BRAIN TUMOR CLASSIFICATION USING MRI IMAGES WITH K-NEAREST NEIGHBOR METHOD

Rafi Haidar Ramdlon
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
rafihaidarramdlon@gmail.com

Entin Martiana Kusumaningtyas
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
entin@pens.ac.id

Tita Karlita
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
tita@pens.ac.id

***Abstract*—The accuracy level in diagnosing tumor type through MRI results is required to establish appropriate medical treatment. MRI results can be computationally examined using K-Nearest Neighbor method, a basic science application and classification technique of image processing. Tumor classification system is designed to detect tumor and edema in T1 and T2 images sequences, as well as to label and classify tumor type. Data interpretation of such system derives from Axial section of MRI results only, which is classified into three classes: Astrocytoma, Glioblastoma, and Oligodendroglioma. To detect tumor area, basic image processing technique is employed, comprising of image enhancement, image binarization, morphological image, and watershed. Tumor classification is applied after segmentation process of Shape Extration Feature is undertaken. The results of tumor classification obtained was 89.5 percent, which is able to provide information regarding tumor detection more clearly and specifically.**

Keywords— Brain Tumor, Image Processing, K-Nearest Neighbor, Magnetic Resonance Imaging

I. INTRODUCTION

Tumors are a commonly used term, but not specific, for neoplasms. The word tumor only refers to the mass. Tumors are general terms that can refer to benign (generally harmless) or malignant (cancer) growth [1]. Cancer is a term used for cell growth (tumors) that are malignant or continue to grow uncontrollably which can threaten the health of the body.

Cancer is the second death cause in the world with 9.6 million cases in 2018. Globally, around 1 in 6 deaths are caused by cancer [2]. These numbers has increased compared to the number of deaths caused by cancer in 2016 amounting to 8.9 million people.

The initial cause of the appearance of cancer cells in each person might be different, some come from the genetic factors of each individual and some come from external environmental factors. The main cancer risk factors throughout the world are tobacco use, alcohol use, unhealthy eating patterns, and the lack of intensity of physical activity. Between 30-50% of cancers can be prevented by avoiding risk factors and applying detection as early as possible. Detection as early as possible can reduce the burden of cancer and increasing chance of healing. Early detection can be done by conducting a medical examination. Medical technology that applied in cancer detection includes X-ray mammography, Digital mammography, Ultrasonography, Magnetic resonance imaging (MRI), Positron Emission Tomography (PET), CT scans, etc. In this study the dataset used in the form of images from cancer detection using Magnetic Resonance Imaging (MRI) technology.

Magnetic resonance imaging (MRI) is one of the most advanced imaging techniques to date. MRI is used to produce images of body parts in detail by applying the imaging plane. Compared to medical X-ray-based diagnostic techniques, MRI does not use ionizing radiation but uses radio frequency [3].

Clinical diagnosis images must demonstrate the contrast between normal anatomy and anatomy accompanied by pathology. If there is no difference, then abnormalities in the body's tissue structure cannot be detected. One of the advantages of MRI is that when compared with other depictions, the appearance of soft tissue is very good. The contrast characteristics of each image depend on the number of variables, and it is important to know each mechanism that can affect contrast in the MRI results [4]. The process of reading MRI results, especially in parts of the head and brain organs in most hospitals in Indonesia, uses a manual method by comparing all the results of MRI images obtained from Axial, Sagital, and Coronal images. The accuracy of the translation results affects the subsequent medical treatment according to the type of stage of the cancer. The process of translating computationally generated MRI images can assist radiologists in increasing accuracy to classify the type of cancer suffered based on its stage. This study classifies MRI images of brain tumors into 3 classes so that the medical treatment that will be carried out is further targeted and as needed. Computational cancer detection has been carried out many studies. However, the detection of cancer is automatically only detected by one type of image sequence, T1 contrast, which only shows tumor detection. Whereas the manual process is carried out from all types of sequences resulting from the MRI process. Each sequence displays a different type. On the results of the MRI sequence T2 FSE can be a brain image that shows the extent of edema from the tumor. Edema itself is a brain organ that swells from contracting a tumor. This area of edema is also used as a radiologist to determine the type of class of the tumor.

Fast Spin Echo (FSE) uses a flip angle of 90º with some pulse reshaping of 180º to produce several SEs in a given TR. Each echo is a pulse encoded with different amplitude,

978-1-7281-4449-8/19/$31.00 ©2019 IEEE

so data can be collected on each echo and stored in a different line k space. FSE can be used to produce one to two echoes in one SE. The echo circuit can be exchanged so that the data collected from some echo trains can get a second echo. This method is generally used to produce PD and T2 images which illustrate similar measurements in SE. On the other hand, T2 images can also be obtained without PD images [5].

