the result on test base that which patient have disease and which not. The model generated and validated here further implemented in next step in python language.

E. Implementation

Development of tool process is of two part designing and coding. The tool was developed in Python shown in Fig 3 Designing part of tool was done with help of software Glade3 for python.

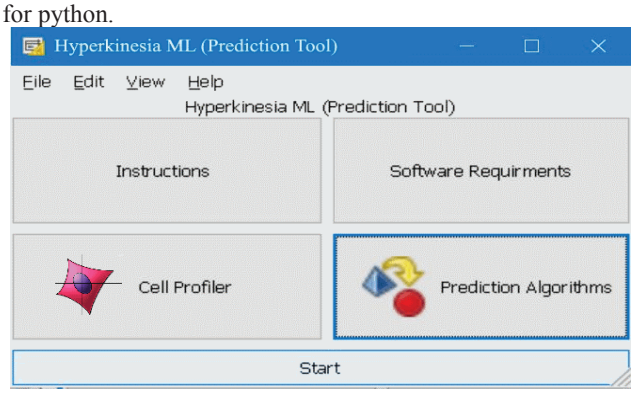

Fig 3 Tool

Glade is a rapid application development tool to allow fast and simple development of user interfaces for the GTK+ toolkit and the GNOME desktop environment. The user interfaces designed in Glade are saved as XML, and by using the GtkBuilder GTK+ object these can be loaded by applications dynamically as needed. By using GtkBuilder, Glade XML files can be used in numerous programming languages including C, C++, C#, Vala, Java, Perl, Python, and others. The tool designed here is having firstly four options showing Instruction, software requirement, CellProfiler and training and testing algorithm.

The instruction option shows instruction to use the tool. There are various steps to follow for the use of tool which is also come under the instruction part. The next is Software requirement option in tool prediction which shows different software needed by tool to make prediction. The next part is the CellProfiler this is used to extract feature and make csv files further needed by training and testing algorithm. The next option is testing and training algorithm which further gives two options that are Train/test and Classify. First option is Train test as it makes a model file which next used by classify mode Train test mode use output generated by CellProfiler and further make a pcalda model and result summary which have accuracies. In the last there is classify mode in which the model created in last phase and the feature file is given as an input to get a predict_lable file. The software is available for worldwide at https://github.com/lalaantika/HyperkinesiaML.

III. RESULTS

The steps are followed to obtain best accuracy of input MRI images. First of all, training of system is done by using various data set or sample. And then system is tested for few of the given sample, and accuracy is measured. The image set was divided into two portions. The first part is for training the system and the second was for testing purpose. 60% of the image set was used as training though the left over 40% was taken for the testing set done by CPCHARM. For each picture set, features were figured out and stored for training of data. The result summary shows as in TABLE I

TABLE I: -Accuracies

Accuracies					
Image set of 30 patients	***No of rounds in training/testing mode***	***Accuracies***	***Mean performance***	***Median performance***	***Standard deviation***
Merge of 25 images of patients	0	96.66	97.16	96.66	1.97
	1	93.33			
	2	96.66			
	3	96.77			
	4	98.33			
	5	95.0			
	6	98.33			
	7	100.0			
	8	96.66			
	9	100.0			
1st image of all patients	0	80.0	86.16	88.33	3.94
	1	88.33			
	2	88.33			
	3	80.0			
	4	83.33			
	5	83.33			
	6	88.33			
	7	88.33			
	8	91.66			
	9	90.0			

Accuracies					
Image set of 30 patients	***No of rounds in training/testing mode***	***Accuracies***	***Mean performance***	***Median performance***	***Standard deviation***
2nd image of all patients	0	91.66	87.5	88.33	4.90
	1	93.33			
	2	86,66			
	3	91.66			
	4	93.33			
	5	81.66			
	6	90.0			
	7	80.0			
	8	85.0			
	9	81.66			
3rd image of all patients	0	91.66	86.83	86.66	3.90
	1	85.0			
	2	86.66			
	3	91.66			
	4	81.66			
	5	88.33			
	6	85.0			
	7	86.66			
	8	80.0			
	9	91.66			
4th image of all patients	0	83.33	85.0	84.16	4.08
	1	78.33			
	2	81.66			
	3	86.66			
	4	91.66			
	5	91.66			
	6	81.66			
	7	86.66			
	8	85.0			
	9	83.33			
5th image of all patients	0	76.66	83.5	84.16	6.075
	1	88.33			
	2	86.66			
	3	80.0			
	4	83.33			
	5	85.0			
	6	71.66			
	7	95.0			
	8	83.33			
	9	85.0			
6th image of all patients	0	88.33	86.0	85.00	2.26
	1	88.33			
	2	85.0			
	3	86.66			
	4	85.0			
	5	81.66			
	6	85.0			
	7	85.0			
	8	85.0			
	9	90.0			
7th image of all patients	0	85.0	83.83	85.83	4.01
	1	88.33			
	2	78.33			
	3	78.33			
	4	81.66			
	5	86.66			
	6	86.66			
	7	88.33			
	8	78.33			
	9	86.66			
8th image of all patients	0	80.0			
	1	83.33			
	2	91.66			
	3	86.66			
	4	80.0			

