

Morphology-based feature classification between childhood medulloblastoma and Normal brain cells using Neural Networks

Daisy Das¹, Dr. Lipi B. Mahanta², Dr. Shabnam Ahmed³

^{1,2}Central Computational and Numerical Sciences Division,

Institute of Advanced Study in Science and Technology, Guwahati (India)

³Pathology Department, Guwahati Neurological Research Centre,

Sixmile, Guwahati (India)

ABSTRACT

In this paper, we proposed a framework for cell level classification between normal and abnormal biopsy samples of childhood brain tumor with emphasis to childhood medulloblastoma, a most common childhood brain tumor. Till best of our knowledge, all work on childhood medulloblastoma is on its architectural classification using texture-based features. The paper presents morphology-based classification between normal and abnormal samples using feedforward neural network and cascaded neural network. Since cells are an essential biomarker for any malignancy of biopsy samples, we tried to evaluate the virulence of childhood medulloblastoma based on its shape and size compared to standard cell samples.

Keywords—*medulloblastoma, morphology, cell level, classification, neural network, etc.*

I.INTRODUCTION

Childhood medulloblastoma cells are highly malignant and constitute about 20%[1] of childhood brain tumors. Medulloblastoma is a densely packed cellular, midline cerebella a tumour that grows over the roof of the fourth ventricle. It is the primary brain tumor in childrens[2]. The world health organization has classified it into four subtypes viz: classic, desmoplastic, nodular and large cells medulloblastoma. However, the classic pattern is mostly prevalent. Past work on medulloblastoma[3-5] contributes only towards the textual property of medulloblastoma. However, cell morphology conveys much more vital information in tissue samples. Microscopic view of the tissue level has scanty cytoplasm with sheets like an arrangement of tightly packed cells and irregular dark size nuclei. Also, not all brain tumors are malignant and do not invade in other body parts. Tumors are a group of cells that have variation in their properties due to various regions. They grow slowly, look normal and less harmful. To analyze the difference between normal and abnormal tissue samples of childhood brain tumor we have made a study on the histopathological images of normal and malignant tissue samples

1.1 Motivation

Childhood brain tumors need most attention as the rate of survival of children can be increased. This is primary Central nervous system tumors and thus are very significant and can have 100% mortality. Clinical diagnosis does not convey a strong confounding factor as the patients are mostly infants that cannot express the severity. The diagnosis often varies from pathologist to pathologist due to experience and subjective review. Therefore an accurate classification is important as to have a proper prognosis while not ending in over or under treatment.

1.2 Contribution:

The paper contributes towards the following:

a) Generation of Database: A database of childhood medulloblastoma histological slide images were created with images using oil emersion in 100x magnification for cell classification. The individual cells were segmented based on ground truth details and 1272 cell level image database was created. The biopsy tissue blocks of childhood medulloblastoma were collected from the neurological department of Guwahati Medical College and Hospital. It is the foremost Public Sector Super-specialty Hospital in the region catering to the general public and treating all diseases, and the slides were prepared at Ayurshundra private healthcare Ltd. Ayursundra Healthcare Pvt. Ltd is one of the biggest private healthcare centers in the region with an active pathology unit. The ground truth was marked by a certified pathologist at Guwahati neurological research center. We were successful in creating 1272 image database of normal and abnormal cells and also 50 images of 100x resolution for normal and abnormal samples. Since it is challenging to get normal brain samples we have considered those cells that were part of the abnormal sample but contains a normal region with pathologist advice. The details are given in **Table 1**.

