ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

An Efficient Classifier Approach for Brain Tumor Prediction.

Sumathi A*, Poornima M, Kalaichelvi V, Manimozhi K and Meenakshi P.

SASTRA University, Kumbakonam, Tamil Nadu, India.

ABSTRACT

Decision tree induction algorithms have been used in many application areas one of these area is medical diagnosis. Several Decision tree algorithms have been used to classify the brain tumor and its types. A non-backtracking decision tree algorithm c4.5 adopts greedy approach in which decision trees are constructed in top down recursive divide and conquer manner. This algorithm predicts categorical class label with discrete attributes. But some of the cases it is hard to classify the subtypes because of distance based attribute and it leads to less prediction accuracy, which is the major drawback. To overcome this issue, this projected work proposed a new technique called Minimum Spanning Tree Classifier (MSTC) which improves the classification accuracy in terms of removing irrelevant features from distance valued attributes, which leads to achieve an efficient prediction. This proposed MSTC helps to predict the category of the brain tumor with accuracy and which leads to proper discretion for medical treatment.

Keywords: Decision Tree, Information Gain, C4.5, MSTC

Corresponding author

ISSN: 0975-8585

INTRODUCTION

The term 'Brain tumor' is a disease which affects brain and central spinal canal. They are classified into primary and secondary, in which the primary brain tumor occurs due to environmental and genetic disorders and the secondary brain tumors [1], occurs due to cancer in other parts of the body spreading to the brain cells. Data Classification [2] is one of the familiar techniques which are used to categorize the large data base with the use of discrete attributes. The goal of classification is to accurately predict the target class for each case in the data. Decision trees (DT)[3] is a classification technique which provides possible solutions, based on sequence conditions of continues and discrete attributes. Decision tree [4] induction algorithms have been used in many application areas one of these area is medical diagnosis. Several Decision tree algorithms have been used to classify the brain tumor and its types. In the existing system c4.5 [5] demonstrates better accuracy for biological data classification. However this work has identified that the classification accuracy of c4.5 is not suitable for distance based attributes and some of the cases it is hard to classify the subtypes because of distance based attribute and it leads to less prediction accuracy.

METHODOLOGY

This work proposed an efficient classifier to classify brain tumor dataset. To achieve this, this proposed work is implemented c4.5[6] with minimum spanning tree approach called minimum spanning tree classifier which improved the classification accuracy in terms of removing irrelevant features from hard training dataset which leads to achieve an efficient prediction. It improves classification accuracy by find the false positive data at minimum level and it removed from the dataset and apply the greedy search for find data easily which helps to predict type of tumor in an efficient way. This proposed model was evaluated with tumor dataset. Patient age, gender, tumor location and tumor growth rate are taken as input.

Information Gain

This splitting criterion minimizes the information needed to classify the attributes and also minimizes the expected number of tests needed to classify the attributes. It calculates using the following formula

Gain(A)= Info(D)- Infox (D)

where Info(D) and Infox(D) are given as,

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2 p_i$$

$$Info_A(D) = \sum_{i=1}^{v} \left(\frac{|D_i|}{|D|} \right) \times Info(D_i)$$

The attribute which have high information gain is chosen as the splitting attribute and it is the root node. But Information Gain bias towards the attributes on large number of values.

C4.5 Classification

C4.5 algorithm was developed with the information gain attribute selection measure for construct decision tree for classification and the input data to be classified based on this method to find the type of tumor and calculated the classification accuracy using confusion matrix.

Proposed Methodology

MSTC Classification

The new technique called MSTC (Minimum Spanning Tree Classifier) algorithm was developed for classification. Minimum spanning tree is to start with an empty value and try to add minimum related values one at a time, always making sure that what is built remains acyclic. The resulting values are always is a subset of some minimum spanning tree. First find the some sure false positive values of dataset then remove that

ISSN: 0975-8585

data from the database and construct the decision tree based on c4.5 approach and apply minimum spanning tree approach for find particular growth rate from the dataset then predict the type of tumor and its states. Classification accuracy is calculated using confusion matrix **[Fig. 1].**

MSTC Algorithm

Get training dataset D with n data object

if D is not empty
    If Dj is false data
        Remove Dj from D
if attribute is discrete attribute
        attribute_list = attribute_list+attribute

Apply attribute selection method (D , attribute_list)
Generate Decision tree(D , attribute_list)

repeat
    Apply Minimum spanning tree in terminal node
    If (input_data== Dj)
        Predict entity for class variables
        Break;
Until j not greater than n

