

Brain Tumor Diagnosis Support System: A Decision Fusion Framework

Kalifa Shantta^{1#} & Otman Basir²

^{1,2}Electrical and Computer Department, University of Waterloo, 200 University Ave W,
Waterloo, ON N2L 3G1, Canada.

#corresponding author.

Type of Review: Peer Reviewed.

DOI: <http://dx.doi.org/10.21013/jas.v15.n3.p1>

How to cite this paper:

Shantta, K, Basir, O. (2020). Brain Tumor Diagnosis Support System: A Decision Fusion Framework. *IRA International Journal of Applied Sciences* (ISSN 2455-4499), 15(3), 30-47. doi:<http://dx.doi.org/10.21013/jas.v15.n3.p1>

© Institute of Research Advances.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License subject to a proper citation to the publication source of the work.

Disclaimer: The scholarly papers as reviewed and published by the Institute of Research Advances (IRA) are the views and opinions of their respective authors and are not the views or opinions of the IRA. The IRA disclaims of any harm or loss caused due to the published content to any party.

Institute of Research Advances is an institutional publisher member of *Publishers International Linking Association Inc. (PILA-CrossRef)*, USA. The institute is an institutional signatory to the *Budapest Open Access Initiative, Hungary* advocating the open-access of scientific and scholarly knowledge. The Institute is a registered content provider under *Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH)*.

The journal is indexed & included in *CAS Source Index of Chemical Abstracts Service of American Chemical Society (USA)*, *Index Copernicus (IC Value 85.27)*, *WorldCat Discovery Service (USA)*, *CrossRef Metadata Search (USA)*, *WorldCat (USA)*, *OCLC (USA)*, *Open J-Gate (India)*, *EZB (Germany)* *Scilit (Switzerland)*, *Airiti (China)*, *Bielefeld Academic Search Engine (BASE) of Bielefeld University, Germany*, *PKP Index of Simon Fraser University, Canada*.

ABSTRACT

The early and accurate detection of brain tumors is important in providing effective and efficient therapy and thus can result in increased survival rates. Current image-based tumor detection and diagnosis methods depend heavily on the interpretation of the neuro specialists and/or radiologists. Therefore, it is quite possible for the interpretation process to be time-consuming, and prone to human error and subjectivity. Automatic detection and classification of brain tumors have the potential to achieve efficiency and higher degree of predictable accuracy. However, it is well established that the accuracy performance of automatic detection and classification techniques varies from technique to technique, and tends to be image modality dependent. Thus, it is prudent to explore the variability in the performance of these techniques as a means to achieve consistent high accuracy performance. This paper presents a framework for fusing multiple tumor classifiers. The fusion process is based on the Dempster Shafer evidence fusion theory. Several tumor classifiers are employed. Experimental results will be presented to validate the efficiency of the proposed framework. It is concluded that fusing the classification decisions made by the various classifiers it is conceivable that efficient and consistent high accuracy classification performance can be attained.

Keywords: Multi-Classifier, Decision Fusion, Discreet wavelet transform, Dempster Shafer Evidence Fusion Theory, Accuracy Classification.

Introduction

Medical imaging has proven useful as a tool for detecting and classifying brain tumors. The imaging process for detection and classification often requires experienced radiologists to read the images. Due to the inherent subjectivity of the human decision process, the diagnostic process may be erroneous. For example, MRI imaging offers high resolution and hence an ability to display clear brain structures, tumor size, and position. Also, MRI is particularly useful in classifying brain tissues, whether cancerous or not. The amount and complexity of information contained in MRI can be overwhelming to radiologists. Brady et al. [1] show that radiology includes decision-making under a situation of doubt, and consequently, is not guaranteed correct explanation and opinions.

Furthermore, Radiologist involves decision-making under conditions of ambiguity, and therefore, is not guaranteed to yield accurate explication or statements. For instance, in the Niagara region, Canada, some hospitals have decided to audit 4000 CT scans, MRI, and mammograms after discovering mistakes made by a radiologist in interpreting and classifying some images. It is estimated that it took four months to complete this audit, 2015 [2]-[3]-[4]. Two hospitals in Mississauga, Canada, in the period of one year from April 2012, to March 2013, have determined that in 3,500 CT scans and mammograms, the medical specialist had incorrectly classified and interpreted scan results [5]-[6]. In Vancouver general hospital, between October 2016 and January 2017, 5,278 images from 8,400 CT scans found inconsistencies with the reading and reporting medical photos. As modern imaging modalities become more complex, especially CT and MRI, it is presently conventional to translate clinical photographs to take a longer time to process them. The analysis of these images and make the right decision is an important task. Therefore, machine-learning algorithms/applications are demanded in the medical imaging field. Also, medical imaging applications are becoming more complicated, with a more substantial need to automate the analysis and introduce machine learning techniques to classify images faster and automatically.

Background and Previous Methods

Recently several techniques have been offered for image segmentation and classification procedures of brain cancers. Most of them are based on one or two algorithms with different feature extraction & selection techniques. This paper introduces Multi-Classifier with decision combination based on the elementary combining classifier and Dempster Shafer theory to achieve high confidence and accuracy. Next is the review of recent methods:

Singh, A. [7], offered data recognition techniques for the classification of MRI images. The suggested method is accomplished in 4 steps: preprocessing, segmentation, feature extraction, and classification. The primary phase, improvement, and skull stripping is completed to develop speed and efficiency. The fuzzy C-mean clustering technique is applied in segmentation. A gray-level matrix does the extraction of the MRI image features. The last stage involves the support vector machine to categorize the given pictures. The result shows high precision and efficiency in the MRI image classification.

Salankar et al. [8], in their paper, presented classification procedures conducted by Support Vector Machines (SVM) in order to distinguish healthy and unusual MRI brain images. Feature extraction from MRI Images was obtained by the grey level of texture features. Extracted features decreased by utilizing the PCA approach and then applying SVM for training and testing. They conclude this achieved more accurate results than did the other techniques.

Abdulrazzaq et al. [9], they proposed an approach for an automatic classification utilizing multi classifier. They extend their previous work that is the achievement of various feature types were studied by applying two classification methods. One is the kNN classifier, and the other is SVM. They are examining the outcome of fusing these two algorithms. Their experiments demonstrate precision enhancements based on utilizing Image CLEF2005 dataset.

