BRAIN TUMOUR CLASSIFICATION USING MACHINE LEARNING THRESHOLDING

PERIYAKARUPPAN K[1], KARTHIK S[2], KARTHIKEYAN N[3]

Abstract

In recent era, it is of the utmost importance to perform early identification and diagnosis by means of MRI due to the high mortality rate associated with brain tumours. Because of the intricate nature of the brain structure and the interconnection of its many different parts, it is famously difficult to treat a brain tumour in an effective manner. The accurate and efficient division of a brain tumour is still a difficult challenge to tackle, despite the numerous various technologies that are currently available for doing so. Despite these advancements, the problem remains difficult. Dealing with the many different kinds of tumours presents its own specific challenges when seeking to divide and classify them. Due to the complexity of the situation, using only a single imaging modality may make it challenging to undertake full segmentation and classification of a brain tumour. In this paper, we model a threshold-based classification based on the selection of certain features set. The classification algorithm used in the study is random forest. The model is trained and tested in python tool. The results of simulation show an increased accuracy than other methods. sets of EEG and the results show that the proposed method has higher range of classification accuracy than the other methods.

KeyWords: brain tumour, thresholding, classification.

DOI Number: 10.14704/nq.2022.20.8.NQ44705 **NeuroQuantology 2022; 20(8): 6810-6817**

[1] Professor, Department of Computer Science & Engineering, SNS College of Engineering, Coimbatore-641107, Tamilnadu, India, kperiyakaruppan@gmail.com
[2] Professor and Dean, Department of Computer Science & Engineering, SNS College of Technology, Coimbatore-641035, Tamilnadu, India, profskarthik@gmail.com
[3] Professor, Department of Computer Science & Engineering, SNS College of Technology, Coimbatore-641035, Tamilnadu, India, profkarthikeyann@gmail.com

Introduction

The central nervous system is the part of the body that is in charge of relaying sensory information to other parts of the body and setting off the appropriate responses. Both the brain and the spinal cord work together in order to accomplish this spreading of the message. The cerebral cortex, the brain stem, and the cerebellum are said to be the three basic areas that make up the human brain, according to the study of brain anatomy. The average human brain weighs roughly in the range of 1.2-1.4 kg and has a volume that ranges from 1130 cm3 for female brains to 1260 cm3 for male brains [1].

The volume of the brain also varies depending on gender. The prefrontal cortex plays a role in executive functions such as decision, the resolution of problems, and the regulation of movement. The parietal lobe is in charge of determining where the body is positioned relative to its environment. Memory and hearing are controlled by the brain temporal lobe, whereas the occipital lobe is responsible for the processes of the brain that are related to visual processing. The grey matter that makes up the cerebral cortex, also known as cortical neurons, can be found in the most superficial layer of the brain, which is referred to as the cerebral cortex [6].

In comparison to the size of the cerebrum, the cerebellum is quite a little smaller. A specific part of the nervous system is in charge of motor control, which is also known as the coordinated management of an organism voluntary actions. This function is also known as coordinative management. The motor cortex is a part of the nervous system that goes by a number of different names. None of the diagnostic tools are able to locate the small lesion region. This is because the stroke territories range in size and severity to varied degrees. When compared to those of other animals, the cerebellum of a human being has a more complex structure and is completely developed [7]. The anterior lobe, the posterior lobe, and the flocculonodular lobe are the components that make up the cerebellum. The structure known as the vermis has the shape of a sphere and is responsible for connecting the two lobes. The greyish cortex of the cerebellum is slightly thinner than the cortex of the cerebrum, and it encircles the core of the cerebellum that is composed of white matter (WM). The distinct sections of the front and back of the brain, sometimes known as lobes, collaborate with one another to allow you to do increasingly difficult motor activities.

The flocculonodular lobe is the one in charge of ensuring that homeostasis is preserved. The brain stem is a structure that extends for approximately 7 to 10 centimetres in length and has the appearance of a stem. The area can be found at the exact middle of the brain. It is because it contains nerve bundles that originate from both the brain and the periphery of the body that it assists with things like breathing, maintaining balance, and directing the eyes. Other functions that it assists with include regulating the heart rate and blood pressure. After exiting the cerebral cortex, the nerve fibres of the thalamus go through the brain stem on their way to the spinal cord [8]. The thalamus is located at the base of the brain. After then, they proceed to other areas of the body and start their invasions there. After that, they leave. The medulla, the pons, and the midbrain together make up the vast majority of what is known as the brain stem. The midbrain is an area of the brain that helps support a broad variety of other brain functions, such as the processing of information that is motor, auditory, and visual, as well as the movements of the eye. While the medulla oblongata is responsible for regulating blood flow and performing other processes including swallowing and sneezing, the pons is responsible for helping with respiration, intra-brain communication, and costlings [9].