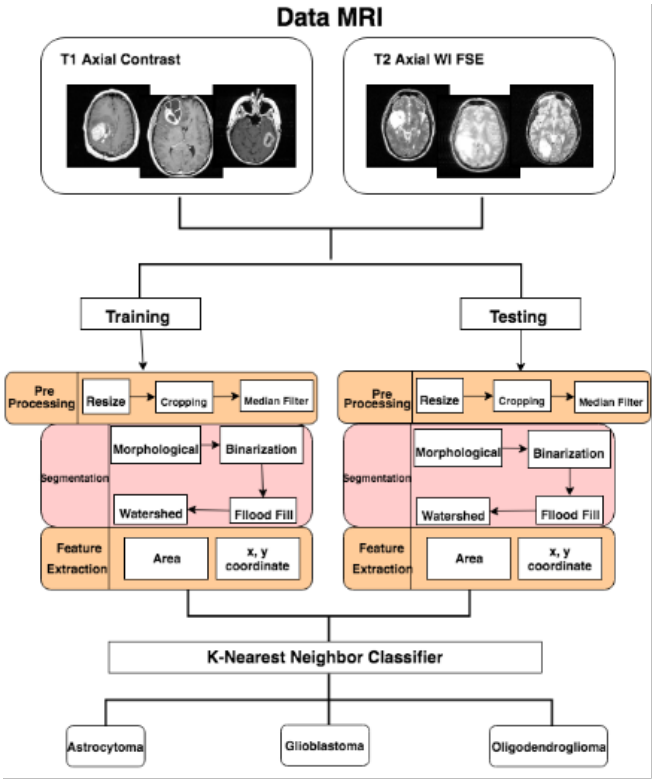


Fig. 1. System Diagram

Some computational research has been done in reading MRI results. Balakumar (2015) [6] performed computational readings on MRI results by using several features during feature extraction including shape features, intensity features, and texture features. In the research of Dilip Kumar Gandhi (2012) [7] feature extraction used uses texture features which are then calculated using the GLCM (Gray Level Co-Occurance) method. Mohammad Havaei (2014) [8] conducted research in computing the readings of MRI results of brain tumors using the classification method, namely the K-Nearest Neighbor method and using a simple vector feature in its feature extraction process. Research Narmada M. Balasooriya (2017) [9] describes the classification of brain tumors into five classes using the MRI dataset with the proposed Convolutional Neural Network method.

This study uses the K-Nearest Neighbor method to classify brain tumor data into 3 classes, namely Astrocytoma, Glioblastoma, and Oligodendroglioma. The classification results aim to increase accuracy in the process of translating magnetic resonance imaging (MRI) images and can be used as reference references using other classification methods. This study uses a Segmentation Technique in the form of Morphological Gradient and Watershed. In the feature extraction process, forms are used in the form of area and centroid so that the two features that make a difference with other studies.

In this paper the problems are explained in several parts. In the second part the completion methods and techniques are used to detect tumors. The results and analysis are explained in the third section and discussion of conclusions is explained in the last section.

II. METHODS

This study was conducted to classify the data obtained from TCIA (The Cancer Imaging Archive) [10], the data taken for this study were images of magnetic resonance imaging (MRI) patients with brain tumors with types of astrocytoma, glioblastoma, and oligodendroglioma. Fig.1 shows a system diagram of a system of classifying data from TCIA (The Cancer Imaging Archive) starting from preprocessing data to the process of determining the class of the inputted image.

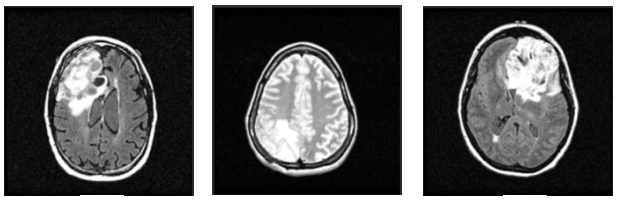
A. Datasets

The data in this study were obtained from TCIA, data containing MRI images of patients with brain tumors along with CSV files. The 3 types of brain tumors contained in this data are: astrocytoma, glioblastoma, and oligodendroglioma. The number of datasets used as training data is 254 in all classes as described in table 1.

TABLE I. LIST OF DATASETS

No	Dataset	Data Source	Amount of Data
1	*Astrocytoma*	TCIA	84
2	Glioblastoma	TCIA	124
3	*Oligodendroglioma*	TCIA	46

Fig. 2 displays an example of the dataset used in this study.



a b c

Fig. 2. Example image of a brain tumor (a)Astrocytoma, (b)Glioblastoma, (c)Oligodendroglioma

This study used MRI data with two different sequences. The first dataset is T1 axial contrast and the second dataset image is T2 axial WI FSE. Draw a dataset using the dcm / Dicom format and must be converted to png format.