Ubaidillah et al. [10], in their paper, offered a comparable study of tumor recognition employing an artificial neural network and support vector machine utilizing four different cancer datasets. The ANN and SVM classification models produced using four stages: input variable selection, data preprocessing and partitioning, and setting of model parameter and model implementation. In the first phase, input variables chose based on the type of datasets. Data normalization and data conversion are implemented in data preprocessing, and data partitioning consists of division of data into two partitions which are training and testing set. Multiple parameters are recognized for the setting of ANN and SVM model. The final stage is model implementation in which classification model is generated for ANN and SVM. The selected classification model is then tested on the testing dataset. Experimental results point that the SVM classifier produced an excellent result for tumor detection.

Chinnu A. [11], proposed the SVM algorithm to classify MRI brain tumor. The segmentation image is done based on a histogram. Offered methodology depends of following central steps: preprocessing, segmentation, feature extraction, and classification. First stage, noise reduction and edge detection are performed using the median filter and Canny Edge detector technique respectively. Histogram thresholding method is utilized for segmentation. Feature extraction from MRI pictures is completed by gray level, symmetrical and texture features. Classification of MR images is conducted using support vector machine. A result shows the improvement of precision degree and a decrease in the fault degree of MRI brain cancer.

Nikame and Shind[12], introduced brain picture classification and detection employing distance classifier scheme. This thesis offers a system for automated prediction of a normal or abnormal utilizing Region growing segmentation by a watershed technique, Euclidean distance algorithm for high-speed calculation, followed with pre-processing and post-processing scheme applied on a database containing both healthy and cancer types of MR brain pictures. This system had two main steps, first, is pre-processing of MRI images and then other post processing operation, which includes operations like noise removal, convert the input image into gray scale image. The segmentation process is applied by threshold method; it is the most popular procedure for detecting significant discontinuities in gray level, the second used morphological operations and featured extracting process. Their work used Euclidean distance classifier; this classifier based on the distance measure is straightforward and not complicated. The results guarantee that the system is effective, and meeting for fast discovery whether the patient is healthy or unhealthy.

R. J.Deshmukh and R.S Khule[13]; they introduced Neuro-fuzzy systems use the fused power of two methods: fuzzy logic and (ANN) applying to identify the brain tumor. This task performed by processing of MRI images of brain cancer for detection and Organization on different kinds of brain cancers. A proper Neuro Fuzzy classifier is used to distinguish the various types of brain tumors. Stages which are conducted out for detection of a tumor are training the neural network, test the MRI image with the knowledge base, and finally, the result will be tumor detect or not detected. Also, the features applied to Neuro fuzzy classifier to detect a candidate circumscribed tumor. Typically, the input zone created from seven nodes communicating to the 7 attributes. The output region formed of single node showing whether the MRI is a candidate circumscribed tumor or not and the hidden zone switches based on the some fuzzy commands that produce greatest classification degree for every set of attributes.

Bruce et al.[14], presented classification techniques predicated on multi classifier and decision fusion for mammogram tumors. The feature dimensions are divided into different smaller sized spaces and applied the different classifier to perform the classification process in each of the partitions, and at the end merges the decision from the classifiers in one decision. They implement this system to categorize the mammogram scans as either benign or malignant.

Methodology

The introduced system's main stages are the Pre-processing stage applying a median filter, image segmentation using thresholding technique, and discrete wavelet transform DWT used in the features extraction stage. A considerable amount of attributes vector, the principal component analysis (PCA) procedure was employed to decrease the features after three DWT decomposing levels for the best accuracy. The Multi-Classifiers stage consists of five algorithms: Naïve Bayes, SVM, K-NN, ANN, and Decision tree classifiers. All of which use the supervised training approach were used for the image classification tasks. They differ actually in their approach on how to classify data.

The decision fusion step is the latest stage where the elementary combining classifier applied to achieve the last decision. The Dempster Shafer theory is the primary fusion technique utilized to fuse the outcomes from multi algorithms. The offered method in this work has been used on the image database downloaded from two web sites. The first data set was downloaded from Harvard Medical School and the second dataset from the Oasis website. The next sections provided a review of pre-processing, image segmentation, wavelet decomposition, principal component analysis, and decision fusion stage.

Pre-Processing Stage

The MRI/CT images contain film artifacts or labels such as a person's surname, age, and marks. In such a broad diversity of image-processing applications, it is necessary to smooth an image while maintaining its edges. The grey levels usually interfere, which causes any of the following stages, such as segmentation, feature extraction, and labeling, to be more challenging. Filtering is possibly the primary procedure in various medical images' classification assignments. The function is to reduce the noise level and to enhance the condition of the picture. The median filter is accomplished by applying the pixel values within a defined window and to place them in numerical order to determine the median value; subsequently, the median value is utilized to substitute the pixel under consideration.

Compared with the mean filter, the median filter depends on the median value rather than the average as in the mean filter. The median of a set is more convincing regarding the presence of noise. In this situation, apply the median filter to eliminate the artifact from the input image. This process will exclude artifacts from the image and will also retain the data from ROI. Figure. 1 [See *Annexure A: Figures & Tables*] depicts two MRIs before and following the application of the median filter. This filter is useful when compared with other techniques and produces output images which are proper for additional processing.

Segmentation Scheme Using Thresholding Technique

The divided volumetric medical image is the segmentation phase's objective, typically anatomic structures (tissue types) that are essential for a particular task. Specifically, it is utilized to split up regions from the rest of the image, to observe or recognize them as objects. The thresholding approaches are applied in the proposed system. Thresholding is the most straightforward and most ordinarily utilized technique of segmentation. A binary region map or binary image is obtained with only one threshold by converting grayscale or colour image. The binary map contains two, possibly separated areas, the first one having pixels intensity with input data values less than a threshold "background," and the second area belonged to the input values that are at or exceeding the threshold "foreground" [15][16]. Figure. 2 [See *Annexure A: Figures & Tables*] shows the MRI image and the result after applying the thresholding technique. The threshold is a value in a gray level that splits pixel strengths into binary portions. An incorrect threshold value results in a low segmentation process [17]. If there is more than one region to extract with different gray levels, more than one threshold is multi thresholding.