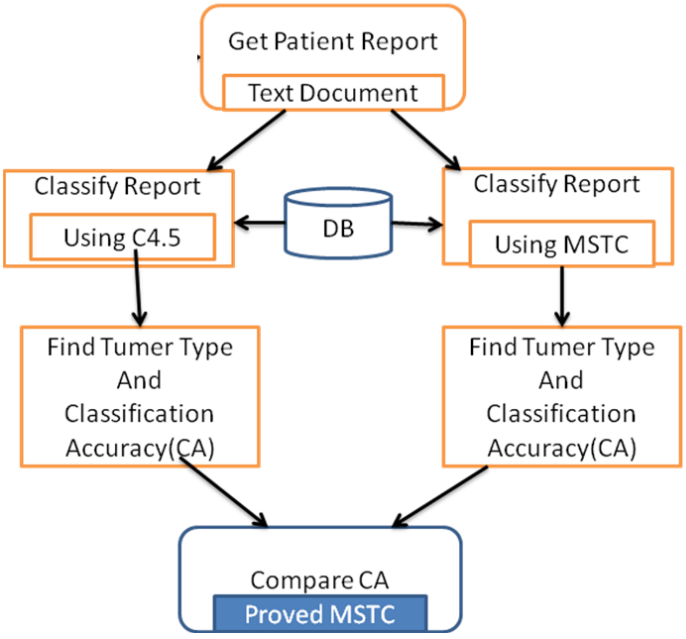


Fig.1 Proposed Methodology

ISSN: 0975-8585

Implementation

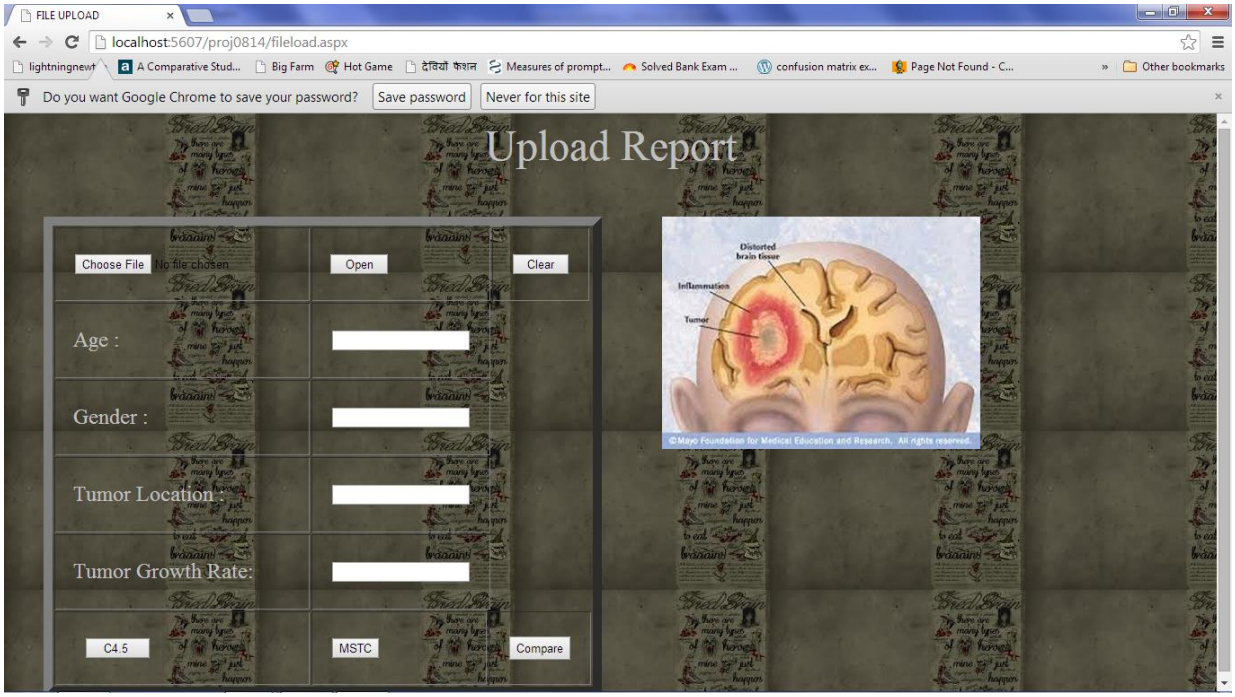