The MRI sequences that are used for the examination of stroke lesions can be categorised according to age, location, and extent of the damage caused by the stroke. An electronic devices might be beneficial in the context of medical treatment for making an accurate diagnosis of the rate at which the disease is advancing. The stroke lesions need to be isolated in order to properly evaluate the entire infected region of the brain and determine which treatment choices are appropriate. Cognitive neuroscientists, who investigate the connection between brain damage and mental disability, have discovered that stroke injuries require segmentation in order to be properly treated. Because of the way in which stroke-related lesions vary in appearance over the course of time, diagnosing and classifying them can be famously challenging. Diffusion weighted imaging (DWI) and fluid attenuated inversion recovery are two examples of the types of MRI sequences that can be utilised to detect stroke damage. This hyperintensity is able to be detected most clearly in DWI sequences obtained from patients who are in the acute stage of stroke, which is the stage during which the primary concern is infection. The degree of perfusion that may be estimated from the mapping is demonstrated by the underperfusion zone.

One may consider the area that is between the two sections to be the eclipse penumbra. Stroke can generate lesions that can appear in many various contexts and take many different forms. These lesions can also express themselves in a broad variety of circumstances. It is possible for several lesions of varying sizes and shapes to form at the same time, and these lesions will not necessarily adhere to any established vascular patterns. This can happen at any moment over the course of the disease. Strokes can generate lesions that are capable of spreading throughout an entire hemisphere and have radii that are in the millimetre range. Because of the unusual shape of the hemisphere, the severity of the disease may differ quite a little from one portion of the affected area to another of the same area. When viewed with the naked eye, white matter hyperintensities and chronic stroke lesions appear to be quite similar to one another. This makes it difficult to do automated stroke segmentation [10]. Also, the recent improvements in the computation [11] [12] field and network quality [13] [14] propomotes this developments possible.

In this research paper, we develop a deep learning segmentation model to find the imaging modalities. In order to reduce the problem of low contrast nature, the study creates a technique based on linear contrast enhancement and optimised it using histogram equalization. The goal of sparse Convolutional Neural Network (CNN) models was to provide a more informative feature vector.

Background

Harati et al. [2] were able to demonstrate that the tumour region could be automatically segmented by selecting seed points by utilising an enhanced fuzzy connectivity (FC) technique. This was accomplished by choosing seed points. Because of this, they were able to demonstrate that the cancer region could be segregated automatically.

Saha et al. [3] developed a strategy that is based on the Bhattacharya coefficient in order to expeditiously discover the border box that covers a tumour or edoema in a short amount of time. Their method of clustering starts with an axial view of a brain image, divides it in half, and then uses a rectangle to compare the left and right halve corresponding regions in order to determine which region within the rectangle is the most distinguishable from the others. This method begins with an axial view of a brain image and ends with an axial view of a brain image.

Within the framework of their semiautomatic method of segmentation, Zhu et al. [4] recommend the application of the ITK-Snap tool for the initial segmentation of the brain tumour. The software approach includes both voxel-based segmentation and deformable-shape-based segmentation as options for the user to choose from.

Sachdeva et al. [5] utilised texture information with intensity in an active contour model (ACM) in order to get around the difficulty that was presented by free vector flow (FVF), boundary vector flow (BVF), and gradient vector flow (GVF). Over-segmentation occurs when inadequate edges are chosen, which is caused by the edoema that surrounds the tumour. This results in an excessive amount of tissue being removed. The edoema is responsible for both the incorrect selection of edges and seeds, which leads to the preconvergence problem. Both of these problems are caused by the edoema.