Feature Extraction Scheme Using DWT

The discrete wavelet transform (DWT) coefficients are used in the proposed system as an input feature. Wavelet transform disintegrates a signal into a collection of basic functions. These basis functions are named wavelets. Thakur [18] states that wavelets' primary interest is that they offer the time and frequency representation, which is especially useful for the classification process. Figure. 3 [See *Annexure A: Figures & Tables*] shows 2D DWT. Wavelets are achieved from a single prototype wavelet called mother wavelet by scaling and shifting [19]:

$$w_{\varphi}(a, b) = \int_{-\infty}^{\infty} x(t) * \varphi_{a,b}(t) dx \quad (1)$$

Where

$$w_{\varphi_{a,b}}(t) = \frac{1}{\sqrt{|d|}} \varphi\left(\frac{t-c}{d}\right) \quad (2)$$

By applying the scaling and shifting, the wavelet $\Psi(c,d)$ can be calculated from the mother wavelet Ψ , where c is the shifting parameter, and d is the scaling factor. The DWT method work for separate the data into various frequency resolutions. Repeating the decomposition method divides the input signal into multiple lower-resolution segments. DWT can be represented as [20]:

$$DWT_{x(n)} = \begin{cases} d_{j,k} = \sum x(n) h_j^*(n - 2jk) \\ a_{j,k} = \sum x(n) g_j^*(n - 2jk) \end{cases} \quad (3)$$

In this study, three levels of decomposition for each image are applied; the Haar wavelet transform is performed. The Haar wavelet is considered as uncomplicated wavelets. The wave representation is period of a square wave. Image size after first decomposition level is 128×128, level two 64×64 and level three 32×32. At this stage, we obtain a linear matrix with an image size of 1×1024 matrix. Next figure.4 shows the four sub band LL (low-low), LH (low-high), HL (high-low), and HH(high-high) of the first level of DWT decomposition. Fundamentally the medical images require higher precision without failure of data. The Discrete Wavelet Transform (DWT) was based on time-scale representation, which offers effective multi-resolution.

Feature Reduction Scheme Using PCA

Large numbers of features increase the execution time and required storage memory. Also, it makes the classification process more complicated. Therefore, it required reducing the number of attributes. PCA is one useful tool to decrease the data set dimension containing many correlated variables; however, holding a maximum of the variance. It transforms the data set into a novel set variable based on their variances or importance. The PCA algorithm has three properties: the components of the input patterns orthogonalized so that no correlation among them, the most considerable variation come first as the resulting of orthogonal components, and ignore those components are offering the smaller to the diversity in the data set. It is well-known that before applying PCA, the input data is normalized to have zero mean and unity variance. The dataset consists of MRI image T2 256 x 256 and more information about the dataset in section VI. Thus, the extracted features were decreased from 65536 to 1024 after applying three levels of decomposition DWT.

Nevertheless, these features are still extensive for the classification stage; therefore, PCA is applied to decrease the size of attributes to a lower degree. It shows that only 13 principal components, which are around 1.85% of the primary data, might maintain 95% of the overall variance. Figure. 5 [See *Annexure A: Figures & Tables*] shows PCA schematically.

Decision Fusion

The objective of all decision fusion systems is to generate a model, which given the smallest number of input data, is able of producing proper decisions. Ludmila [21] groups classifiers output into three types:

1. Abstract or class label: every classifier provides the class label for every single input vector. Every classifier D_i creates a class label $s_{i,i} = 1, \dots, K$. Therefore, for any object $x \in \mathbb{R}^n$ to be classified, the K classifier outputs state that a vector $s = [s_1, \dots, s_k]^T \in \Omega^k$ where $\Omega = \{w_1, w_2, \dots, w_c\}$ is the set of class labels.
2. Class rank: the classifier provides a ranking record of all potential labels for every input pattern. The first location represents the utmost possible class where the final one is the most unlikely class.
3. Measurement or soft/fuzzy outputs: with the information of class rank, the classifier assigns a weight or probability to each class. Every classifier creates a bi-dimensional vector $[d_{j,1}, \dots, d_{i,b}]^T$.

Figure 6 [See *Annexure A: Figures & Tables*] illustrates the applied techniques utilised to combine the multi-classifier based on the classifier output type. For example, if the classifiers' output of rank or abstract type applies, certain fusion techniques may be applied, such as majority vote, weighted majority vote, and Bayesian combination.

However, with regard to the probability schemes such as Min, Max, product, and the average probability as well as the Dempster-Shafer Theory could be applied if the classifier output from the measurement type applies. [22] .

Elementary Combiners

In this research work, the combination rules such as Min, Max, average, and product probability with the majority vote rule are applied to fuse the multi-classifiers as a first technique [23]. The majority vote decision applies when all of the classifiers vote for one class or more than 50 percent, plus one of the classifiers vote for the same category. Therefore, to formulate the concluding classification, decision outputs from every classifier were combined. Following this, there is a majority vote rule that satisfies:

$$R_r(A) = \sum_{j=1}^k d_{c,j} (A) \quad (4)$$

where c is the class, A is the testing pattern vector, j = 1, 2, . . . , K where k number of algorithms. The selection of an odd number to avoid a tie in the majority vote method; dc,j is the paired resolution value {0, 1}, 0 matches the incorrect classification, and 1 the correct category.

Ponti Jr [24] shows that a combination of classifiers on the measurement level of all these rules can be applied:

1. Min: from among the classifiers, calculates the minimum score of each class and sets the unknown testing features to the class which has the maximum grade.

$$R_c^{min} (A) = \min p \left(\frac{c}{A} \right). (5)$$

2. Max: from among the classifiers finds the maximum result of every class and algorithms the unknown sample to the class which has the maximum grade between the maximum score.

$$s R_c^{max} (A) = \max p \left(\frac{c}{A} \right). (6)$$

3. Product: multiply the scores created from every classifier and set the class label of the maximum score to the unknown input attribute.

$$R_c^{product} (A) = \prod_{i=1}^k p \left(\frac{c}{A} \right). (7)$$

4. Sum: adds the grade created by every single classifier and sets the class label of the maximum result to the unknown input attribute.

$$R_c^{sum} (A) = \sum_{i=1}^k p \left(\frac{c}{A} \right). (8)$$

Dempster Shafer Theory of Evidence

The DST concept has been used to deal with ambiguity management and imperfect reasoning. Unlike the Bayesian approach, the DS theory can explicitly model the unknown information. The accumulation of evidence is used to narrow down a set of hypotheses. DS method allows the demonstration of ignorance due to the ambiguity of the proof. If the ignorance's value reaches zero, the DS model is reduced to a standard Bayesian model. X is represented by basic belief m(X) delivered by the source of evidence under consideration. Figure .7 [See Annexure A: Figures & Tables] demonstrate different measurements over a unit interval and has the following features:

1. The green area over the unit interval represents the belief.
2. The red area represents the disbelief.
3. The grey area represents ignorance, which indicates that neither the belief nor the disbelief range is selected.
4. Plausibility, pls(B) is the addition of uncertainty and belief measures. This indicates that most stretches of belief are in ignorance, and do not demonstrate trust nor disprove it.
5. The doubt is entire uncertainty and disbelief.

$$\sum_{x \subseteq \Theta} m(X) = 1 \quad \text{and} \quad m(\emptyset) = 0 \quad (9)$$