Fig.2 Upload dataset

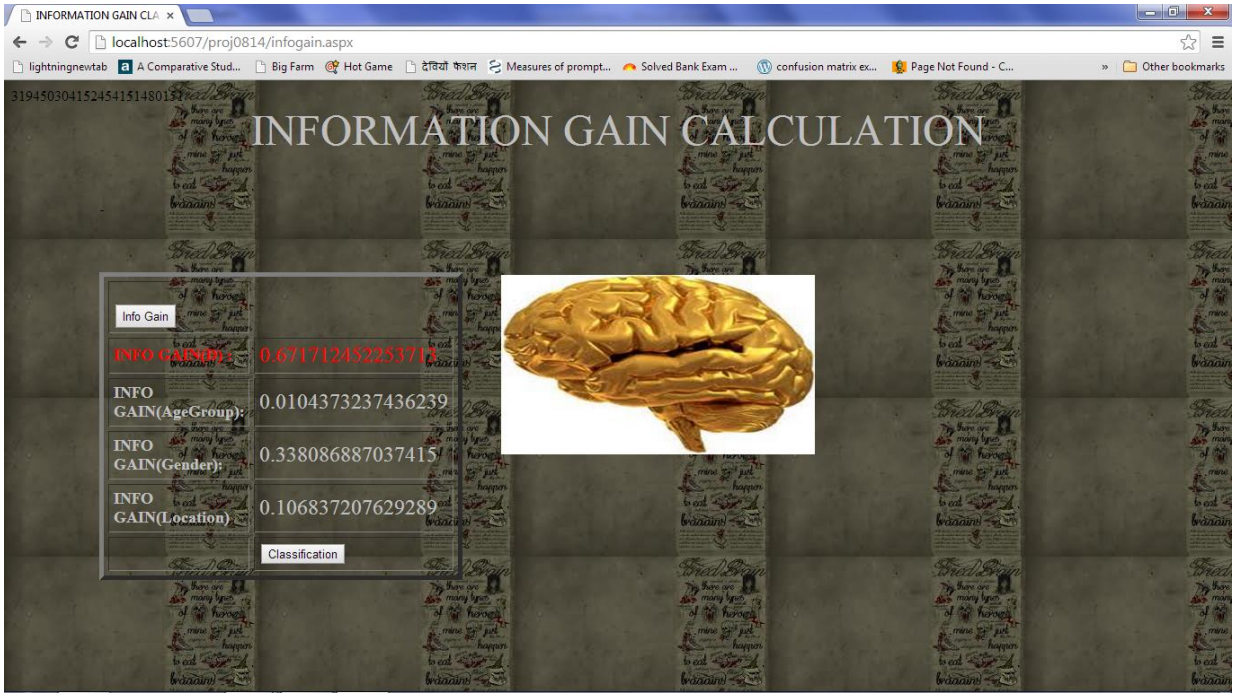
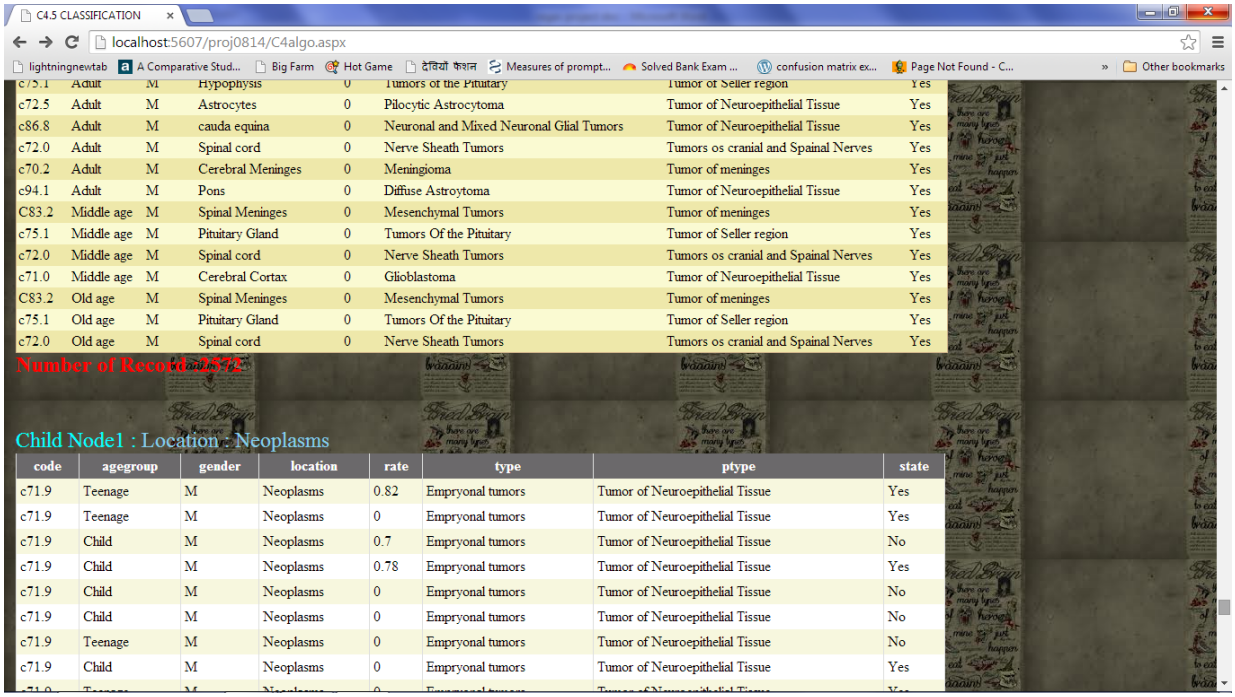