When discussing neurological malignancies, the terms slow-growing and fast-growing can be used interchangeably to describe the disease progression. Tumors can be categorised as either benign (slow-growing) or malignant (aggressive), depending on the rate at which they grow, depending on whether or not they have migrated from the location where they first appeared to other areas. This is determined by whether or not they are aggressive. The World Health Organization gives different types of brain tumours a grade that falls anywhere between I and IV on a scale. In the following few lines, we will discuss the varied degrees of severity that are connected with brain tumours.

- Grade I: Cancers that are categorised according to grade I condition does not rapidly spread. These are connected with a greater chance of survival over the long run, and they can be removed almost entirely through surgical treatments. This particular tumour is referred to as grade 1 pilocytic astrocytoma, and it matches the description quite well.
- Grade II: Although cancers of grade II advance slowly, there is a greater chance that the disease will metastasize, or spread, to other organs as the cancer grows. After a tumour has been surgically removed, there is always a chance that it will return and spread to other parts of the body. This class of tumours includes oligodendroglioma, which is a specific subtype of the disease.
- Grade III: Cancers of grade III progress more rapidly than cancers of grade II and have the potential to spread to tissues that are located in close proximity to the original tumour. As a result of the aggressive nature of these tumours, surgery alone is generally insufficient as a treatment option; postoperative radiation therapy or chemotherapy is typically necessary instead. The subtype of brain cancer known as anaplastic astrocytoma is considered to be one of the deadliest kinds of the disease.
- Grade IV: As a result of their potential to spread to other parts of the body, grade IV tumours are considered to be among the most dangerous kinds. Making use of the blood vascular network could make the process of rapidly expanding go more smoothly. One disease that falls under this category is known as glioblastoma multiforme.
- Ischemic stroke: Stroke caused by ischemia is a serious brain disease that is a leading cause of death and disability around the world. Underperfusion, also known as tissue hypoxia, and the death of tissue can occur within hours after the blood supply to the brain has been cut off. Lesions that are classified as acute occur within the first twenty-four hours after the onset of stroke symptoms; lesions that are classified as subacute occur within the first 24 hours to two weeks after the onset of stroke symptoms; and lesions that are classified as chronic occur after two weeks or more.

Proposed Method

In this paper, we adopt a threshold-based classification model to improve the process of classification of brain tumors.

Figure 1: Proposed Method

Preprocessing

The procedure of extracting the necessary region begins with preprocessing, which is a very critical step. The non-brain tissue is eliminated through the application of an algorithm for brain extraction in two dimensions. As a result of flaws in the radio frequency coil, the bias field presents a significant challenge for magnetic resonance imaging (MRI). Inaccuracies of this nature are referred to as intensity inhomogeneity. Because of the presence of noise and other irregularities in the image, it may be challenging to distinguish between healthy and diseased tissue. As a result, performing direct image analysis may be difficult. As a result of this, segmentation is now handled entirely by software rather than by people. The term for this type of procedure is an automated process.

Thresholding methods

Even though it may be challenging to choose the optimal threshold in low-contrast images, the thresholding method continues to be an essential and effective tool for differentiating target elements. In order to determine threshold values, the histogram of an image intensity must first be analysed. The findings of this analysis are then applied. The two basic varieties of this strategy are referred to as the local thresholding approaches and the global thresholding approaches. The global thresholding approach is the most effective way to segment an image when there is a significant and uniform contrast or intensity difference between the objects and the background.

When trying to determine the value of the threshold that is most appropriate for a specific circumstance, the statistical method known as Gaussian distribution might be of assistance. When a single threshold value does not produce the type of segmentation results that are required or when the threshold value cannot be measured from the entire histogram, multiple methods are utilised. These methods are employed in situations where a single threshold value does not produce the sort of segmentation results that are required. The thresholding method is utilised the most frequently in the first stage of the segmentation process, and many individual sections are recovered from the grayscale photos.

In order for HOG features to function, a image must first be segmented into smaller sections and then those segments must be reassembled into the original image. We begin the procedure by taking the image that has been segmented and applying a Sobel kernel function filter to it. This is the first step. Because of this, we are able to acquire a gradient in both the Gx and the Gy directions. As can be seen in Equations (1) and (2), respectively, both the angle and the amplitude of the gradient are computed for each and every one of the individual pixels in the image.

$$f_{seg|G(i,j)|} = \sqrt{G_x(i,j)^2 + G_y(i,j)^2}$$

$$f_{seg\theta_G} = \tan^{-1}\left(\frac{G_y(i,j)}{G_x(i,j)}\right)$$

where

|G| - gradient magnitude,

θG - gradient angle,

i - columns and

j - rows.