When ϕ is empty, it shows the certainty that an empty set is always equal to zero and that Θ characterizes the entire frame of discernment. The trust function for the occurrence of D is specified by:

$$Bel(D) = \sum m(X), \quad X \subseteq D \text{ and } D \subseteq \Theta \quad (10)$$

In this work, the classifier output provides the evidence and the Dempster Shafer combination rule is able to process this evidence. The source of proof is not dependent; neither does the intersection set presuppose the empty rule. Dempster's combination rule may be applied in the fusing of any two beliefs such as Bel_A and Bel_B in order to produce a unique confidence function. Dempster's Rule of fusion is a technique that fuses proof from several independent sources. Furthermore, the probability mass functions are fused by using the Dempster Rule, on the assumption that Bel A and Bel B are two belief assignments over the event space Θ, with probability masses m_A and m_B , respectively. Therefore, the total possibility mass proposes that c is:

$$m(c) = K \sum_{a_i \cap b_j} m_A(a_i) * m_B(b_j) \quad (11)$$

where,

K is the normalizing factor. This function is called the orthogonal sum of Bel_A and Bel_B, indicated as Bel_A ⊕ Bel_B [25].

This sum can also be denoted as $m_A \oplus m_B$ which is:

$$m(c) = K[m_A(a_i) * m_B(b_j)] \quad (12)$$

The normalising factor, K assists as a measure of the conflict between the two certainty functions which is given by:

$$K = \frac{1}{1-k} = \frac{1}{1-\sum_{a_i \cap b_j} [m_A(a_i) * m_B(b_j)]} \quad (13)$$

Furthermore, k is the so-called amount of conflict between the two belief functions. If Bel_A and Bel_B do not conflict, then $k = 0$. If $K = 1$ then the functions totally contradict and Bel_A ⊕ Bel_B does not exist [26].

This conflict factor will be monitored during the combination of the evidence and if it is equal to the predetermined thresholds, it will contradict, and that the decision is uncertain and that further testing is required. Generally, there are more than two sources of proof for a proposition. In order to fuse several certainty functions, Dempster's Rule is repeatedly applied to pairs of functions. In this work, there are five belief functions: Bel_{ANN}, Bel_{SVM}, Bel_{KNN}, Bel_D, and Bel_{Bayes}. Initially, Bel_{ANN} and Bel_{SVM} are combined and then Bel_{ANN} ⊕ Bel_{SVM} is fused with Bel_{KNN}, and so on. The final sum is Bel_{ANN} ⊕ Bel_{SVM} ⊕ Bel_{KNN} ⊕ Bel_D ⊕ Bel_{Naive}. Nevertheless, the order of fusion is of no consequence. The calculated mathematics of the rule is given by [27]:

$$Bel(C) = \frac{\sum_{A_i \cap B_i = C; C \neq \emptyset} Bel(A_i) \times Bel(B_i)}{1 - \sum_{A_i \cap B_i = \emptyset} Bel(A_j) \times Bel(B_j)} \quad (14)$$

The pairwise fusion method is applied. In the first stage it fuses, for example, the opinions of KNN classifiers (K) and supports vector machine algorithm (S). Tabel. 1 [See Annexure A: Figures & Tables] show pairwise fusion of KNN and SVM classifiers.

In the next stage, it fuses the result from the last combination of the SVM and KNN classifiers where the proof produced by the D-Tree classifier (D). Let $bel_{KNN(B)}$ and $bel_{KNN(M)}$, represent the beliefs from the KNN classifier for both classes as benign (B) and malignant (M). Likewise, for the Naive Bayes classifier evidence is given as $bel_{Bayes(B)}$ and $bel_{Bayes(M)}$, where U_{Bayes} and U_{KNN} are the unbelief or uncertainty of the two classifiers. Bel(M) is a trust mass specified to classify malignantly. This is calculated by a product of benign trust of Bayes and KNN, considering the independence of the pieces of evidence sources. The multiplication of the benign belief of KNN and the uncertainty of Bayes and the uncertainty of KNN and the benign belief of Bayes is added and all these basic opinions are summed. Therefore :

$$bel_{comb(B)} = bel_{Bayes(B)} \times bel_{KNN(B)} + U_{Bayes} \times bel_{KNN(B)} + bel_{Bayes(B)} \times U_{KNN} \quad (15)$$

Steps of Combination

Dempster's combination includes an estimation of belief and unbelief or uncertainty resulting from each classifier. Figure.8 [See Annexure A: Figures & Tables] illustrates the inputs and outputs of each classifier. Output 'K' indicates the belief values obtained from the k-Nearest Neighbour while 'S' indicates the belief values from the support vector machine. 'A' indicates the belief values from the artificial neural network. The 'D' output belief is from the decision tree, and 'N' indicates the belief values from Naive Bayesian. The first phase combines evidence

from the KNN classifier and supports the vector machine classifier outcomes. The uncertainty and beliefs are applied to Dempster's rule as input. S and K are the evidence which provides beliefs from the support vector machine and the KNN classifier respectively. The beliefs $belKNN(B)$ and $belKNN(M)$, where $belKNN$ indicate the belief provided by the KNN and two classes, benign (B) and malignant (M) under study. Likewise, the support vector machine classifier beliefs are given as $belSVM(B)$ and $belSVM(M)$. The uncertainties for two classifiers are $UKNN$ and $USVM$; benign and malign classes are achieved from SVM and k-nearest neighbour classifiers respectively.

Brain Tumor Detection Support System

The proposed solution is to introduce powerful decision fusion frameworks that coordinate multi-classifiers' decisions into a single decision, as shown in Figure. 9 [See *Annexure A: Figures & Tables*].

The brain tumor detection support system is implemented to classify digital MRI and CT images to detect brain tumors as either benign or malignant. Artificial intelligence will build innovative image analysis tools to detect diseases such as brain cancer in a more targeted and effective way to resolve classifying medical images with high accuracy of classification. Additionally, the development of an effective detection and classification system helps physicians know the location and type of tumor at an appropriate time.

The proposed system block diagrams consist of two stages, a training stage, and a testing stage. Both phases consist of the nextphases: pre-processing, segmentation, feature extraction depending on DWT, feature decrease by applying PCA, multi-classifier, and training accuracy assessment; finally, the testing phase, the last stage is the decision fusion, which is responsible for combining the different decisions from that which resulted from the multi-classifier stage. The elementary combining classifier, such as majority voting, minimum probability, maximum probability, product probability, and average probability, is applied to combine considerable evidence from the classification stage. The Dempster-Shafer theory of evidence is used as the principal technique to fuse the multi-decision to arrive at one final decision and express any uncertainty regarding this decision.