Accuracies					
Image set of 30 patients	***No of rounds in training/testing mode***	***Accuracies***	***Mean performance***	***Median performance***	***Standard deviation***
	5	86.66	84.83	85.0	3.97
	6	86.66			
	7	90.0			
	8	83.33			
	9	80.0			
9th image of all patients	0	90.0	85.0	84.16	4.4
	1	85.0			
	2	83.33			
	3	83.33			
	4	90.0			
	5	80.0			
	6	80.0			
	7	85.0			
	8	93.33			
	9	80.0			
10th image of all patients	0	90.0	86.0	85.0	3.0
	1	83.33			
	2	83.33			
	3	85.0			
	4	86.66			
	5	91.66			
	6	85.0			
	7	81.66			
	8	88.33			
	9	85.0			
11th image of all patients	0	90.0	86.5	85.83	2.62
	1	86.66			
	2	88.33			
	3	83.33			
	4	85.0			
	5	86.66			
	6	85.0			
	7	91.66			
	8	83.33			
	9	85.0			
12th image of all patients	0	88.33	83.33	84.16	5.31
	1	90.0			
	2	76.66			
	3	88.33			
	4	83.33			
	5	81.66			
	6	90.0			
	7	73.33			
	8	85.0			
	9	81.66			
13th image of all patients	0	81.66	84.83	85.83	3.60
	1	78.33			
	2	88.33			
	3	81.66			
	4	86.66			
	5	86.66			
	6	90.0			
	7	88.33			
	8	85.0			
	9	81.66			
14th image of all patients	0	76.66	85.5	86.66	5.11
	1	86.66			
	2	90.0			
	3	91.66			
	4	86.66			
	5	76.66			
	6	90.0			
	7	88.33			
	8	86.66			
	9	81.66			

Accuracies					
Image set of 30 patients	***No of rounds in training/testing mode***	***Accuracies***	***Mean performance***	***Median performance***	***Standard deviation***
15th image of all patients	0	88.33	86.16	88.33	5.11
	1	76.66			
	2	88.33			
	3	88.33			
	4	85.0			
	5	88.33			
	6	91.66			
	7	93.33			
	8	83.33			
	9	78.33			
16th image of all patients	0	91.66	87.83	87.5	3.25
	1	85.0			
	2	88.33			
	3	86.66			
	4	88.33			
	5	86.66			
	6	85.0			
	7	95.0			
	8	83.33			
	9	88.33			
17th image of all patients	0	80.0	83.66	85.0	2.86
	1	86.66			
	2	86.66			
	3	86.66			
	4	85.0			
	5	81.66			
	6	81.66			
	7	85.0			
	8	78.33			
	9	85.0			
18th image of all patients	0	86.66	87.16	86.66	83.33
	1	86.66			
	2	90.0			
	3	91.66			
	4	83.33			
	5	91.66			
	6	90.0			
	7	81.66			
	8	86.66			
	9	83.33			
19th image of all patients	0	78.33	83.0	84.16	4.76
	1	85.0			
	2	81.66			
	3	86.66			
	4	86.66			
	5	71.66			
	6	88.33			
	7	83.33			
	8	86.66			
	9	81.66			
20th image of all patients	0	75.0	83.5	84.1	4.17
	1	86.66			
	2	86.66			
	3	85.0			
	4	81.66			
	5	86.66			
	6	83.33			
	7	80.0			
	8	90.0			
	9	80.0			
21st image of all patients	0	83.33			
	1	88.33			
	2	83.33			
	3	86.66			
	4	86.66			
	5	86.66	85.0	85.83	3.72
	6	78.33			
	7	91.66			
	8	80.0			
	9	85.0			
22nd image of all patients	0	85.0	83.83	84.16	2.79
	1	80.0			
	2	88.33			
	3	86.66			
	4	81.66			
	5	83.33			
	6	80.0			
	7	85.0			
	8	86.66			
	9	81.66			
23rd image of all patients	0	96.66	86.66	85.83	5.05
	1	93.33			
	2	83.33			
	3	83.33			
	4	88.33			
	5	83.33			
	6	85.0			
	7	86.66			
	8	88.33			
	9	78.33			
24th image of all patients	0	86.66	84.5	85.0	3.80
	1	80.0			
	2	85.0			
	3	91.66			
	4	80.0			
	5	85.0			
	6	86.66			
	7	80.0			
	8	81.66			
	9	88.33			
25th image of all patients	0	85.0	84.5	84.16	2.89
	1	81.66			
	2	88.33			
	3	81.66			
	4	86.66			
	5	81.66			
	6	85.0			
	7	90.0			
	8	81.66			
	9	83.33			