B. Pre-Processing

Preprocessing is an initial process that is carried out to improve image quality, equalizing image size which is one of the variables that will affect the results because the extraction feature used is a form feature and takes brain objects that are used and discards unnecessary background. *The stages in* preprocessing in this study are Median Filtering, Resizing Image, and Cropping Image.

1) Resizing Image

This step would be will be changed the image pixel value height and width to 512 x 512. This process works so that the segmentation process can be done because the size of the image that is too small causes no objects that can be detected during the segmentation process. The process of

resizing images also functions as a dependent variable because in this study feature extraction used is a form feature so that the image size is very influential with the results obtained.

2) Median Filtering

This stage eliminates noise to improve image quality in the image. The output pixel value is determined by the median of the specified mask environment. Median is searched by sorting the pixel value of the specified mask. Then look for the middle value.

3) Cropping Image

At this stage the image cutting process is carried out by detecting brain objects used in the next process and removing background objects that are not needed for the next process. This process serves to minimize the possibility of factors that can affect the results of features obtained such as the error of the process of detecting tumor objects during the segmentation process.

C. Segmentation Processing

This process is the process of detecting the location of the tumor in the image. This is useful for the next process because the results of the segmentation obtained are used for material from the feature extraction used. Stages in segmentation processing in this study are Morphological Operation, Binarization, Floodfill, Dilatation, Erotion, and Watershed.

1) Morphological Transformation

.At this stage the image that has gone through the pre-processing stage is done by edge detection to find the object that is searched for and separating the object and background. Edge detection is done to process T2 images. In picture T1, color detection is used in the image using the Morphological Open.

The important thing to detect objects is the kernel value must be obtained first. The kernel value is obtained through the functions provided by open cv, Morphological Transformations. In this library, the kernel is provided in the form of a rectangle, ellipse, and cross. In this study the kernel used was the ellipse kernel. The kernel is a matrix used for blurring, sharpening, embossing, and edge detection. The kernel is obtained from the initiation of the large dimensional array needed. Here is an example of an array of kernel types:

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

(a) Rectangle

0	0	1	0	0
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
0	0	1	0	0

(b) Ellipse

0	0	1	0	0
0	0	1	0	0
1	1	1	1	1
0	0	1	0	0
0	0	1	0	0

(c) Cross

2) Binarization

The binarization stage is done to change the pixel in the image into a binary number using the thresholding method. At this stage, the detection of binary pixels is to be thresholded. Initialize first the binary value you want to find and change by determining the lower limit and upper limit of the binary value first. The smaller the binary lower bound value is initialized or close to the value 0, then the changed color is close to black and changed to white or vice versa if the lower limit initialized approaches 255 then the color is changed to white and colors that are not initialized in binary numbers will be changed turned black.

3) Flood Fill

Flood fill is an algorithm to determine the area connected to a node given in a multi-dimensional array. The purpose of the flood fill method used in this study is to combine background colors in images with background colors on brain objects so that the only remaining black images in the background and white are detected as objects in the brain, either tumors or organs in the brain such as the brain stem . The input used is a binary image, a substitute color binary number, and the image's height and width values. The process is done with 2 checks, namely checking the rows and columns. Checking continues until the size of the rows and columns.

4) Opening dan Closing

This stage is useful for eliminating small noises that are not the object being sought, namely tumors. Small noises that are erased are small organs in the brain that are also detected in the previous process. Opening is the same as erosion but opening is erosion followed by dilation. Images from flood fill are used as input along with kernel values. Then followed by the closing stage. The closing stage is the opposite of the opening stage, namely the dilation followed by erosion of its function is to eliminate the black point or noise on the selected object so that the detected object becomes clearer. Input from this stage is the result of the opening process that has been done and the kernel value.

5) Dilation

At this stage the image that has been cleaned is then dilated to strengthen the image of the object obtained because the image of the previous object has undergone an erosion process and it could be that the image of the object undergoes a slight erosion. The input from this process is the result of the previous closing process, kernel values and iterations.

6) Watershed

Wathershed has an understanding of hills and valleys where hills are of high intensity if grayscale images are seen as the surface of topography and valleys are of low intensity. If a valley is filled with water until it is fully filled, the water will fill the other valleys. Then a barrier is needed where the water will blend. In this study watershed is used to give different colors so that the colors of each other between objects do not mix so that it becomes apparent between the object and background. In Opencv, the watershed algorithm has advantages over other libraries, namely the use of markers which function to determine which points are in the valley that must be merged and not. The color used is the color of HSV.

D. Shape Feature Extraction

The input for this process is the result of the segmentation process. Where only the detected tumor object is white and the background image is black. After that the input data is processed using first-line detection to find the object being processed.