Fig.3 Information Gain Calculation

ISSN: 0975-8585

code	agegroup	gender	location	grate	type	ptype	state
c72.5	Teenage	M	Astrocytes	0	Pilocytic Astrocytoma	Tumor of Neuroepithelial Tissue	Yes
c71.0	Old age	M	Cerebral Cortax	0	Glioblastoma	Tumor of Neuroepithelial Tissue	Yes
C75.1	Older	M	Cranial	0.21	Craniopharyngioma	Tumor of Seller region	No
c71.0	Child	M	Ventricle	0.65	Glioma maliganant, NOS	Tumor of Neuroepithelial Tissue	No
C93.9	Child	M	Choroid Plexus	0	Choroid Plexus Tumors	Tumor of Neuroepithelial Tissue	Yes
c71.9	Teenage	M	Neoplasms	0.82	Empryonal tumors	Tumor of Neuroepithelial Tissue	Yes
c71.0	Middle age	M	Cerebral Cortax	0.14	Glioblastoma	Tumor of Neuroepithelial Tissue	Yes
C30.0	Old age	M	nasal Cavity	0.12	Olfactory Tumor	Tumor of Brain	No
c71.0	Middle age	M	Cerebral Cortax	0.3	Glioblastoma	Tumor of Neuroepithelial Tissue	Yes
C71.7	Teenage	M	Mid Brain	0	Brain Stem Glioma	Germ Cell tumor	No
c86.8	Teenage	M	cauda equina	0.33	Neuronal and Mixed Neuronal Glial Tumors	Tumor of Neuroepithelial Tissue	No
c71.0	Middle age	M	Cerebral Cortax	0	Glioblastoma	Tumor of Neuroepithelial Tissue	Yes
c71.0	Teenage	M	Ventricle	0.4	Glioma maliganant, NOS	Tumor of Neuroepithelial Tissue	No
c75.1	Middle age	M	Pituitary Gland	0.23	Tumors Of the Pituitary	Tumor of Seller region	Yes
c72.0	Old age	M	Spinal cord	0.28	Nerve Sheath Tumors	Tumors os cranial and Spainal Nerves	No
C75.3	Middle age	M	Pineal Gland	0	Tumor of The Pineal Region	Tumor of Neuroepithelial Tissue	Yes
c71.9	Teenage	M	Neoplasms	0	Empryonal tumors	Tumor of Neuroepithelial Tissue	Yes
c72.5	Adult	M	Astrocytes	0	Pilocytic Astrocytoma	Tumor of Neuroepithelial Tissue	No
c72.5	Teenage	M	Astrocytes	0.92	Pilocytic Astrocytoma	Tumor of Neuroepithelial Tissue	Yes
C95.9	Teenage	M	Lymphatic	0.35	Lymphoma	Tumor of Lymphoma	Yes
c72.5	Child	M	Astrocytes	0.9	Pilocytic Astrocytoma	Tumor of Neuroepithelial Tissue	Yes

Fig.4 C4.5 Classification



code	agegroup	gender	location	rate	type	ptype	state
c71.9	Teenage	M	Neoplasms	0.82	Empryonal tumors	Tumor of Neuroepithelial Tissue	Yes
c71.9	Teenage	M	Neoplasms	0	Empryonal tumors	Tumor of Neuroepithelial Tissue	Yes
c71.9	Child	M	Neoplasms	0.7	Empryonal tumors	Tumor of Neuroepithelial Tissue	No
c71.9	Child	M	Neoplasms	0.78	Empryonal tumors	Tumor of Neuroepithelial Tissue	Yes
c71.9	Child	M	Neoplasms	0	Empryonal tumors	Tumor of Neuroepithelial Tissue	No
c71.9	Child	M	Neoplasms	0	Empryonal tumors	Tumor of Neuroepithelial Tissue	No
c71.9	Teenage	M	Neoplasms	0	Empryonal tumors	Tumor of Neuroepithelial Tissue	No
c71.9	Child	M	Neoplasms	0	Empryonal tumors	Tumor of Neuroepithelial Tissue	Yes