The votes from each individual cell are distributed among a number of different categories using a gradient that is based on the angle. In the end, a normalisation vector is created by employing each histogram block as a building component to make the final product. As shown in the following equation (3), the HOG feature descriptor takes as its inputs the eight bin cells that make up the segmented image. This equation gives the HOG feature description that you were looking for. 6814

$$f_{segV^{N_i}} = \frac{V_i}{\sqrt{V_2^2 + \in^2}}$$

∈ refers to a minor constant that is larger than zero and does not divide evenly by zero. All of the histograms that are contained within a block are displayed by the non-normalized vector, which is denoted by V. The HOG feature vector is obtained by multiplying the normalisation vectors of each block together to generate the final product. This produces the final product. We also compute the average, as well as the standard deviation and the variance, for each individual characteristic.

Classification of MR images

An example of an ensemble method is the Random Forest algorithm, which generates a classifier by combining several distinct decision tree (DT) variants into a single model. Random Forest is an example of an ensemble approach. An ensemble approach is represented here by this particular method. DT results in the production of a structure that is analogous to a tree. This structure nodes and edges have been labelled, and values have been assigned to the attributes of the nodes and edges. The combination of two hundred DT classifiers with the technology of flexible boosting (Ada-Boost) results in the creation of an ensemble. Each of these classifiers has an unsatisfactory track record of performance. The number of incorrect predictions is reduced to almost nothing when the results of multiple tree classifiers are combined into a single algorithm.

An improvement on the original Random Forest algorithm, the iterative Random Forest algorithm creates successive forests of weighted decision trees by iterating over the algorithm previous steps. This approach is used for tasks involving regression and classification that involve the importance of variables and the variable-interaction space. These tasks entail determining which variables are significant and which variables interact with one another. The information about the patterns that are contained in these woods is mined for information. The ability of this strategy to generate a very large number of decision trees is where the majority of its power lies, which is why it is considered to be its greatest strength.

You are able to do many computations on trees simultaneously provided that each tree is assigned to a distinct core of the central processing unit (CPU). If, on the other hand, more trees are required than can be constructed in a single iteration of the process, this will slow down the process of developing the forest as a whole. The process of constructing a forest can be massively parallelized if the process of growing trees is distributed across numerous nodes in a high-performance computing cluster. This makes the development of the forest much faster. This makes it possible for the process to take advantage of a large number of cores, which, in turn, speeds up the prediction and analysis of decision patterns. Because of these developments, it will now be possible to conduct research in fields that call for the collection of enormous amounts of data in real time and that have the potential for massive combinatorial impacts. Finding the genetic links in the complete genomes of biofuel producers that are responsible for higher production could be one approach to achieving this goal.

Results and Discussions

In order to evaluate the many different methodologies that have been suggested, the researchers make use of a wide variety of datasets that are available to the general public. In this section of the article, we are going to look into several different data sets that are both significant and difficult. When compared to any other MRI database, the BRATS datasets constitute the single largest challenge. The BRATS Challenge is conducted on an annual basis, and the resolution of the competitions that are held in subsequent years will be increased by 1 mm3 voxel in each of those years.

In one of the experiments, performance was evaluated by using the fivefold cross-validation methods for the detection of tumours on training samples that were selected at random. This was done within the context of one of the experiments. In order to check that the data have not been overfit, the method of cross-validation is utilised. The portion of the fivefold process that deals with evaluation makes use of only one set of data, while the other four sets are put to use in the process of instruction and learning, respectively. This process will continue to be repeated several times over the course of a chain of cycles. The entire collection of data is first split in half to create two separate sets. The first half is utilised for the purposes of evaluation, while the second half is put to productive use in the practise setting. In the second experiment, ground-truth annotations were utilised to assist in the generation of pixel-by-pixel findings.

