DERIVATIVE INVESTIGATION OF SARCOMA THROUGH MACHINE LEARNING

Mashal Tariq[1], Shehla Andleeb [2]
[1,2] Electrical Engineering Department, Usman Institute of Technology
Str-13, Block 7, Gulshan-e-Iqbal, AbulHassan Isphahani Road, Karachi, Pakistam
[1]mtariq@uit.edu, [2] shandleeb@uit.edu

ABSTRACT- *Computer aided detection or diagnostics technology has transformed the medical field. Its knowledge and application provide huge assistance to medical professionals, especially, pathologists. Classification of medical data is yet another considerable area of research where computer technology shows meaningful and decisive development. It has been discussed over a decade and still more expert knowledge is required in finding features and formulating system of classifiers. Brain Tumor is critical disease which has a brunt consequences on human quality of life. Early tumor detection is an extensive field of research. Our work deals with classification of brain tumor using machine learning WEKA software. Our targeted tumor type is Solid cum cystic. Our training sets consist of four normal and tumor cases with mean intensity and variance features selected. The features are calculated from images acquired from Magnetic Resonance Imaging modality. Proposed study is a hybrid methodology for tumor data classification.*

Keywords: Brain, Computer aided diagnoses, tumor, machine learning, classification

I. INTRODUCTION

Medical data analysis has been under discussion for decades. Recorded Data Type whether it is narrative, numerical, recorded signals or pictures is considered of higher priority for prognosis and treatments. Recording of data is generally done manually which is time consuming. Automation in same respect is requirement of this time. In today's era the fully automated and semi automated system of data recording and sorting is high in demand. Detection of patterns of existing data and future data through machine learning is discussed comprehensively [1]. It includes background material of probability and linear algebra as well as discussions on recent developments. Medical Statistical analysis [2] is explained in details inclusive medical literature with appropriate approach. Research in medical data analysis has revolutionized and assist medical professionals. In addition, it raises the quality of patients examinations. It is a ongoing investigation area of work. Machine learning is a technology which provides immense support to therapeutic field. Classifiers systems play a vital role in machine learning. It has wide range of application irrespective of data types for machine learning [3] are described with various examples.C4.5 has cases of nominal and numerical properties with distinctively described reliably discriminated classes. Narrative data [4] of motor vehicles accidents categories are analyzed through Bayesian in fencing. The proposed method is able to identify pre-crash and crash type out of 3686 motor vehicles crashes. Narratives is defined with 63 keywords. The results obtained from Bayesian model and keyword search varied from 12% to 67%. Importance of Medical data inclusive ethical and legal aspects are discussed in detail. Mathematical knowledge concerning medical data differs from its collection activities [5]. Artificial intelligence has application in very unique areas. Emotions is very crucial to quantify. Emotions can be distinguished with \artificial intelligence help [6].

Brain is a convoluted and elaborative structure. Tumor is brain has a great impact on its functioning which results affecting quality of human life. Tumors are due to accumulation of abnormal growth of normal cells. They can be: benign or malignant. Malignant is a cancerous one. Magnetic Resonance Imaging is a valuable modality to analyze brain development and diseases diagnoses [7]. There is a need of formulating an automated system for tumor detection. Mashal and Shehla [8] discussed various techniques for tumor segmentation. Mu Zhou [9] proposed Synthetic Minority Over Sampling Technique for survival rate

using imbalanced data set(skewed distribution) of multi modal MRI of 42 patients with Glioblastoma Multiforme (GBM) tumors with 95.24% accuracy.

Our study comprises of numerical data. We consider mean intensity and variance values generated from Magnetic Resonance Imaging images of normal or tumor cases. The classifiers used in our work are discussed [10, 11, 12, 13] briefly. This paper consist of -----section: first deals with brief introduction with literature review, second brief materials used, third describes the proposed methodology.

II. MATERIALS

MATLAB is used for feature selection and calculation Magnetic Resonance imaging images acquired from General Electric Signa HDe 1.5 Tesla are used with their respective radiology reports. Four normal case and cases with tumor are selected in axial plane with three slices. T2 weighted MRI imaging modality is considered in this work. The dimensions of all images are 512 × 512 pixels in JPEG format. Solid cum Cystic type is selected. WEKA is used for machine classification.