Fig.5 C4.5 Classification

ISSN: 0975-8585

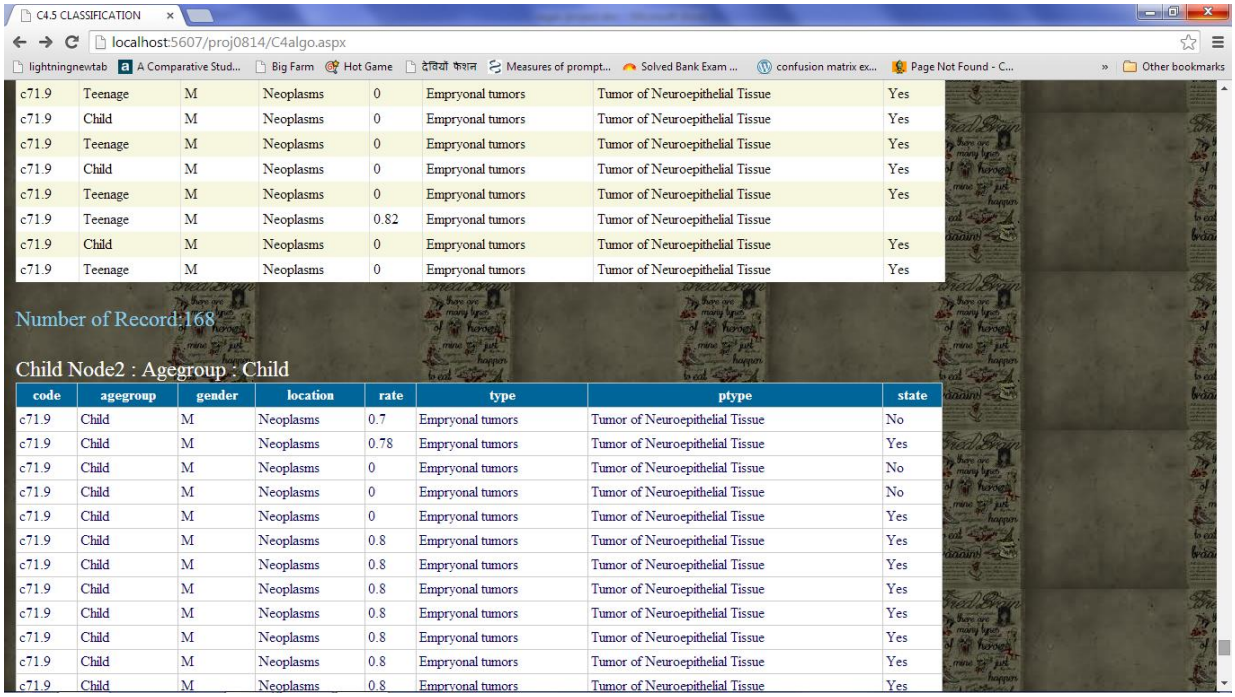


Fig.6 C4.5 Classification

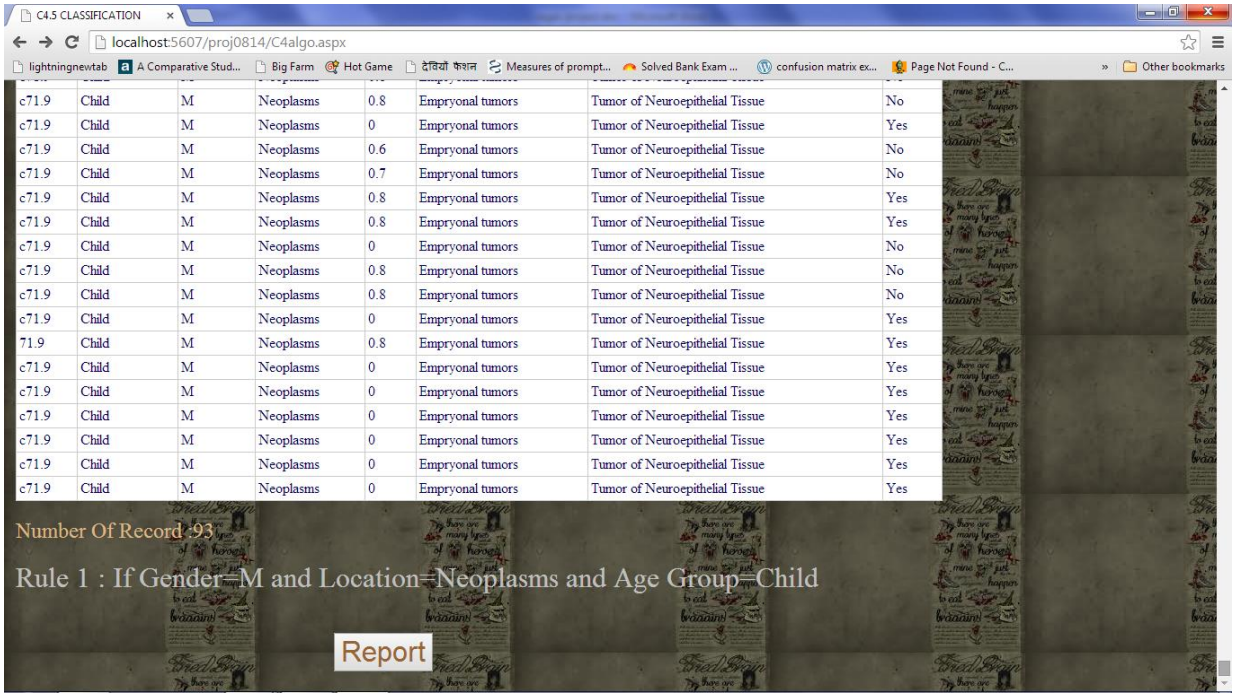