1) Filled Area

The filled area function functions to calculate the area of the object that has been detected. This function can calculate the area even though the object has an irregular shape. The use of this function is very suitable for the form of a tumor that is not always round. First, the input will be labeled first. One input can consist of several labels depending on the results of detection through edge detection. The results of the calculation of the area of the object labeled are not used as a feature used in the classification process, the results of the area are compared first with the area around or the area of the image. The result is a percentage comparison value, the value used for one feature.

2) Centroid

Centroid function aims to find the position where the tumor is located by determining the midpoint of the object detected. The midpoint is used as a benchmark for the position of the x and y coordinates by drawing the x and y lines from the midpoint. The input of this function is the same as input for the function of the filled area. The results of the centroid function are x and y coordinate values. This value is not directly used as a feature for the classification process. The coordinate value is further processed to determine the area where the tumor is located. First the x and y lines are divided into several regions. The x and y coordinate values that have been obtained are compared with the respective boundary values of the region consisting of regions 1, 2, 3, 4, 5 which are at x and y. Each region 1, 2, 3, 4, and 5 has their respective range values. The result value for coordinates with the number of regions included in the region, and the area that is used as a feature for classification.

E. Classification with K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a supervised learning algorithm where the results of newly classified instances are based on the majority of the closest K-neighbor categories. The purpose of this algorithm is to classify new objects based on attributes and samples from training data. The K-Nearest Neighbor algorithm uses neighborhood classification as the predictive value of the value of the new instance. Classifiers do not use any model to match and are only based on memory.

The KNN method algorithm works based on the shortest distance from the query instance to the training sample to determine the KNN. The sample training is projected into a large dimension space, where each dimension represents the features of the data. This space is divided into sections based on the training sample classification. A point in this space is marked by a certain class if that class is the classification most commonly found in the nearest neighbor k from that point. The distance near or far the neighbor can be calculated based on Euclidean Distance which is represented as follows.

$$\mathrm{E}(x,y) = \sqrt{\sum_{i=0}^{n}(xi - yi)^2} \quad (1)$$

The KNN algorithm depends on the data to get the best K value. In general, a high k value will reduce the effect of noise on classification, but make the boundaries between each classification increasingly blurred. Parameter optimization can be used to get good k values, for example by using cross-validation. Special cases where classification is predicted based on the training data closest (in other words, k = 1) [11].

The classification process uses the input results from the feature extraction process, namely the percentage value of the comparison of area, coordinate area of x, and y are and labels obtained from training data. In the classification process there are two different steps, namely during training and at the time of classification. In the process of training data, some of the training data are used to test the accuracy of data classification. Retrieval of training data is done randomly and taken as much as one third of the amount of available training data. Some k neighbor values are used to train data in the classification process. The classification process will divide the data into three classes, namely Astrocytoma, Glioblastoma, and Oligodendroglioma. The training data process serves to determine the value of accuracy of the methods tested for the case studies studied. Accuracy values are calculated based on:

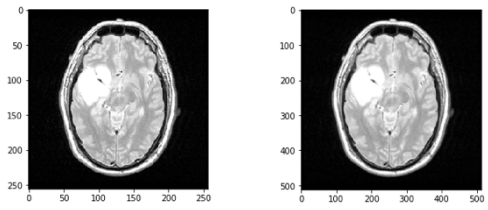
$$Accuracy(\%) = \frac{|True\ Value - Analysis\ Results|}{True\ Value} x100 \quad (2)$$

III. Results and Analysis

The experiment in this paper was carried out through two stages, namely system testing and user testing. At the stage of testing the system is divided into two processes, namely the process of training data and testing data. In the stage of system testing, the process of training data is carried out to all training data available to obtain high accuracy values. The testing process is done after the training process the data gets a high accuracy value. For user trials, it is done through a data validation process by radiologists.

A. Pre-Processing

At this stage the image data is processed to improve the image quality and equalize the variables of each training data image so that there are no variables that change the image feature results to be significantly different. First the data with the format dicom is converted to image data in the png format.



a. Original Image b. After Resize

Fig. 3. Image after image resizing process

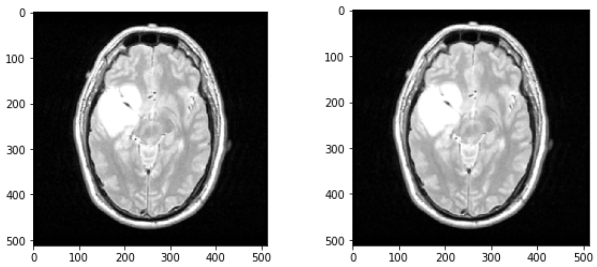
1) Resizing Image

Image data converted to format has a different image size for each file. Size equalization is needed to be 512 x 512 in size to get good and accurate segmentation results. If the image size is too small, the segmentation process cannot

read tumor objects in the brain. Fig. 3 images from the resize process.