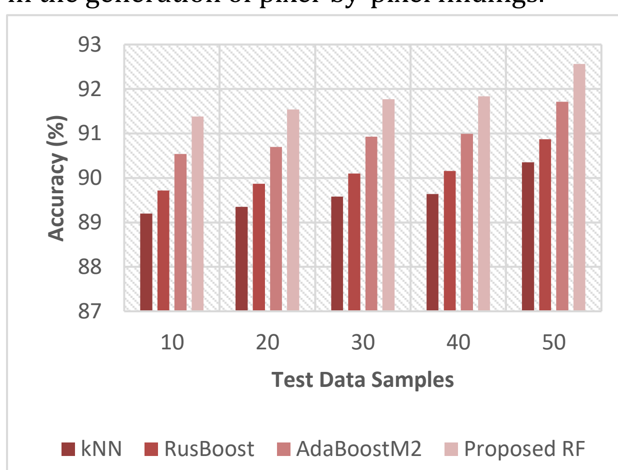

Figure 2: Accuracy

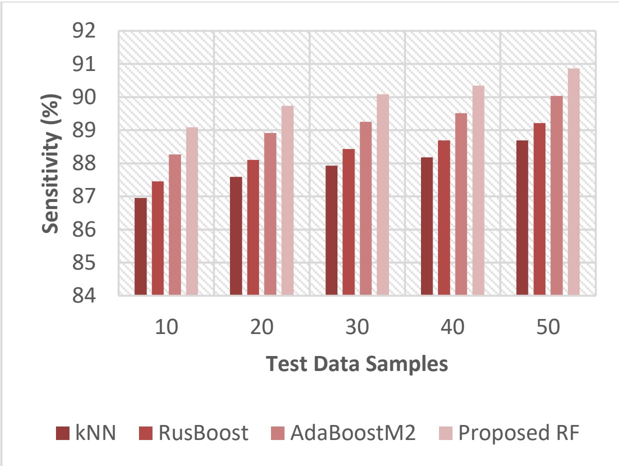

Figure 3: Sensitivity

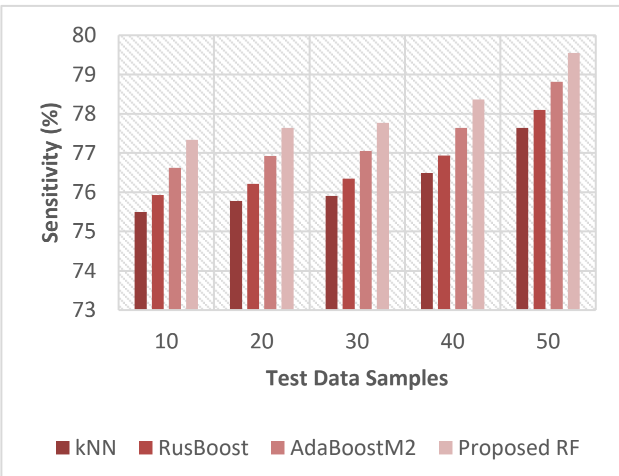

Figure 4: Specificity

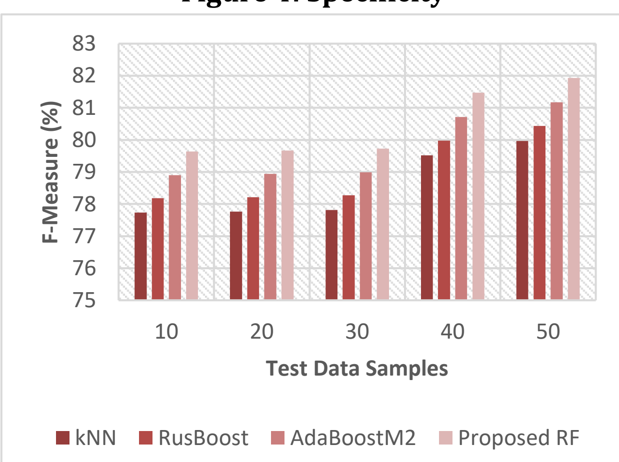

Figure 5: F-Measure

The first set of experiments that we carried out involved us conducting a leave-one-out cross-validation analysis on a portion of the BraTS data. This allowed us to determine the various experimental options that were available to us (four actual HG patients). We did this in the hope that applying the same decisions to the entire BraTS dataset would yield results that are comparable to those produced by the previous steps. We started by running our tests on a relatively modest portion of the overall data in order to cut down on the amount of time and effort that would ultimately be required in the long term.