III. METHODOLOGY

The proposed methodology is shown in Figure 1. The details of important steps are as under.

1) All cases consist of 19 slices. For our analysis we selected 1^{st}, 10^{th} and 19^{th}.

2) Features are studied. Mean intensity and variance of MRI images are calculated with MATLAB. Mean and Variance is given by:

Mean intensity and variance is given by:

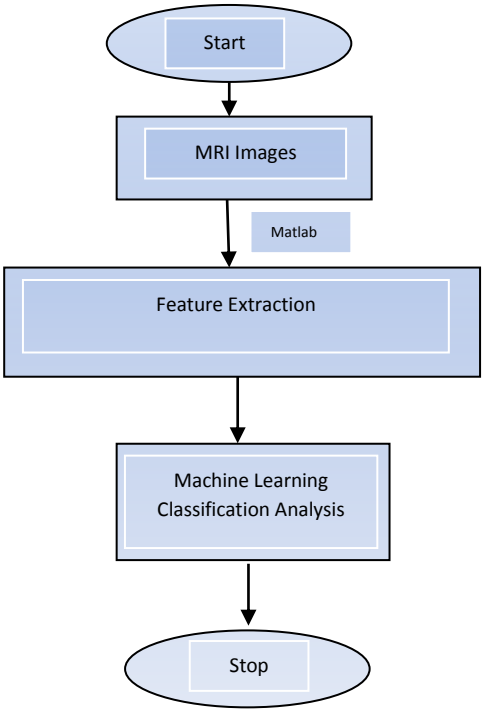
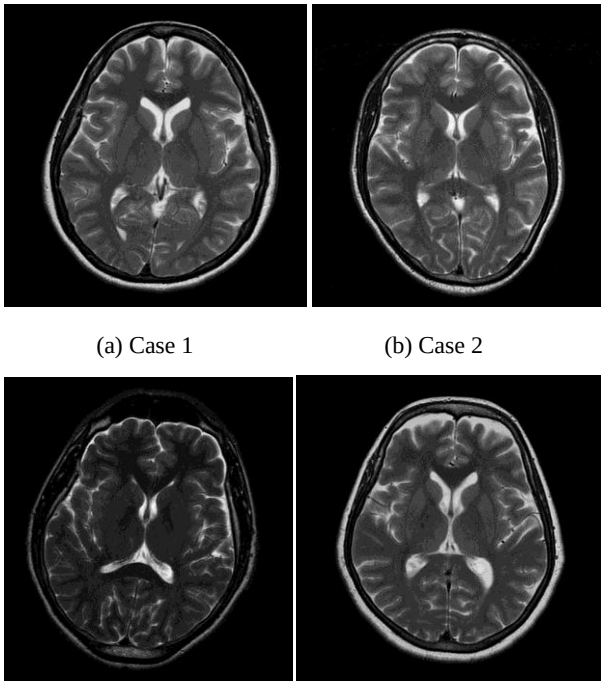


Figure 1. Proposed Methodology



(a) Case 1 (b) Case 2

(c) Case 3 (d) Case 4

Figure 2 Intensity Variation in Normal MRI images

Table I. Mean Intensity And Variance Vaues Table

	Intensity			Variance		
	Slice 1	Slice 10	Slice 19	Slice 1	Slice 10	Slice 19
	Normal			Normal		
Case 1	33.01661	44.56847	37.42167	2610	2580	2400
Case 2	36.54915	41.53649	36.29208	2400	2540	2310
Case 3	21.93443	27.54558	25.35077	1780	1170	1160
Case 4	31.27333	45.36121	40.21143	2330	2330	1160
	Tumor			Tumor		
Case 1	30.2626	42.58475	46.342	1160	2760	1970
Case 2	22.51599	41.85863	30.11685	1000	2380	1000
Case 3	22.63299	36.05094	26.73034	1580	2840	1900
Case 4	14.66106	33.5486	23.56134	2110	2850	3370