Training Phase

The first stage is pre-processing, applying the median filter followed by image segmentation, applying the threshold technique, and feature extraction by using the DWT technique. The length of the feature vectors is reduced by implementing the PCA method. The set of compact feature vectors and the class label are utilized to learn the multi-classifier group. A cross-validation technique is applied for a successful generalization capability of the system. Additionally, the training accuracy evolution is used later in the decision combination system; the combination function allows robust classifiers to participate with greater involvement in the combination task.

Testing Phase

The user or the radiologist submit the brain MRI image, which is to be classified. The segmentation and feature extraction is applied and also the PCA to reduce the feature dimension. This decreased feature vector of dimension is used as the input to the multi-classifier. The multi-classifier group is a set of algorithms that produce "local" choices combined into a one-class label (abnormal or normal), thereby utilizing a suitable decision combination rule. The most critical stage is the decision fusion. All the local decisions are fused into one conclusion: normal or abnormal (benign or malignant) using different elementary combinations such as majority voting, weighting averaging, minimum, maximum, and probability schemes. The primary combination method is the Dempster-Shafer theory of evidence.

Experimental Work

Complete experiments were carried to assess the achievement of the designed framework in brain tumour diagnostic support systems. The data sets are two-weighted 256 x 256 in-plane resolution. One data set was obtained from the Harvard Medical School website, (<http://med.harvard.edu/AANLIB/>), and the second was downloaded from the OASIS website (<http://oasis-brains.org/>). These are benchmark data sets which are utilised in brain MRI image analysis tasks, and contain both benign and malignant MRI brain images. The first benchmark data set contains 66 brain MRI images (18 benign and 48 malignant). The second dataset consists of 160 MRI brain images, of which 20 are normal and 140 abnormal as shown in Table 2 [See *Annexure A: Figures & Tables*]. These images have been

labeled by experts. The experiment was conducted and divided into two stages. Firstly, each classifier was evaluated individually by the same datasets, and also the number of the features required achieving the highest accuracy. Secondly, the combination of the multi-classifier is undertaken by different methods such as: minimum and maximum probabilities, majority vote, and average probability. The confusion matrix Figure 10 [See *Annexure A: Figures & Tables*] was applied to determine the performance of the algorithms both in individual and combination classifiers' tasks. There are four possible outcomes from a two class predication: True positive TP, True Negative TN, False Positive FP and False Negative FN. The benign and malignant images are correctly classified as true negative and true positive respectively.

A False Positive represents the classification of all incorrect results as being malignant where they are benign. False Positive is the false signal in the recognition task. A False-negative represents the classification of all incorrect results as benign where they are malignant. In this system:

1. Sensitivity TP: the possibility that a detection examination is positive when the patient actually has a tumour.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (16)$$

2. Specificity TN : the possibility that a detection examination is negative when the patient is cancer free.

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (17)$$

3. Accuracy: the possibility that a detection examination is correctly completed.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad (18)$$

4. Precision: the fraction of abnormal images with correct results.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (19)$$

where:

TP= Number of malignant cases-correctly labelled.

TN= Number of benign cases-correctly labelled.

FP= Number of benign cases designated as malignant.

FN= Number of malignant cases designated as benign.

Result and Discussion

First of all, we evaluate the number of features that produce the classification process's high accuracy performance and feature reduction effectiveness. The introduced system is based on the DWT decomposition for feature extraction. After three levels of decomposition configuration, the size of the LL is 32×32 . PCA is utilized to decrease the feature vector size from 65536 to 1024, where these features are still extensive. Thus, principal components are used to reduce these features. The variances against the number of principal components from one to twenty shows at the latest nineteen principal components, which are simply 1.85% of the principal attributes, could preserve 95.4% of the total variance.

To find out the correct number of principal components, which give the best result, the Multi-Classifier's performance was experimented with different principal components up to 18. The graphs of Figure. 11 and Figure.12 [See *Annexure A: Figures & Tables*] display the achievement of the algorithms in terms of accuracy versus the number of components. The proposed system achieves the highest accuracy, with only 12 principal components for input images. Also, this section provides the results of each classifier and the classifier combination utilizing the Dempster-Shafer theory of evidence. These results illustrate in Table 3 and 4 as well as Figure 13 [See *Annexure A: Figures & Tables*] in bar graph format. The classes are normal (benign) and abnormal (malignant), which are indicated as 0 (benign) and 1 (malignant), respectively. The class indicated by 2 identifies to uncertainty in image begins classified and consequently, "more testing required." This classification is invaluable when the case of false negatives is extremely high. Moreover, it may be the best option to warn an expert of uncertainty rather than of an unconvinced and possibly incorrect decision.

The outcomes are given in the configuration of a confusion matrix. The classification outcomes are prepared for five algorithms: k- nearest neighbor denoted as KNN, Naive Bayesian as indicated as Bayes, decision tree as indicated as D-tree, artificial neural network marked as ANN, support vector machine shown by SVM, and the Dempster-Shafer Theory. Class 0 represents the normal (benign); class 1 represents abnormal (malignant), and class 2 corresponds to

the uncertainty classification. The KNN algorithm offers the highest accuracy in categorizing classes related to class 0 (benign), where 30 images are correctly classified as benign, and 8 images are misclassified as malignant. The D-Tree algorithm offers the highest performance in the categorization related to class 1 since 174 images were correctly classified as malignant, while 14 images were misclassified as benign. Figure. 13 clarifies the diversity of categorization results between the classifiers. The worst classifier was Naive Bayes, which in both classes, only 21 and 160 images were correctly classified in class 0 and class 1, respectively. Every classifier is given with equal training and testing data. The evidence fusion procedure facilitates a more robust classification over multiple data sets. Figure.14 shows the combination rules of max, sum, majority voting, average probability, and product probability. The achievement of the fused suggested scheme is significantly better than the single classifiers. Table 5 demonstrates the DST confusion matrix, where there is no determined threshold for the conflict factor while combining the beliefs. This shows that no uncertainty is represented in class 2. In this case, 33 and 180 images are correctly classified as classes 0 and 1, respectively. The false-negative rate is when a classified image is shown as benign, where it is, in fact, malignant, and the false positive rate is when an image is classified as malignant when it is benign. A false positive can result in unnecessary therapy, and a false negative can result in an inaccurate diagnosis, which is particularly dangerous since illness has been ignored. These results' impact means improvement is necessary for the tool applied to combine the decision; this assists in reducing the numbers of false-positive and false-negative results. Table 6. indicates that four images from class 0 (benign) and five images from class 1 (Malignant) require further tests (uncertainty), meaning that the contradiction between the evidence attains the predetermined threshold and cannot produce a confident decision. This is possibly a helpful mechanism for the evaluation of uncertainty and reliability in brain tumour detection when it is undesirable to achieve a correct measurement from experiments. A significant feature of this theory is the combination of evidence obtained from various sources and the modeling of opposition between them. The achieved result demonstrates that the Dempster-Shafer Theory enhances the classification process.