Fig.7 C4.5 Classification

ISSN: 0975-8585

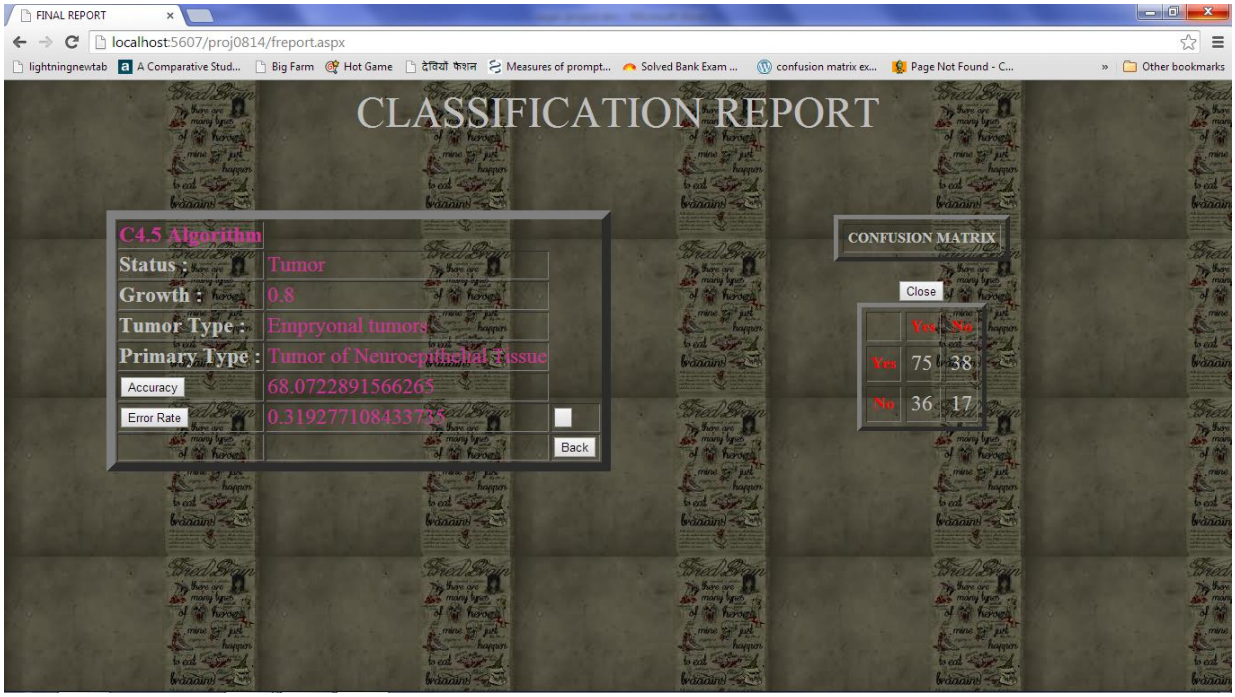
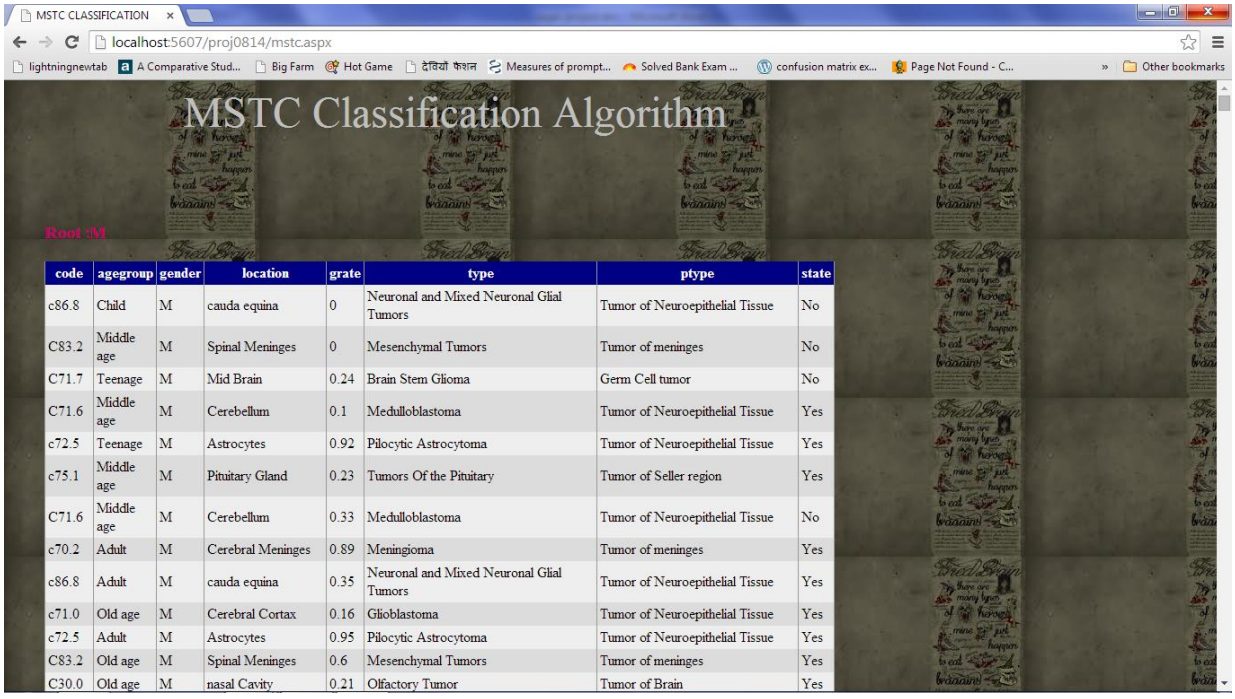