I. Table Ii. Classification Accuracy & Time Effciency

Features	Classifier	Time to built model (sec)	Classification accuracy
Mean Intensity	Multi layer Perceptron	0.04	55%
	Classification Via Regression	0.12	50%
	Nested Dichotomies	0	55%
	Rotation Forest	0.02	55%
	PART	0	55%
	Random Forest	0	55%
	Multiclass Classifier	0	46%
Variance	Multi layer Perceptron	0.06	29%
	Classification Via Regression	0.1	38%
	Nested Dichotomies	0.01	29%
	Rotation Forest	0.03	33.33%
	PART	0.01	25%
	Random Forest	0.02	29.1667%
	Multiclass Classifier	0.01	25%
Mean Intensity & Variance	Multi layer Perceptron	0.05	55%
	Classification Via Regression	0.1	46%
	Nested Dichotomies	0.01	42%
	Rotation Forest	0.03	59%
	PART	0.01	25%
	Random Forest	0.02	46%
	Multiclass Classifier	0.01	46%

$$I_M = \frac{\sum_{n=1}^{N} I_n}{N} \quad (1)$$

and

$$V = \frac{\sum (I - \bar{I})^2}{n} \quad (2)$$

Whereas I_M represents mean intensity in equation 1. In equation 2, $\bar{I}$ the mean intensity value and n is the total data values. In our case n = 512.

3) Classification accuracy is studied with WEKA machine learning software.

4) The following classifiers are tested

i) Multilayer Perceptron (MLP)

ii) Classifier via Regression

iii) Nested Dichotomies

v) Random Forest

vi) Multiclass Classification.

vii) Part Classifiers

5) Results are examined and analyzed. Tables are tabulated and results is displayed analytically as well as graphically.

IV. RESULTS & DISCUSSIONS

Table I represents features values generated by MATLAB of MRI images. For Slice 1 (normal) intensity ranges from about 31 to 36.All values are somehow very close to each other. There is asmall fluctuation of values are seen in some cases. These are due to the brightness of original images as can be seen in Figure 2.It shows slices from normal cases, it is clear that case 3 is darker in brightness. Sometimes size of a skull also matters. But for tumor it is clearly shown the difference is larger as the tumor differs in shape and size. Even its location also varies from case to case and slice to slice. Also, cystic and solid part also varies in all respect.

The comprehensive experimental results of machine learning classifiers are shown in Table II generated by WEKA. The reason of using 7 classifiers is to compare performance with each other for effective results. As can be seen mean intensity shows accuracy of 46% with Multiclass classifier and 55% with Multilayer Perceptron. Accuracy with variance a s feature doesn't show promising result ranging from 25% to 38%. The third training set with mean intensity and variance combined show better results of 59% with Rotation Forest

and 55% with Multi layer Perceptron. Overall analysis shows that mean intensity is a strong feature in classification of tumor. For solid cum cystic type of tumor Multiple layer Perceptron classifier can be considered.

All the classifiers result are very time efficient. Multiple layer Perceptron shows 0.04 for mean intensity, 0.06 for a variance and 0.05 for mean intensity and variation combined. Figure 3 provides the comprehensive statistical analysis of intensity based features of brain individually and in combination and it can be clearly observed that mean intensity is need performing a significant role performance of learning classifiers. Blue color representing mean, pink variance and green intensity and variance .