2) Median Filtering

The median filter has better noise cleaning results than other filtering techniques for this study. Fig. 4 images resulting from the filter median process.



a. Before Median b. After Median

Fig. 4. Image after the median filter process

3) Cropping Image

Cropping techniques are needed so that images that are processed focus on the object. The cropping process is needed to remove background images that are not brain objects. The size of the image that has been resized will change according to the detected brain object but the size of the pixel does not shrink so that the brain object can still be detected inside.

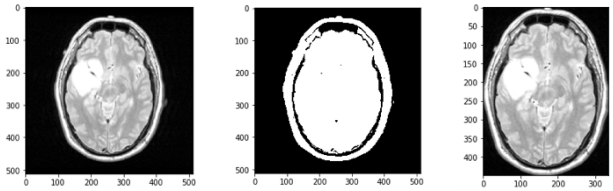


Fig. 5. Image after the auto cropping process

The cropping process is done automatically by doing the thresholding process to detect where the position of the brain is if it is detected, the object will be converted into binary number 1, which is white. While the background is changed to binary number 0. Detection of this brain object is based on the value of the binary color you want to search. This cropping does not fully adjust the shape of the brain object but is rectangular in shape by detecting the length of the brain object at the end and the width of the brain object at the end. This cropping process is done by checking with parameters x and y or length and width. If in one long or wide series no binary number 1 is found, the line is removed or cut. Fig. 5 images resulting from the cropping process.

B. Segmentation Processing

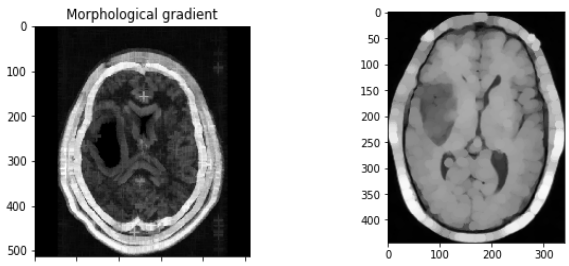
The segmentation process aims to get objects by labeling detected images.

1) Morphological Transformation dan Binarization

This process is done to get or detect objects that are on the inside of brain objects. In the T2 sequence image this process uses edge detection to detect edema in the brain. In the image of the T1 sequence, color detection is used to detect tumors. Use of edge detection on T2 to detect edema because in the T2 sequence image, the edema looks very clear and looks white. So if edge detection is carried out, the detected object has clear shapes and edges and the edges of the object are not broken so that if done binarization results will appear. Different cases if on the T1 sequence image.

In the image of the T1 sequence, there are white and black tumor objects in the white object. When edge detection is carried out, there are many objects in the object. So if a binarization process is carried out then the object will not be detected.

In the T1 sequence image, the threshold value is used with a distance between 90 - 128. This figure is obtained after going through the Exhaustive Searching process, because the color in the T1 sequence image has a grayish black threshold range, while for the brain stem color has a range value deep black. In the T2 sequence image the threshold value used 0-20, the threshold value is obtained after going through the Exhaustive Searching process because after the edge detection process with morphological gradient the color on the object is changed to black with a white object edge. In this process a kernel array value is needed that has different functions in each function. Fig. 6 images of the morphological process in the T1 and T2 sequence images.

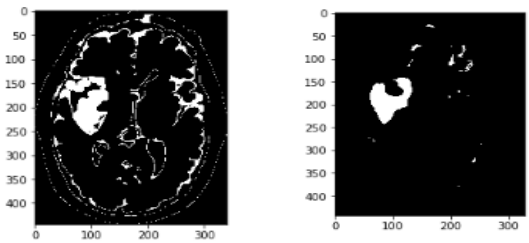


a. b.

Fig. 6. Image after the morphological transformation process. a. Image of T2 sequence, b. Image of T1 sequence.

2) Flood Fill

The floodfill algorithm functions to eliminate images that are not objects by selecting the matrix value for each pixel based on the length and width of the image size. In the floodfill algorithm all pixels are checked to the limit of the length and width of the image. If the pixel checked is not the result of the detected object, the pixel value is changed to a black pixel value of 0. If an object is a fixed pixel value it is 255 or white. Fig. 7 images from the Flood Fill process in the sequence image



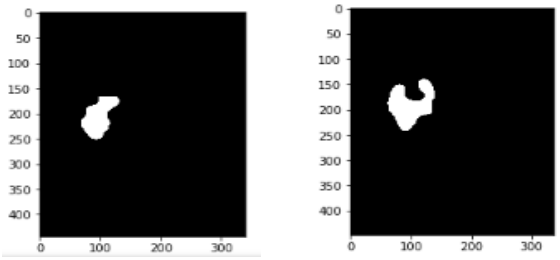
a. b.