Table1: Database Creation

Database Generated		
Type	Abnormal	Normal
Magnification	100x	100x
Resolution	800X600	800X600
No. of Images	44	6

b) Cell morphology analysis: The tissue level cell morphology was studied based on individual cells area, perimeter, irregularity, etc. For both normal and abnormal cell samples of ground truth images. We have used ground truth data for the study of the properties of the cells since no segmentation methodology can be 100%

accurate and using experienced knowledge of pathologist for marking of the ground truth we can have a more precise study of the cell level.

c) Classification: A binary classification of childhood medulloblastoma cells was made where the inclination of each cell towards normal or abnormal samples was achieved. The work till now on childhood medulloblastoma is done only at architectural level using texture features. We in this work attempted to have a cell level classification of the tissue samples.

d) Reproducibility: The work is open for more improvement and comparison in future reproducibility by researcher and experts using more sophisticated methodologies and high-level features. All the experiments were performed in Matlab 2016b.

II.MATERIALS AND METHODS

Ethical Statement: All experiments were performed by the guidelines of institutional ethics board (Registration number ECR/248/Indt/AS/2015 of Rule 122DD, Drugs and Cosmetics Rule, 1945 of India) of Institute of Advanced Study in Science and Technology, Guwahati. Patient consent was taken in the format of Guwahati Medical College, from where the samples were obtained, as part of their regulations.

Implementation: The experiments were implemented in Matlab® (R2016b, Mathworks, and Natick, MA, USA) and All-in-one HP pc (2.70 GHz, Intel Core i5, 4 GB Ram) and built-in Matlab function and modification of open source code.

2.1 Overview of the work:

We divided our study into five phases. Phase 1 is identification of cells by professional pathologist from whole slide image where they labeled the cells as normal and abnormal and marked them accordingly, Phase 2 is segmentation of ground truth cells from the images, Phase 3 is extraction of morphological features of the cells from the segmented images and Phase 4 is finally classification of the cells between normal and abnormal. Below in **Fig 1**, we give a diagrammatic representation of our work:

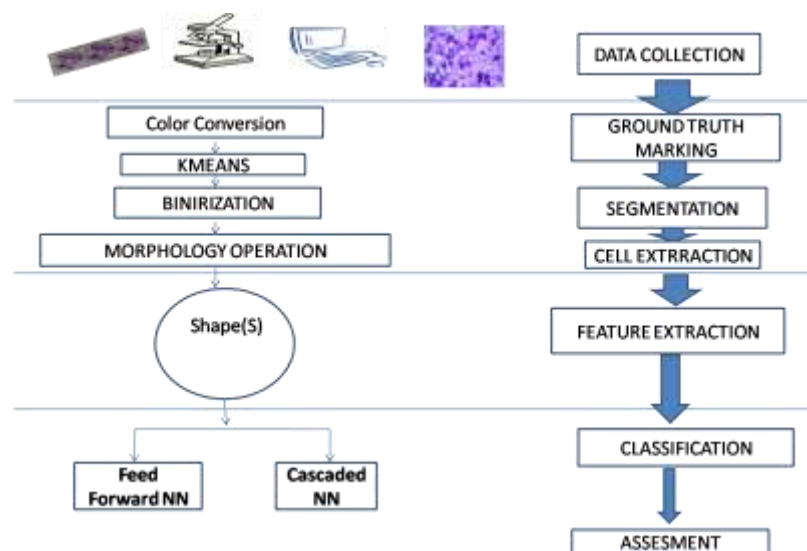


Fig1: Block diagram of the work

2.2 Database Generation:

The collection of tissue samples was from Guwahati Medical College and Hospital, GMCH. The slides were prepared at Ayurshundra Healthcare Pvt. Ltd by staining it in Hematoxylin and Eosin (H&E) solution which marks the nucleus as blue and cytoplasm as pink. After the preparation of slides, the slides were observed under a microscope at the pathology department of Guwahati Neurological Research Center. We collected 50 images at 100x magnification out of which 6 are normal and remaining abnormal samples. We extracted a total of 1272 cells out of which 275 were normal cells. So a total number of cells studied were 275 normal cell, and 275 abnormal cells were examined. The data collected for this purpose was in 100x magnification for detail study of the cellular structure of the tissue samples. However, pathologist prefers to observe the cell structure at 40x magnification.

