

Bone Cancer Detection Using Feature Extraction with Classification Using K-Nearest Neighbor and Decision Tree Algorithm

Satheesh Kumar, B^{a,1} and Sathiyaprasad. B^b

^{a, b} *Research Scholar, Dept of CSE, Annamalai University, India*

Abstract. The malignant cells that cannot be controlled from spreading throughout the body is Cancer. Among which the cancer occurs in bone is their type. It is malignant disease occur in bone of human body where their growth cant be controlled from growing. This bone cancer is very critical of all the cancer types since the malignant cells are not identified at their earlier stage and it is the major challenge. Bone cancer is highly common for children and teenagers. For earlier detection of this cancer the correlation of medical imaging has been adapted with image processing and machine learning techniques where maximum accuracy can be obtained similarly even for bone cancer. This paper proposes the detection of bone cancer from the dataset taken from clinical dataset. Here the proposed design comprises of 2 phases in predicting the disorder with higher accuracy. The first stage is extracting the feature of segmented bone image using Gray-Level Co-occurrence Matrix (GLCM) method is applied to extract the features in terms of statistical texture-based and the second phase is classification of extracted feature using K-NN with decision tree algorithm. The simulation results show the enhanced classification results and extracted output with higher accuracy.

Keywords: Cancer, Bone cancer, X-ray, MRI, CT imaging, features, Gray-Level Co-occurrence Matrix (GLCM), K-NN.

1. Introduction

A tumour is an abnormal growth of new tissue and that can be formed in any of the organs in our body. There are many different kinds of cancer like lung cancer, brain cancer, and bone cancer. Nowadays bone cancer is considered to be one of the most dangerous and serious cancer in the world, with the smallest survival rate after the diagnosis. There are two types of bone cancer, noncancerous (benign) and cancerous (malignant) [1]. Obtaining an accurate result in bone cancer detection is very important in many imaging application. It mainly helps to plan for the treatment at the earlier stage and for the evaluation of the therapy. The early detection of bone cancer will decrease the mortality rate. To obtain more accurate results, we divided the whole process into three stages, image processing stage, image segmentation, feature extraction, and classification [2].

¹B. Satheesh Kumar, Research Scholar, Dept of CSE, Annamalai University, India.
E-mail: vbsatheesh30@gmail.com

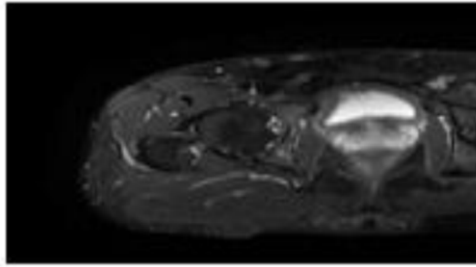


Figure1. MRI Image of the bone with tumour

In the kernel the hyper plane and optimal plane has been divided by the SVM or support vector machine classifier which is discriminative [3]. For feature based classification, KNN classifier is used in obtaining higher similarity and this technique is non-parametric classification [4]. The integration of neural network with fuzzy logic is ANFIS which is together known to be neuro fuzzy classifier. In this classifier, the features of fuzzy design have been obtained through the neural network [5]. The main objective of our proposed method is to have a fast and robust system for detecting the bone cancer in early stage and to obtain a more accurate result than many other existing techniques. Rest of the article has been continued as follows, section 2 discuss the related works which is done in field of bone cancer detection. Section 3 presents the detailed explanation on the proposed work in detecting the bone cancer. Experimental results and their discussion have been discussed in section 4. Final part of paper in section 5 gives the conclusion and the future work for this research.

2. Related Works

Kishore Kumar ReddyC et al. determined the technique in bone cancer detection based on their tumour size and the phases of growing cancer. For segmentation of bone MRI they have used seeded region growing algorithm [6]. KrupaliDMistry et al. explained the method in identifying the bone cancer by combining the feature extraction using clustering algorithm for input MRI images. The tumor cells has been identified after segmentation through mean pixel intensity where it is used in detecting enchondroma bone cancer [7]. MadhuriAvula et.al. Proposed the technique in detecting bone cancer by segmentation of image using k-means clustering algorithm where the tumor has been detected on basis of pixel intensity. The simulation results has been obtained from processing 400 bone MR images [8]. Eftekhar Hossain[9] has proposed an Detection & Classification of Tumor Cells from Bone MR Imagery Using Connected Component Analysis & Neural Network. The bone tumor can be detected by using connected component labeling algorithm. Akash Pandey et al [10] have done a paper on A Survey Paper on Calcaneus Bone Tumor Detection Using different Improved Canny Edge Detector. Here Computer Aided Diagnosing (CAD) is used to analyse Computed Tomography images.

3. Research methodology

The research design consists of phase namely pre-processing, feature extraction, training, and testing. Firstly the input image of bone MRI has been processed by filtering and this image which is filtered has segmented. Then this segmented image and the features namely statistical texture-based features of disorders are extracted from segmented image and the texture features are used to predict the category of the bone disorder using machine learning classifiers. The architecture for proposed extraction and classification technique has been given in Figure 2.