We have obtained the features through a series of tests using four different classifiers: kNN, RusBoost and AdaBoostM2 and came to the conclusion that RF performed the best when applied to the classification of brain tumours. This led us to the conclusion that RF was the best classifier to use. When it comes to the actual dataset, both HG and LG undergo a singular iteration of leave-one-out cross-validation in a fashion that is distinct from the other. In addition to this, we carried out a comprehensive classification, which consisted of breaking the tumour down into three distinct parts: the entire tumour, the core tumour, and the augmenting tumour. The outcomes of utilising the proposed method are higher than the outcomes of utilising other methods that are of a comparable nature, or even improve upon them as in Figure 2-5.

Conclusions

In this study, we provide a hierarchical classification strategy for the tumour by splitting it into three unique components: the overall tumour, the core tumour, and the augmenting tumour. Multi-modality MRI research involves retrieving and utilising characteristics such as intensity, intensity difference, neighbourhood information, and wavelet features collected with a variety of classifiers. According to the quantitative results of our proposed method, combining wavelet-based texture characteristics with an RF classifier has shown that it is possible to get classification accuracy that is on par with or even better than what is currently considered to be the state of the art. This was demonstrated by the fact that our method was able to achieve this result.

References

[1] Kang, J., Ullah, Z., & Gwak, J. (2021). Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors, 21(6), 2222.

[2] Harati, V., Khayati, R., & Farzan, A. (2011). Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images. Computers in biology and medicine, 41(7), 483-492.

[3] Saha, B. N., Ray, N., Greiner, R., Murtha, A., & Zhang, H. (2012). Quick detection of brain tumors and edemas: A bounding box method using symmetry. Computerized medical imaging and graphics, 36(2), 95-107.

[4] Zhu, Y., Young, G. S., Xue, Z., Huang, R. Y., You, H., Setayesh, K., ... & Wong, S. T. (2012). Semi-automatic segmentation software for quantitative clinical brain glioblastoma evaluation. Academic radiology, 19(8), 977-985.

[5] Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., & Ahuja, C. K. (2012). A novel content-based active contour model for brain tumor segmentation. Magnetic resonance imaging, 30(5), 694-715.

[6] Díaz-Pernas, F. J., Martínez-Zarzuela, M., Antón-Rodríguez, M., & González-Ortega, D. (2021, February). A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. In Healthcare (Vol. 9, No. 2, p. 153). MDPI.

[7] Ayadi, W., Elhamzi, W., Charfi, I., & Atri, M. (2021). Deep CNN for brain tumor classification. Neural Processing Letters, 53(1), 671-700.

[8] Sharif, M. I., Khan, M. A., Alhussein, M., Aurangzeb, K., & Raza, M. (2021). A decision support system for multimodal brain tumor classification using deep learning. Complex & Intelligent Systems, 1-14.

[9] Kumar, R. L., Kakarla, J., Isunuri, B. V., & Singh, M. (2021). Multi-class brain tumor classification using residual network and global average pooling. Multimedia Tools and Applications, 80(9), 13429-13438.

[10] Abd El Kader, I., Xu, G., Shuai, Z., Saminu, S., Javaid, I., & Salim Ahmad, I. (2021). Differential deep convolutional neural network model for brain tumor classification. Brain Sciences, 11(3), 352.

[11] Dr. Grandhi Suresh Kumar, Dr. Pratik Gite, Dr. M. Prasad, Dr. Madiajagan M, Dr. S. Selvakanmani, Dr. V. Saravanan, Design of Autonomous Production Using Deep Neural Network for Complex Job Shops, International Journal of Grid and Distributed Computing, Vol. 14, No. 1, pp. 813-824, 2021.

[12] Shakir Khan, V. Saravanan, et., al., "Privacy Protection of Healthcare Data over Social Networks Using Machine Learning Algorithms", Computational Intelligence and Neuroscience, vol. 2022, 8 pages, 2022.

[13] Sumathi A, Saravanan V, “Bandwidth based vertical handoff for tightly coupled WiMAX/WLAN overlay networks”, Journal of Scientific & Industrial Research, vol. 74, pp. 560-566, 2015.

[14] Saravanan V & Sumathi A, “Dynamic handoff decision based on current traffic level and neighbor information in wireless data networks”, Fourth International Conference on Advanced Computing (ICoAC), IEEE, pp. 1-5, 2012.