Conclusion

In this paper, the brain tumor detection support system: a decision fusion framework is implemented. Different stages are constructed to build the system and validate the overall performance of the complete system. First of all, MRI and CT brain images were downloaded from two Web sources to create the datasets to be applied later to the classification task. The preprocessing stage is used to remove unwanted marks or labels that can interfere in post-processing, especially in the classification phase. The median filter algorithm implements to each image will remove artifacts from the picture. The segmentation stage is the next stage, where a region of interest is segmented from the given image. The thresholding technique is applied at this stage. Discrete wavelet transform and principal component analysis are used for feature extraction and reduction stages, respectively. This method also presents multi-classifiers, namely Naïve Bays, k-NN, SVM, ANN, and decision tree, approached from a different perspective because of each algorithm's diverse theoretical framework. The Multi-Classifiers combination was designed and implemented using the Dempster-Shafer theory. The multi-classifier system obtains extremely high sensitivity (89%) and specificity (97%) compared to the most significant individual classifier, which was the D-tree regarding specificity (86%) and the K-NN regarding sensitivity (85%). The DST outperforms the other techniques to combine the multi-decision with sensitivity 98%, specificity 94%, and overall accuracy up to 98%.

References

- [1]. Brady, A., Laoide, R. Ó., McCarthy, P., & McDermott, R. (2012). Discrepancy and error in radiology: concepts, causes and consequences. *The Ulster medical journal*, 81(1), 3.
- [2]. https://www.cihi.ca/en/financial_report_fy09-10_q1_en.pdf
- [3]. Ziaei, A., Yeganeh, H., Faez, K., & Sargolzaei, S. (2008, May). A novel approach for contrast enhancement in biomedical images based on histogram equalization. In *2008 International Conference on BioMedical Engineering and Informatics* (Vol. 1, pp. 855-858). IEEE...
- [4]. <http://www.stcatharinesstandard.ca/2015/06/22/hospital-reviewing-work->
- [5]. <http://news.nationalpost.com/news/canada/ontario-hospitals-waited-five-months-to-publicize-fears-about-possible-errors-in-reading-of-3500-mammograms-and-ct-scans>
- [6]. Ontario patients alerted to possible errors as 3,500 CT scans. [Online]. <http://www.ctvnews.ca/health/health-headlines/ontario-patients-alerted-to-possible-errors-as-3-500-ct-scans-mammograms-reviewed-1.1451352>
- [7]. Singh, A. (2015, February). Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM. In *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 98-102). IEEE.
- [8]. Vector machine. In *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science* (pp. 1-6). IEEE.

- [9]. Abdulrazzaq, M. M., Noah, S. A., & Fadhil, M. A. (2015, December). X-Ray medical image classification based on multi classifiers. In *2015 4th International Conference on Advanced Computer Science Applications and Technologies (ACSAT)* (pp. 218-223). IEEE...
 - [10]. Patil, P. G., & Karande, K. J. (2016). Brain tumor detection techniques: A Survey. *Int Res J Eng Technol (IRJET)*, 3(Issue), 10.
 - [11]. Chinnu, A. (2015). MRI brain tumor classification using SVM and histogram based image segmentation. *International Journal of Computer Science and Information Technologies*, 6(2), 1505-1508.
 - [12]. Nikam, P. B., & Shinde, V. D. (2013). MRI brain image classification and detection using distance classifier method in image processing. *International Journal of Engineering Research and Technology*, 2(6), 1980-5.
 - [13]. Deshmukh, R. J., & Khule, R. S. (2014). Brain tumor detection using artificial neural network fuzzy inference system (ANFIS). *International Journal of Computer Applications Technology and Research*, 3(3), 150-154..
 - [14]. Prasad, S., Bruce, L. M., & Ball, J. E. (2008, August). A multi-classifier and decision fusion framework for robust classification of mammographic masses. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3048-3051). IEEE.
 - [15]. Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
 - [16]. Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.*, 73(16), 5261-5267.
 - [17]. Sahoo, P. K., Soltani, S. A. K. C., & Wong, A. K. (1988). A survey of thresholding techniques. *Computer vision, graphics, and image processing*, 41(2), 233-260.
 - [18]. Thakur, M. L. (2009, July). Genomic biomarkers for molecular imaging: predicting the future. In *Seminars in nuclear medicine* (Vol. 39, No. 4, pp. 236-246). WB Saunders.
 - [19]. Hiremath, P. S., Shivashankar, S., & Pujari, J. (2006). Wavelet based features for color texture classification with application to CBIR. *International Journal of Computer Science and Network Security*, 6(9A), 124-133.
 - [20]. Bouchaffra, D., & Amira, A. (2008). Structural hidden Markov models for biometrics: Fusion of face and fingerprint. *Pattern Recognition*, 41(3), 852-867.
 - [21]. Kuncheva, L. I. (2014). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.
 - [22]. Gargiulo, F., Mazzariello, C., & Sansone, C. (2013). Multiple classifier systems: theory, applications and tools. In *Handbook on Neural Information Processing* (pp. 335-378). Springer, Berlin, Heidelberg.
 - [23]. Alexandre, L. A., Campilho, A. C., & Kamel, M. (2001). On combining classifiers using sum and product rules. *Pattern Recognition Letters*, 22(12), 1283-1289.
 - [24]. Ponti Jr, M. P. (2011, August). Combining classifiers: from the creation of ensembles to the decision fusion. In *2011 24th SIBGRAP Conference on Graphics, Patterns, and Images Tutorals* (pp. 1-10). IEEE.
 - [25]. Kuncheva, L. I. (2014). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.
 - [26]. Al-Ani, A., & Deriche, M. (2002). A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence. *Journal of Artificial Intelligence Research*, 17, 333-361.
 - [27]. Al-Ani, A., & Deriche, M. (2002). A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence. *Journal of Artificial Intelligence Research*, 17, 333-361.
-

Annexure A: Figures & Tables

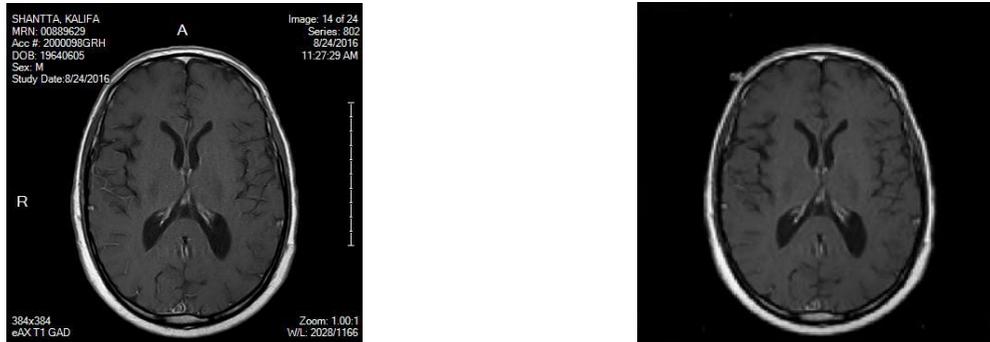


Fig.1. MRI before applied median filter and after.