Fig. 7. Image after the Flood Fill process. a. Image of T1 sequence, b. Image of T2 sequence.

Whereas in T1 the results of morphological and binary on the background the image is black or the pixel value is 0 so there is no significant difference after the floodfill algorithm is done.

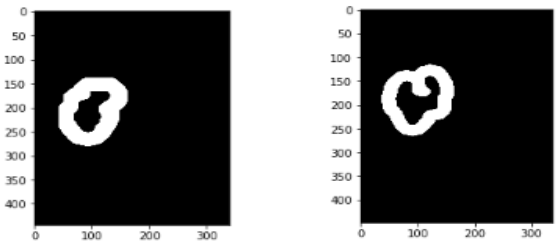
3) Opening dan Closing

In the process of opening and closing the object is removed. Objects that do not fit the size are deleted because they are considered as noise. The process of opening and closing functions is used in the process. The opening function removes noise outside the detected object or image background while the closing function removes the noise inside the detected object. The size of the function removed by the function depends on the dimensions of the kernel array entered. Fig. 8 images from the open and close process on the T1 and T2 sequence images.



a. b.

Fig. 8. Image after the opening and closing process. a. Image of T1 sequence, b. Image of T2 sequence.



a. b.

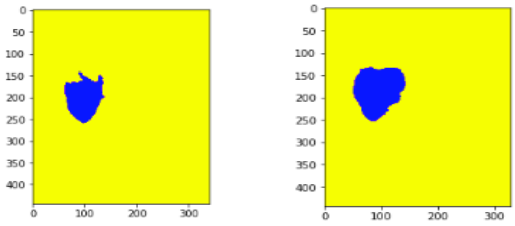
Fig. 9. Image after the dilation process. a. Image of T1 sequence, b. Image of T2 sequence.

4) Dilation

This process aims to restore the detected object shape to its original form in the initial detection process because it has previously undergone several noise cleaning processes. At Fig. 9 found that the detected tumor object is getting wider from the previous process. In the picture, the black object inside is the result of the object after repeated filtering and the white object is the object of the dilation process.

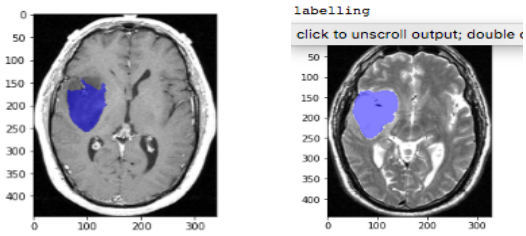
5) Watershed

The watershed process is done by setting the markers first to determine the color position of the different objects so that they are not mixed. After getting different colors between objects and background, then the object segmentation results are moved to the original image by converting the HSV image to RGB. Fig. 10 images from the markers selection process and the coloring of T1 and T2 sequences. Colored markers are then transferred to the initial image to indicate where the tumor is located. Fig. 11 images of the final results of the image segmentation process of T1 and T2 sequences.



a. b.

Fig. 10. Image of the results of the process of determining markers and coloring a. Image of T1 sequence, b. Image of T2 sequence.



a. b.

Fig. 11. Draw the final results of the segmentation process. a. Image of T1 sequence, b. Image of T2 sequence

C. Feature Extraction

At this stage a feature is used to classify the results of the segmentation that has been done. The results of feature extraction are in the form of an area that has been converted into a percentage comparison of tumor area with the area of the image and the x coordinate and the y coordinate obtained from the centroid value obtained from the tumor object.

1) Filled Area

The filled area function gets the area value of object labeling obtained from the markers value in the segmentation process. Labeling objects is done using props. Props are circular labels, so that objects can consist of more than one prop. The area of the props that meet the object is calculated by the function of the filled area. However, the area obtained is not used as a feature value for classification because in the pre-processing process cropping images is done so that the different image sizes are obtained. Different image sizes result in changes in tumor area so that the value of the features obtained is inaccurate. The area of the tumor obtained from the filled area function is then compared with the area of the image that has been cropped. This broad comparison results in a percentage of tumor area to image area. This percentage value is used as a feature value for the classification process because the value obtained is accurate to the tumor area found in the image. This process is performed on both the T1 sequence image and the T2 sequence image. Tables II, III, and IV are the results of the filled area process that shows the value of the tumor area, the area of the image and the broad percentage of the images of T1 and T2 sequences.