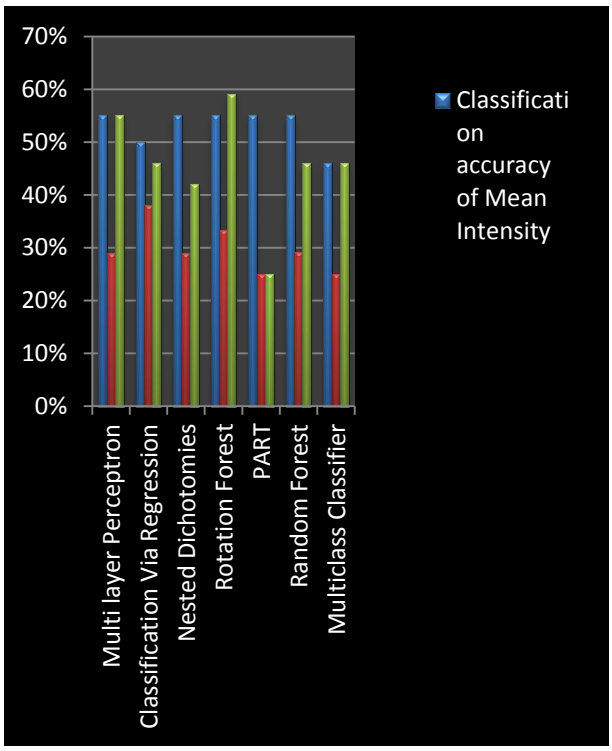


Figure 3 Graphical Representation of Classification Accuracy of Features combined

V. FUTURE WORK

The same work can be extended to all slices of same cases and other types of brain tumor can also be considered. Other features can be added.

REFERENCES

[1] Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective-*Adaptive computation and machine learning series*", ISBN 0262018020, 9780262018029, MIT Press, (2012)

[2] Kirkwood, B. R. "Essentials of medical statistics. Essentials of medical statistics", pp. xii + 234pp.ISBN:0-632-01052-5, (1988)

[3] J. Ross Quinlan, "C4.5: Programs for Machine Learning-*Morgan Kaufmann series in machine learning*", ISBN0080500587, 9780080500584, *Elsevier*, (2014)

[4] Lehto MR , Sorock GS , "Machine learning of motor vehicle accident categories from narrative data". (PMID:9019094), *Methods of Information in Medicine* ,Europe PMC, **35**(4-5):309-316, (1996).

[5] Krzysztof J. Cios, G. William Moore, "Uniqueness of medical data mining", *Artificial intelligence in medicine* **Volume 26**, Issues 1-2, Pages 1–24, , © 2002 Elsevier Science B.V. Published by Elsevier September–October, (2002).

[6] Syed Abbas Ali, Maria Andleeb Siddiqui, Anas Khan, "Analyzing the influence of Prosodic Features on Emotional states Recognition in Regional Languages of Pakistan" , *International Multi-Topic Conference*,Mehran University, Jamshoro,Pakistan, (2015).

[7] Sudipta Roy, Samir K. Bandyopadhyay, "Detection and Quantification of Brain Tumor from MRI of Brain and it's Symmetric Analysis", *International Journal of Information and Communication Technology Research*, **Volume 2** No. 6,pg:477-483, ISSN 2223-4985 (June 2012)

[8] Mashal Tariq, Shehla Andleeb, Dr. Hira Andleeb, "Analysis of Tumors through Image Processing Techniques" *Conference of Life Science Research*-2015, Abstracts, Asian J Agri Biol, Special Issue-2015,pg: 18.(2015)

[9] Mu Zhou ; Lawrence O. Hall ; Dmitry B. Goldgof ; Robert J. Gillies et al., "Imbalanced learning for clinical survival group prediction of brain tumor patients", *SPIE Proceedings, Medical Imaging 2015: Computer-Aided Diagnosis* ,**Volume 9414**, (March 20, 2015)

[10] S. Bibi' , G. Tsoumakas , I. Stamelos , I. Vlahavas ,"Regression via Classification applied on software defect estimation", *Expert Systems with Applications*, Elsevier Ltd, **Volume 34**, Issue 3, Pages 2091–210,(April 2008)

[11] M.C. Fairhurst and M.S. Hoque, "Moving window classifier: approach to off-line image recognition", *Electronics Letters*, **Volume 36,** Issue 7, p. 628 – 630, Print ISSN 0013-5194, Online ISSN 1350-911X, (2000)

[12] Andy Liaw and Matthew Wiener, "Classification and Regression by random Forest", *R News*, **Vol. 2/3**,ppg# 18-21, (December 2002).

[13] Rodriguez, J.J.; Kuncheva, L.I.; Alonso, C.J., "Rotation Forest: A New Classifier Ensemble Method," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , **vol.28,** no.10, pp.1619-1630, (Oct. 2006),