2.3 Marking of the ground truth:

Our first step was to identify the malignant and normal cells in a histological image sample. So we extracted the ground truth from expert pathologist to determine the targeted malignant cells from the tissue samples. The medulloblastoma cells when observed under the microscope has irregular large size oval carrot like nuclei with a scanty cytoplasm. The individual cells were observed by experience pathologist at Pathology Department of Guwahati Neurological Research Center, and the ground truth was marked under the supervision of a pathologist. Simple Microsoft Paint was used to mark the ground truth in red. The overlapped and doubtful nucleus were excluded from our study and only well-stained nuclei were used to make the classifier. Since marking of the ground, the truth is a tedious and time-consuming task, so we initially extracted 1272 nucleus from 50 smear level images for our purpose. Out of which we have 63,425,377 and 122 of type desmoplastic, nodular, large and classic subtype. Remaining 275 were normal cells. **Fig2** shows the marked ground truth images at 100x

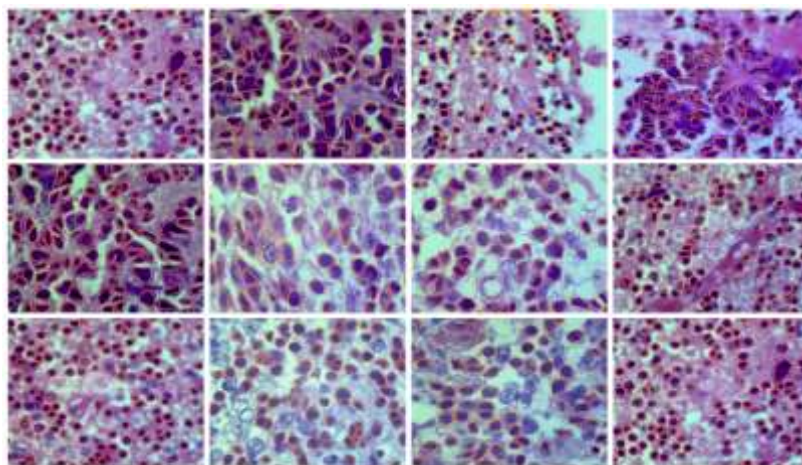


Fig2: Childhood medulloblastoma images at 100x magnifications

2.4 Segmentation of ground truth nuclei:

Our next task was segmentation of the ground truth cells from the background for our study. Kmeans color segmentation was used to segment out the ground truth marked nuclei. Kmeans color segmentation is a cluster based segmentation method used to segment out the different colors present in an image into different clusters of colors having lesser intra cluster variance than inter cluster variance. We run a few experiments with varying numbers of cluster to see the segmentation output and found that five clusters were appropriate for our purpose. This was because the histological tissue samples have many variations of colors from the cytoplasm to cells. After detecting the color border of the ground truth image, it was converted to a binary image with white border and black background. The region of the border was filled with pixels of intensity 255 followed by open morphology operation to get noiseless ground truth image. **Fig 3** describes the extraction of targeted cells.