Fig. 2. MRI before applied thresholding and after.

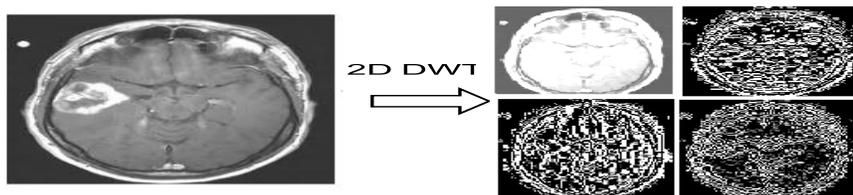


Fig. 3. Two level of DWT

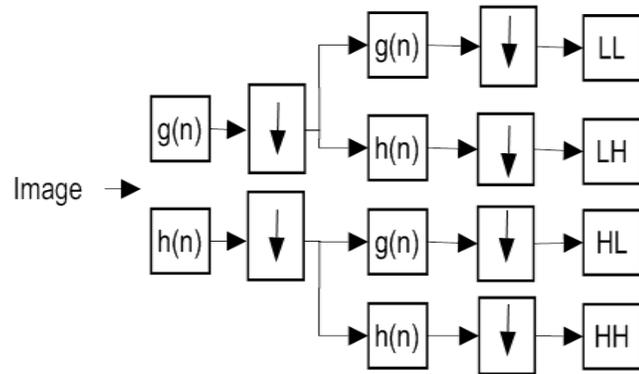


Fig. 4:DWT block diagram

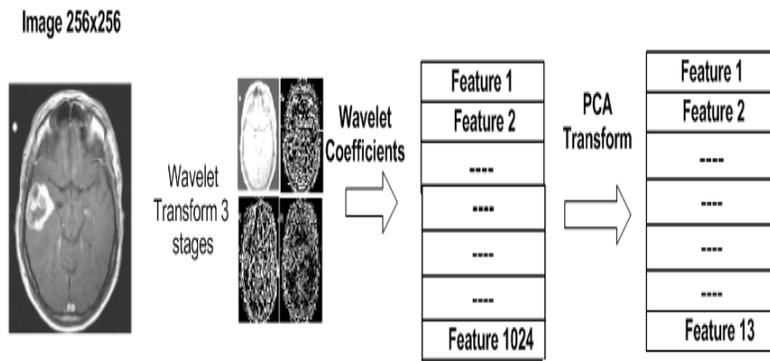


Fig. 5: PCA schematically

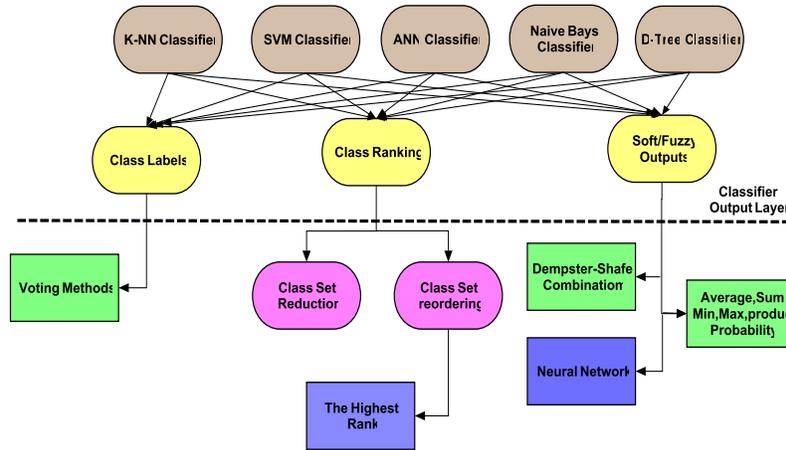


Figure 6 Fusion techniques category.

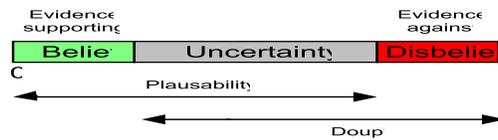


Fig. 7. Different Measurements over a Unit Interval.

Tabel. 1 Fusion of KNN& SVM classifiers.

		KNN classifier beliefs		
		$Bel B$	$Bel M$	U_{KNN}
SVM Classifier beliefs	$Bel B$	$Bel B$	$Bel \emptyset$	$U_{KNN} \times bel_{SVM(B)}$
	$Bel M$	$Bel \emptyset$	$Bel M$	$U_{KNN} \times bel_{SVM(M)}$
	U_{SVM}	$bel_{KNN(B)} \times U_{SVM}$	$bel_{KNN(M)} \times U_{SVM}$	$Bel U$