Fig.8 Classification Report of c4.5



code	agegroup	gender	location	grate	type	ptype	state
c86.8	Child	M	cauda equina	0	Neuronal and Mixed Neuronal Glial Tumors	Tumor of Neuroepithelial Tissue	No
C83.2	Middle age	M	Spinal Meninges	0	Mesenchymal Tumors	Tumor of meninges	No
C71.7	Teenage	M	Mid Brain	0.24	Brain Stem Glioma	Germ Cell tumor	No
C71.6	Middle age	M	Cerebellum	0.1	Medulloblastoma	Tumor of Neuroepithelial Tissue	Yes
c72.5	Teenage	M	Astrocytes	0.92	Pilocytic Astrocytoma	Tumor of Neuroepithelial Tissue	Yes
c75.1	Middle age	M	Pituitary Gland	0.23	Tumors Of the Pituitary	Tumor of Seller region	Yes
C71.6	Middle age	M	Cerebellum	0.33	Medulloblastoma	Tumor of Neuroepithelial Tissue	No
c70.2	Adult	M	Cerebral Meninges	0.89	Meningioma	Tumor of meninges	Yes
c86.8	Adult	M	cauda equina	0.35	Neuronal and Mixed Neuronal Glial Tumors	Tumor of Neuroepithelial Tissue	Yes
c71.0	Old age	M	Cerebral Cortax	0.16	Glioblastoma	Tumor of Neuroepithelial Tissue	Yes
c72.5	Adult	M	Astrocytes	0.95	Pilocytic Astrocytoma	Tumor of Neuroepithelial Tissue	Yes
C83.2	Old age	M	Spinal Meninges	0.6	Mesenchymal Tumors	Tumor of meninges	Yes
C30.0	Old age	M	nasal Cavity	0.21	Olfactory Tumor	Tumor of Brain	Yes