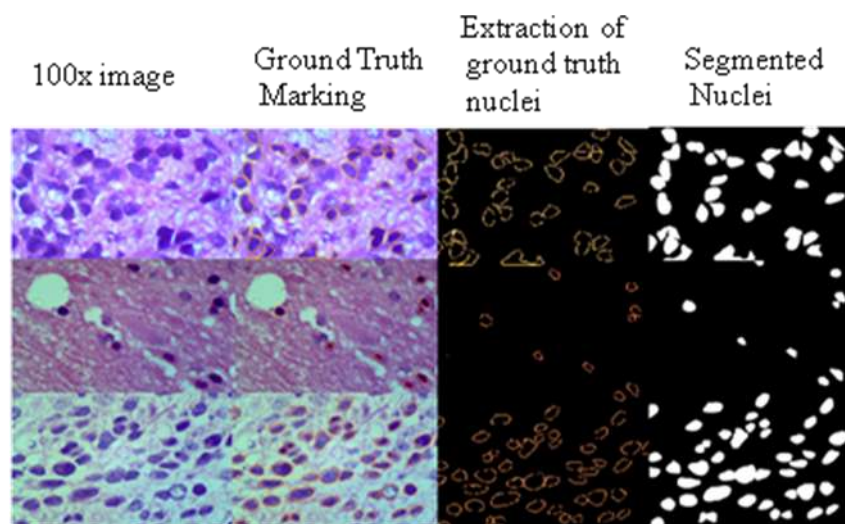


Fig3: Extraction of ground truth cells.

2.5 Feature Extraction

To quantify the visual information feature extraction was used. The morphological features of the cells that signify its shape, size irregularity was studied for both normal and abnormal cell samples. Area, EulerNumber, Orientation, Extent, Perimeter, FilledArea, Solidity, Eccentricity, major axis length, EquivDiameter, minor axis length were considered as morphological features of the cell image. These features were used to see the irregularity, enlargement, and distortion in shape of the cells.

The features are described below:

Area: Number of pixels in the segmented region.

Euler Number: Smallest rectangle that may contain the segmented region.

Orientation: Angle between x-axis and major axis of the ellipse that may be fitted in the segmented region.

Extent: ratio of total pixels in the segmented region to the pixels in the bounding box.

Perimeter: Distance around border by calculating the distance between each adjacent pair along the border.

FilledArea: Number of pixels in the filled image.

Solidity: Proportion of the pixels in the convex hull that is also in the segmented region.

Eccentricity: Degree of roundness.

MajorAxisLength: Length of the major axis of the ellipse of segmented image.

EquivDiameter: Diameter of a circle with the same area as in the region.

MinorAxisLength: Length of the minor axis of the ellipse of segmented image.

All total we had 11 feature vectors from a single cell for the study.

2.6 Classification

Next, we performed two classification experiments for binary classification of the cells. For one experiment we used the Feed Forward Neural Network, and for the second experiment, we used Cascaded Forward Neural Network. Cascade-forward networks include a connection from the input and every previous layer to following layers whereas a Feedforward networks consist of a series of layers where the network input is connected to the first layer and each subsequent layer has a connection from the previous layer[6]. The final layer produces the network's output[7]. Our cascaded neural network had two layers where five neurons were in the first layer and 10 in the second layer. For Feed forward neural network we used three layers with a number of neurons as 5,10 and 15 in respective layers. The number of neurons and layers were considered by hit and trial method. For this purpose, 5 fold cross-validation was performed where the total data was divided into five equal subsets and each.95% of the data at one time was used as the training and 5% as testing.

III.RESULTS AND DISCUSSIONS

We evaluated the classifier's accuracy by quantifying the proportion of correctly classified samples by total samples given by

$$\text{Accuracy} = \{(\text{True positive} + \text{True negative}) / \text{Number of samples}\} * 100$$

Where,

True Positive: No. of Cells identified to belong to the class which it belongs to.

True Negative: No. of Cells identified that do not belong to the classified class and do not belong to.

Number of Samples: Total no. of cell present from all class.

Table 2: FeedForward and Cascaded neural network performance

Classifier	Number of Hidden Layer	Number of neurons	Number of Normal cells	Number of Abnormal Cells	Accuracy
FeedForward Neural Network	3	[5 10 5]	275	275	85%
Cascaded Neural Network	2	[5 10]	275	275	86%

The accuracy of both classifications are almost similar, but we obtain a little higher accuracy in case of the cascaded neural network. In case of a number of layers cascaded neural network requires a lesser number of layers than feedforward neural network.