TABLE II. EXTENSION OF TUMOR CALCULATION TABLE TOWARDS A PICTURE SAMPLE FROM ASTROCYTOMA CLASS

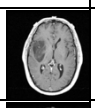
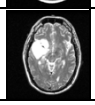
No	Sequence	Image	Image Area	Tumor Area	Percentage
1	T1		115600	5528	4,782
2	T2		107584	7842	7,289

TABLE III. EXTENSION OF TUMOR CALCULATION TABLE TOWARDS A PICTURE SAMPLE FROM GLIOBLASTOMA CLASS

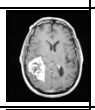
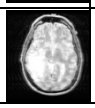
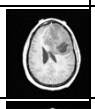
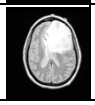
No	Sequence	Image	Image Area	Tumor Area	Percentage
1	T1		128164	10010	7,810
2	T2		128881	12848	9,969

TABLE IV. EXTENSION OF TUMOR CALCULATION TABLE TOWARDS A PICTURE SAMPLE FROM OLIGODENDROGLIOMA CLASS

No	Sequence	Image	Image Area	Tumor Area	Percentage
1	T1		106929	2737	2,560
2	T2		107584	19374	18,008

2) Centroid

At this stage the value of the centroid or the midpoint of the object detected with the same function is regionprops. The regionprops centroid function is an array that stores x coordinates and y coordinates. The values of the x and y coordinates are not the values used for the feature values used in the classification process. The value of features used in the classification process is the position of the x and y coordinates. The position of the x and y values is obtained by dividing the image by n. The value of n is obtained from the initialization of the number of divider values from the desired area for x and y. The value of the quotient marks the length per region. Tables V, VI, and VII show the value of the centroid feature extraction process.

TABLE V. COORDINATE X, Y VALUE TABLE AND TUMOR AREA POSITION IN THE PICTURE SAMPLE FROM ASTROCYTOMA CLASS

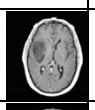
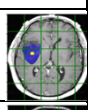
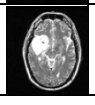
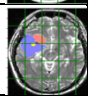
No	Sequence	Image	Coordinate x,y	Position x,y	Result
1	T1		x=98,638 y=201,466	1, 2	
2	T2		x=96,292 y=183,057	1, 2	

TABLE VI. COORDINATE X, Y VALUE TABLE AND TUMOR AREA POSITION IN THE PICTURE SAMPLE FROM GLIOBLASTOMA CLASS

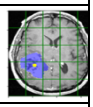
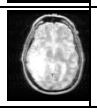
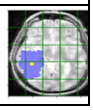
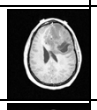
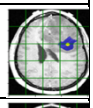
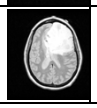
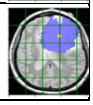
No	Sequence	Image	Coordinate x,y	Position x,y	Result
1	T1		x= 106,931 y= 288,015	1, 3	
2	T2		x= 95,504 y= 285,585	1, 3	

TABLE VII. COORDINATE X, Y VALUE TABLE AND TUMOR AREA POSITION IN THE PICTURE SAMPLE FROM OLIGODENDROGLIOMA CLASS

No	Sequence	Image	Coordinate x,y	Position x,y	Result
1	T1		x=224,302 y=164,734	3, 2	
4	T2		x= 194,813 y= 134,236	2, 1	

D. Classification

In the classification process, model validation is used to divide the data into training and testing in order to obtain the accuracy of testing results from the K-Nearest Neighbor classification method. The second model validation technique is Stratified Random Sampling. This model validation technique is a model validation technique by taking testing data from each existing label or class with a predetermined amount of testing data. In this technique the features that have been obtained from the dataset are stored in CSV which varies according to the class. Each CSV file is processed sequentially by taking testing data as much as the desired data. Data was taken randomly, the previous permutation process was carried out. After all labels have testing data then training data and testing data for each class are combined into one Frame Data. An illustration of the validation technique of the Stratified Random Sampling model is in Table VIII.

TABLE VIII. ILLUSTRATION OF VALIDATION OF STRATIFIED RANDOM SAMPLING MODELS

No	Class	Amount of Data	Amount of Testing Data
1	Astrocytoma	84	(n=0,33)*84 = 28
2	Glioblastoma	124	(n=0,33)*124 = 41
3	Oligodendroglioma	46	(n=0,33)*46 = 15

Testing on each validation method is done three times per n_neighbor value with random testing data of one third the amount of data used or 33% and with n neighbor values 1, 5, and 10. The results of the accuracy of testing data for each model validation method are Table IX.

The validation technique of the Stratified Random Sampling model has high accuracy results with the k_neighbor value because of the spread of training data and testing data that are evenly distributed so that when testing data using the K-Nearest Neighbor classification, the data obtained from the training data process is sufficient so that

there is a slight error in labeling new class during the data testing process.