Fig.9 MSTC Classification

ISSN: 0975-8585

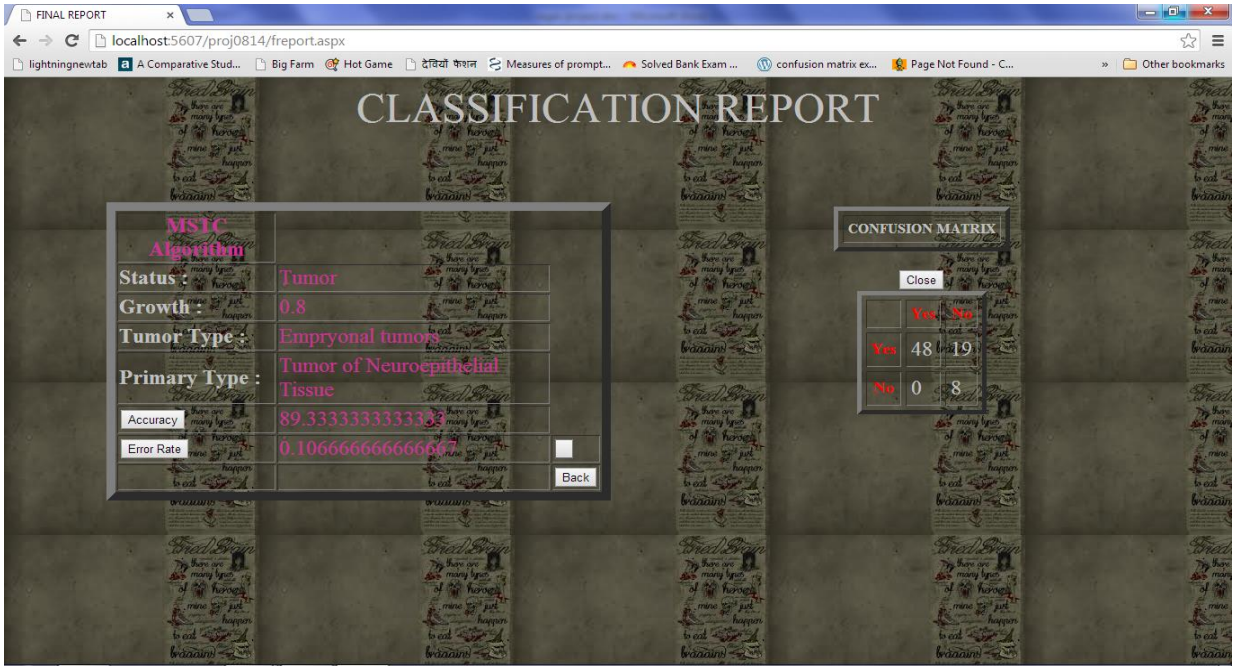


Fig.10 Classificatiion Report of MSTC

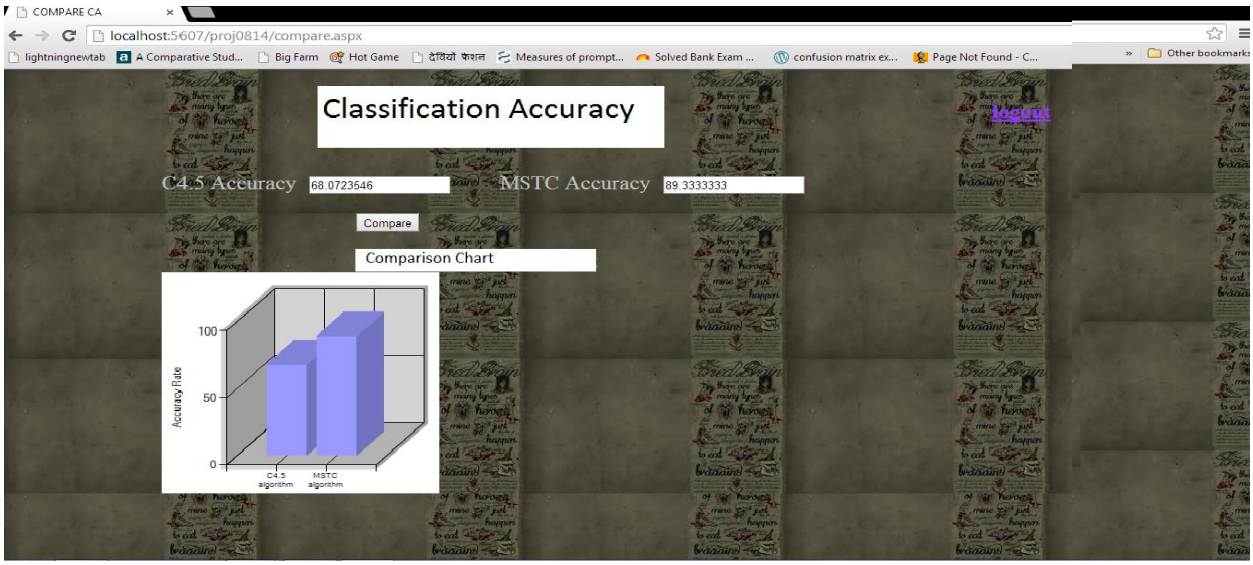


Fig.11 Comparison of Classification Accuracy

Fig.12 Classification Accuracy of c4.5

ISSN: 0975-8585

Fig.13 Classification Accuracy of MSTC

CONCLUSION

C4.5 is one of the familiar decision tree techniques. C4.5 algorithm performs well in constructing decision trees and extracting rules from the biological dataset. However, for biological dataset Minimum Spanning Tree Classifier algorithm is needed in order to provide better classification accuracy. Minimum spanning tree concept is used for finding data in dataset as much as faster than exiting system. From the experiment results, MSTC achieved high classification accuracy than c4.5 and it also gives less error rate. MSTC helps to predict the category of the tumor with accuracy and which leads to proper discretion for medical treatment.

REFERENCES

[1] Eman M. Ali et al. International Journal of Computer Applications 2015; 131(2).
[2] P Rajendran , M Madheswaran. Journal of Medical Systems 2012; 36(2):747-764.
[3] Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers,2008;3rd reprint.
[4] R Jin, and G. Agrawal, In Proc. of ACM SIGKDD, 2003; 571-576.
[5] A Sumathi et al. Journal of Theoretical and Applied Information Technology 2012; 35(2):208-213.
[6] P Shanthakumar and P Ganeshkumar. Computers & Electrical Engineering 2015; 45: 302-311