IV. Discussion

Hence although over the years a big number of prediction tools are present but there is no such tool which works on cardiac diseases. There are some workings in area of MRI brain prediction yet, to date to our knowledge, there is no resource available that provides prediction tools for heart diseases. Therefore, this research has compiled a method, where manual creation along with information from other resources have been integrated to provide a tool that will allow researchers, doctors and clinicians to get prediction of diseases on bases of MRI test. This prediction tool is a user friendly and free available resource tool which provides predictions on the bases of MRI test. There is hope that the availability of this tool would save effort and time of specialists involved in the field and thus will facilitate

the doctors, clinicians for heart diseases diagnosis, treatment and prevention.

MRI data is taken from IGMC Shimla which is in form of Dicom images. There are total 60 patients on which work is followed. The MRI test report of each patient has 900-1200 images. Features are extracted from these images and then testing and training part is done. The feature extracted from images was 957. After testing it gives accuracies of prediction. This tool helps doctors and radiologist to make prediction in an easy way.

As tool is created for clinicians, researchers, doctors which are working in different locations, it is important to provide an easy, open source way to diagnose disease. Thus a tool give prediction of disease is the best solution.

V. CONCLUSION

Global hyperkinesia is a foremost health problem in whole world which leads to various other severe complications for example heart attack. The fact that only some of people testified being identified as an effect of screening indicates that are missing opportunities to prevent heart disease. There are various imaging tests present for heart in present research MRI images is used. Considering the importance of heart diseases, numerous computational techniques have been created for prediction of imaging tests and also associated complications. These image prediction methods are relying on machine learning. The collection of data is from hospital. There are 30 patients of each category which are taken under consideration. Having set of 900-1200 Dicom images. The images can't be used for machine learning directly so CellProfiler is used for feature extraction part. In the projected study CP-CHARM is cast-off for the model building using MRI images, with this model 80-85% accuracy is achieved with the 25 image set and 97% accuracy by merging imaging using Amide no other method until now is used for the heart disease prediction. This developed method is implemented in python and available for people in worldwide web.

As per less data availability of data prediction accuracy can be increased by using more patients' data. This approach can be comprehensive for development of software that will be GUI which helps Doctors to predict diseases. Similar tool (methods) may be developed for other complex disease of heart like hypotropic cardiomyopathy etc.

REFERENCES

[1] Cellprofiler.org. (2016). CellProfiler. [Internet] Available at: http://cellprofiler.org/ [cited 1 Sep. 2015].

[2] Global Hypokinesis - Symptoms, Definition, Causes, Treatment. [Internet] Medical Treasure 2014 July 26 [cited 14 March 2016]. Available at: http://medicaltreasure.com/global-hypokinesis/.

[3] Gonzalez, Rafael C and Richard E Woods. Digital Image Processing. Upper Saddle River, N.J.: Prentice Hall. Print; 2002.

[4] Jackway, P.T., 1996. Gradient watersheds in morphological scale-space. IEEE Transactions on Image Processing, 5(6), pp.913-921.

[5] Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M. et al.American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics–2015 update: a report from the American heart association[internet]. Circulation. 2015;131(4):e29–322. [cited 14 Apr 2016] Available at: http://www.ncbi.nlm.nih.gov/pubmed/25520374.

[6] Padmavathi, G., Muthukumar, M. and Thakur, S.K., 2010. Nonlinear image segmentation using fuzzy c-means clustering method with thresholding for underwater images. IJCSI, 7(3), p.35.

[7] R.C. Gonzalez, R.E. Woods and S.L.Eddins, "Digital image processing using MATLAB", Second edition, Gatesmark publishing, USA, 2009

[8] Uhlmann, V., Singh, S. and Carpenter, A.E., 2016. CP-CHARM: segmentation-free image classification made accessible. BMC bioinformatics, 17(1), p.1.

[9] Vincent, L. and Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE transactions on pattern analysis and machine intelligence, 13(6), pp.583-598.