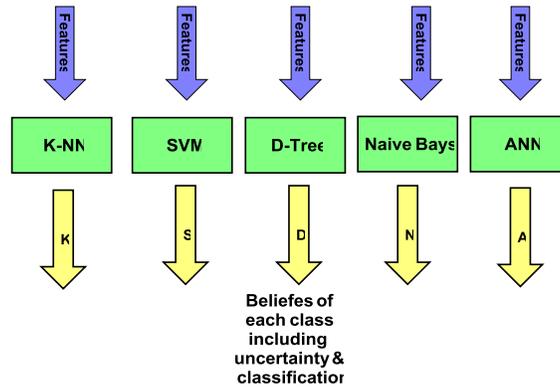


Figure.8 . Block Diagram of Individual Classifiers

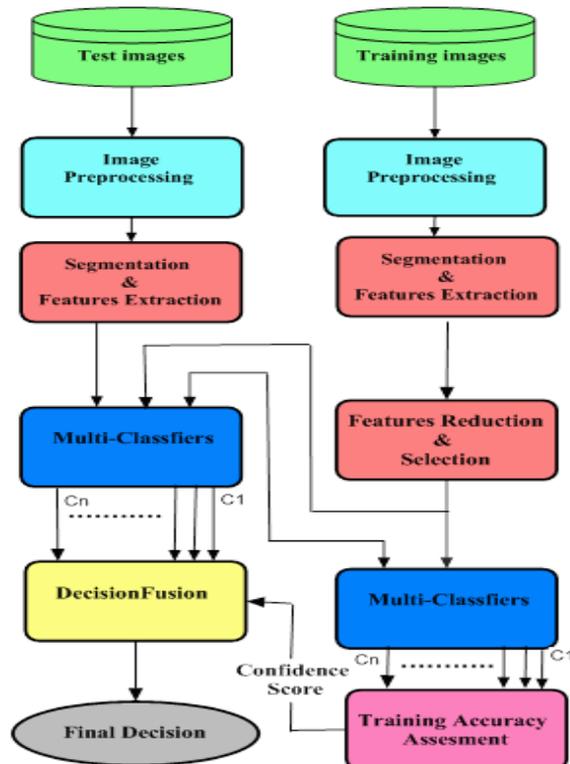


Fig. 9:Block diagram of proposed system

Table 2 Datasets.

Datasets		Normal	Abnormal
160		20	140
66		18	48
Total	226	Total 38	Total 188

		Predictor	
		C	1
Actua	C	TN	FF
	1	FN	TF

Fig.10. Confusion Matrix

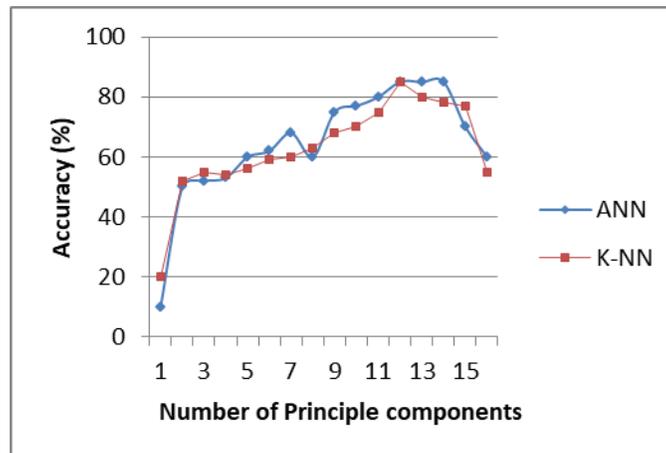


Fig.11. ANN and K-NN Classifiers performance vs Principle Component

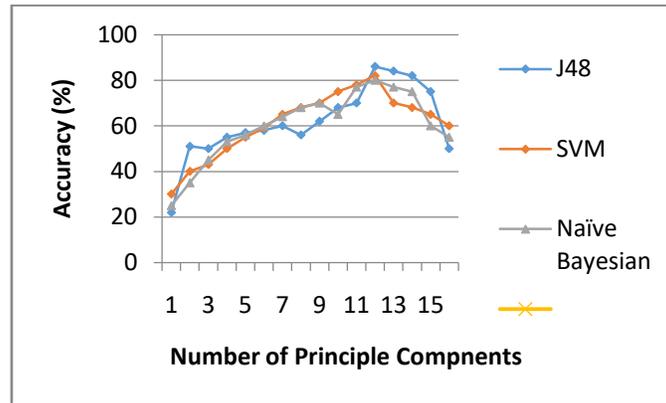


Fig.12. Naïve Bays, SVM, and J48 Performance vs No. of Principle Components

Table 3 Confusion Matrix of SVM, ANN, & KNN classifier

	ANN			SVM			KNN			
	0	1	2	0	1	2	0	1	2	
0	27	11	0	0	28	10	0	30	8	0
1	18	170	0	1	20	168	1	18	170	0

Table 4 Confusion Matrix of D-tree, Naïve Bays Classifier

	D- Tree			Naïve Bays		
	0	1	2	0	1	2
0	26	12	0	21	17	0
1	14	174	0	28	160	0

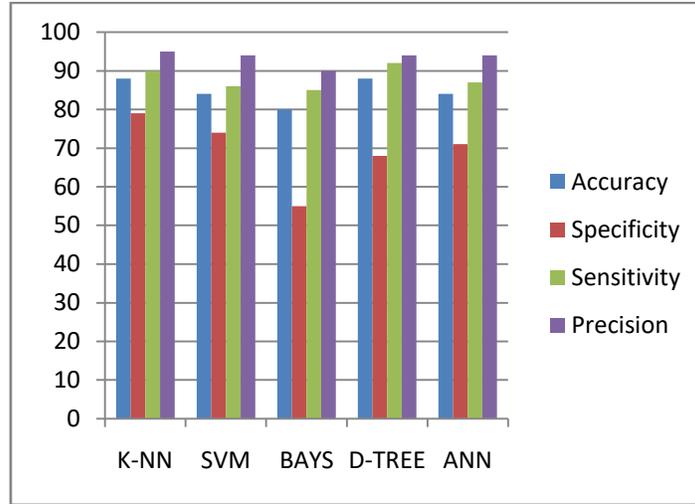


Fig. 13 .Comparison of Multi-Classifier.

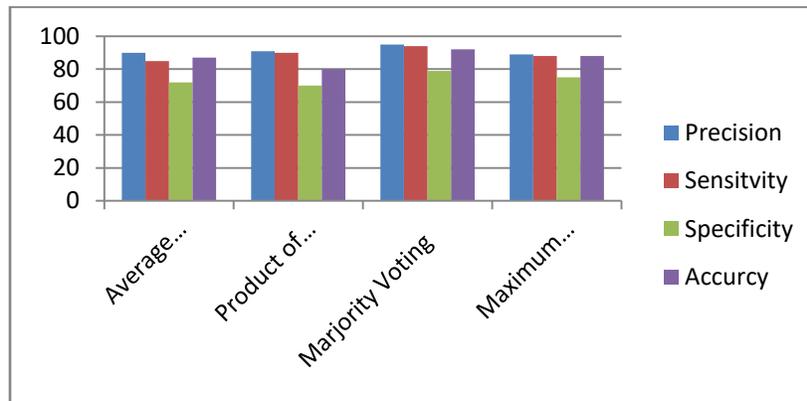


Figure 14 Comparison of several Multi-Classifier Combination Rules

Table 5 DST Confusion matrixes

	DST Fusion		
	0	1	2 Uncertainty (More Test Required)
0	33	5	0
1	8	180	0

Table.6 DST Confusion matrixes with setting k = 0.47 conflicting factor

	DST Fusion		
	0	1	2 Uncertainty (More Test Required)
0	32	2	4
1	4	179	5