TABLE IX. CLASSIFICATION ACCURACY RESULTS USING K-NEAREST NEIGHBOR METHOD

No	k_neighbor	Trial (%)			Average Accuracy (%)
		1	2	3	
1	1	89,9	89,9	88,6	89,5
2	5	75,9	81,0	64,6	73,8
3	10	69,6	73,4	60,8	67,9

The value of k used in the classification process using the K-Nearest Neighbor method depends on the amount of data used. The more data used, the greater the k value that can be tested. The amount number of substraction values in the accuracy of k value or the number of neighbors could depends on the number of things that must be considered by the KNN algorithm. If there are any values differences in the closest neighbors, The KNN algorithm will classify the classes based on the most dominant class in the nearest neighbour results. Fig. 12 Desktop application display of brain tumor classification results of user testing.

Fig. 12. Desktop application for classification of brain tumors

Table X shows a comparison between the accuracy of the K-Nearest Neighbor method with the other methods, known as the Gaussian Naïve Bayes method and the PNN (Probabilistic Neural Network). This accuracy is obtained by comparing the classification methods using the same training data and testing data, as well as the pre-processing, segmentation, and feature extraction processes.

TABLE X. COMPARISON OF CLASSIFICATION METHODS

No	Classification Method	Accuracy (%)
1	K-Nearest Neighbor	89,5
2	Gaussian Naïve Bayes	71,1
3	Probabilistic Neural Network	76,5

IV. CONCLUSION

Accuracy results obtained from the classification of brain tumors using the K-Nearest Neighbor method are relatively high at 89.5 percent using the Stratified Random Sampling technique in the data validation model when compared with other classification methods, Gaussian Naive Bayes and Probabilistic Neural Network. However, the accuracy of the classification process can be decreased due to the spread of training data and testing or models of validation and features used. In this study feature extraction is very dependent on the results of the segmentation process obtained. Factors that can influence the results of segmentation can be done in the pre-processing process so that the accuracy of the results of the segmentation process is obtained. The low accuracy obtained is due to the lack of features used and the unprocessed segmentation process is more complex. The features used in this study are only a small part of the factors used to determine tumor class. In the process of segmentation the use of color detection and object edge detection are still considered lacking to obtain good and accurate segmentation results. Machine learning process is needed in the segmentation process to get good, accurate and auto detection segmentation results.

ACKNOWLEDGMENT

The author expresses his gratitude to SMF Radiology Hospital (Regional General Hospital) Dr. Soetomo Surabaya for assistance with data collection techniques, methodologies, and constructive suggestions for writing this manuscript.

REFERENCES

[1] R. Hruban, "What Are Tumors," John Hopkins University, 2018. [Online]. Available: http://pathology.jhu.edu/pc/BasicTypes1.php. [Diakses 22 January 2019].

[2] WHO, "Cancer," World Health Organization, 12 September 2018. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer. [Diakses 22 January 2019].

[3] A. C. A. M. N. B. A. Kwan-Hoong Ng, "Magnetic Resonance Imaging: Health Effects and Safety," dalam *Proceedings of the International Conference on Non-Ionizing Radiation at UNITEN*, Malaysia, 2003.

[4] C. K. R. J. T. Catherine Westbrook, MRI in Practice, Southern Gate, Chichester: John Wiley & Sons, Ltd, 2011.

[5] C. Westbrook, Handbook Of MRI Technique, Southern Gate, Chichester: John Wiley & Sons, Ltd, 2014.

[6] M. S. P. R. G. D. .. Balakumar B., "An Automatic Brain Tumor Detection and Segmentation Scheme for Clinical Brain Images," *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS),* 2014.

[7] K. S. S. V. G. Dilip Kumar Gandhi, "Detection of Brain Tumor and Extraction of Texture Features using Magnetic Resonance Images," *International Journal of Engineering Innovation & Research,* vol. 1, no. 5, 2012.

[8] P.-M. J. H. L. Mohammad Havaei, "Efficient Interactive Brain Tumor Segmentation as Within-Brain KNN Classification," dalam *IEEE International Conference on Pattern Recognition*, Stockholm, 2014.

[9] R. D. N. Narmada M. Balasooriya, "A Sophisticated Convolutional Neural Network Model for Brain Tumor Classification," dalam *IEEE International Conference on Industrial and Information Systems*, Peradeniya, 2017.

[10] B. V. K. S. J. F. J. K. P. K. S. M. S. P. D. M. M. P. L. T. F. P. Kenneth Clark, "The Cancer Imaging Archieve (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging,* vol. 26, no. 6, pp. 1045-1057, 2013.

[11] M. Gelar, "Analisis Sentimen Mengenai Berita Menggunakan Twitter Studi Kasus Detik.com dan Kompas.com," *Universitas Widyatama,* 2015.