IV.COMPARISON

Author	Year	Work	segmentation	Feature	Classifier	Accuracy
Galaro et. al	2011	anaplastic and nonanaplastic medulloblastoma	N/A	Haar Wavelet and MR8 filter	KNN	80%
ai et. al	2011	anaplastic and nonanaplastic medulloblastoma	N/A	Haar, Haralick, and Laws textural features	random forest	91%
Cruz – Roa et. al	2015	anaplastic and nonanaplastic medulloblastoma	N/A	IBCa-CNN VGG-CNN	Softmax	89.8% and 76.6%

Table 3: Past work on medulloblastoma

It is seen that most of the work till date are texture based and is a binary classification between anaplastic and nonanaplastic samples[3-5]. However, no work is found that studies the cell level significance of childhood medulloblastoma tissue samples. Therefore an exact comparison was not possible. **Table 3** below states the work till present. However, the past work is described in the table below

V.CONCLUSION

With correct diagnosis, an improved 2- and 5-yr survival rate has been reported for childhood medulloblastoma[8]. We have studied the cell features between normal and abnormal samples. Using feedforward and cascaded neural network we obtained an accuracy between 85%-86%.Cell-based classification can also be used to differentiate between the various subtypes. As childhood medulloblastoma is highly sensitive in nature computer-aided methods can help in better evaluation.

Authors' contributions

Daisy Das had done the practical survey of the histological processing with practical implementation and analysis of the histological slides through image processing techniques.

Dr. Lipi B. Mahanta supervised the whole idea of the work and was a conceptual support team member.

Dr. Shabnam Ahmed was the medical expert in the study and analysis of our work.

Conflict of Interest

No conflict of interest.

VI.ACKNOWLEDGEMENT

We would like to convey our sincere thanks to Dr. Basanta Kr. Baishya, Head of the department of neurosurgery, Guwahati medical college, and Dr. Inamul Haque, MCH Trainee from the department of neurosurgery, Guwahati Medical College for providing us the tissue blocks and Dr. Anup Das from Ayurshundra Healthcare Pvt. Ltd. For processing the slides. We would further thank Dr. Shabnam Ahmed of Guwahati Neurological Research Centre, Sixmile for dedicating her time and effort and helping us in image acquisition and providing the ground truth. We are grateful to Institute of Advanced Study in Science and Technology (IASST), Guwahati for giving us the platform to perform our research.

REFERENCES

- [1] AP Polednak, JT Flannery, (1995) Brain, Other central nervous system, and eye cancer, Cancer. 75: 330-337.
- [2] T Furata, A Tabuchi, Y Adachi, et al., (1998), Brain Tumor Pathol. 15: 7.
<https://doi.org/10.1007/BF02482094>
- [3] J Galaro , Judkins AR, D Ellison, J Baccon , A Madabhushi , An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas, Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS 3443–3446. 2011;doi:10.1109/IEMBS.2011.6090931
- [4] AC Roa, J Arévalo , A Judkins, A Madabhushi , F González, A method for medulloblastoma tumor differentiation based on convolutional neural networks and transfer learning. Proceedings Volume 9681, 11thInternational Symposium on Medical Information Processing and Analysis; 968103.; 2015;doi: 10.1117/12.22088
- [5] Y Lai, S Viswanathan, J Baccon , D Ellison, AR Judkins, A Madabhushi , A Texture-based Classifier to Discriminate Anaplastic from Non-Anaplastic Medulloblastoma., IEEE 37th Annu. Northeast Bioeng.

Conf. 2011;doi:10.1109/NEBC.2011.5778641

[6] <http://in.mathworks.com/help/nnet/ref/cascadeforwardnet.html>, Matlab Help Page, accessed 12 Oct 2016

[7] <http://in.mathworks.com/help/nnet/ref/feedforwardnet.html>, Matlab Help Page, accessed 18 Oct 2016

[8] FG Davis, S Freels, J Grutsch, S Barlas, S Brem, (1998), Survival rates in patients with primary malignant brain tumors stratified by patient age and tumor histologic type. An analysis based on surveillance, epidemiology and results (SEER) data, 1973-1991. J Neurosurg 88:1-10.