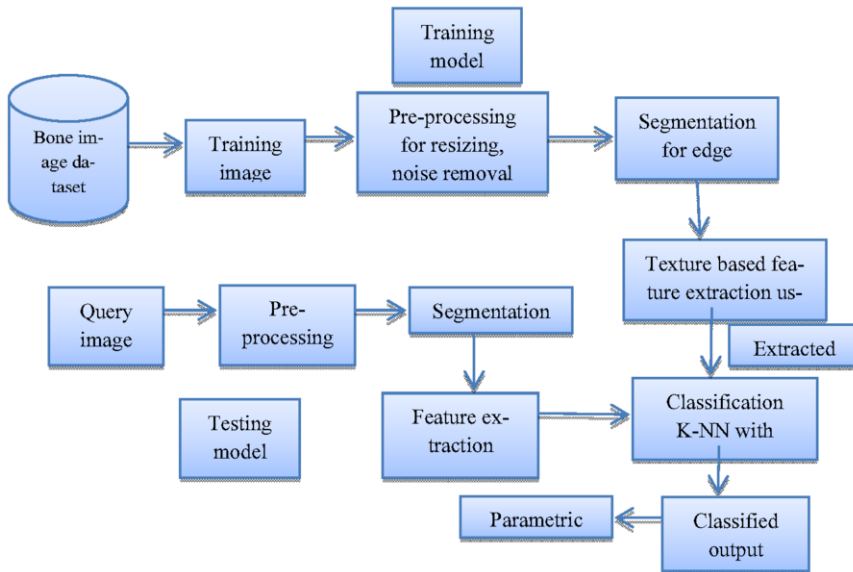


Figure2. Architecture for proposed technique

3.1 Feature extraction using GLCM

Image feature extraction is vital phase of PC based framework. Feature extraction gives certain parameters, based on which PC framework takes choice. After the segmentation is performed on lung area, the features can be acquired from it and the conclusion control can be intended to distinguish nodules in the lung. The whole component which are ascertained from the picture, pass on some data in regards to lung nodule. This data is exceptionally useful in identifying lung nodule as threatening or non-harmful. In this manner the features separated from the CT picture can be utilized as demonstrative pointers. The features that are utilized as a part of this investigation are Texture features utilizing Co- occurrence matrix representation. The GLCM is a second-order factual technique for the feature extraction based on texture. To start with, the picture is converted into an l-grey-level image and GLCM is produced through including incidences of intensity pairs among the existing and neighbor pixels for each scale and alignment. The standardized GLCM is computed in the equation beneath:

$$G(i, j) = \frac{N(i, j)}{\sum_{m=0}^{l-1} \sum_{n=0}^{l-1} N(m, n)} \quad (1)$$

Where i and j are grey values in the l - grey level picture. $N(i, j)$ is the co- occurrence relative recurrence frequency matrix by the equation underneath:

$$N(i, j) = \text{num} (\{(x_1, y_1), (x_2, y_2)\}) \\ |x_2 - x_1 = d \cos \theta, y_2 - y_1 = d \sin \theta, I(x_1, y_1) = i, I(x_2, y_2) = j \} \quad (2)$$

Where (x_1, y_1) and (x_2, y_2) are pixel positions, and $I(\cdot)$ is the grey level of the pixel. $\text{num} \cdot$ means the quantity of the pixel matches that fulfil the comparing conditions. In our work, 15 texture features are ascertained from the comparing GLCM.

3.2 Classification using KNN-decision tree

A non-parametric technique which is used for the problem in classification is done through KNN algorithm. The technique points the data by the neighbor for the output labels of classification. KNN is a very simple algorithm that stores all possible class cases, and works to classify new cases based on the functions of the distance (similarity measure) , KNN in the field of statistical estimation and pattern recognition. The class is classified by the most votes of its neighbors, with the case being assigned to the most popular class among its K nearest neighbors measured by the function of the distance, if the value of k is one, the class is simply assigned to the nearest neighbor class, and classification is easy. K -nearest neighbor algorithm steps:

1. Search for the nearest neighbor numbers that is denoted as K .
2. Calculate the training sample and query image distance.
3. On basis of minimum distance, it has been collected and evaluated by nearest neighbors.
4. From nearest neighbor, class Y has been collected.
5. As the number of prediction class is given for query image, the majority of nearest neighbor has been used.

3.3 Decision tree

The training-set feature extracted images are chosen to become familiar with the feature behaviour related to parameters as well to figure out data gaining for various feature sets. Features with high data gaining are chosen as Decision tree features.

The pseudo code for Decision-Tree algorithm is as follows:

- i. The input characteristic is given as the optimal feature by their root node of the tree.
- ii. The dataset has been classified for training into sub-classes.
- iii. The classified sub class has been carried out for every subclass where the data which has same value as input characteristic.
- iv. After all the above steps, step 1,2,3 has been iterated for every subclass until the node part in each branch in tree has been identified. Since this decision tree is used for classification, whole training data for input has been considered to be

root. The extreme evaluation of decision tree. When this is maximum, there will be problem of over-fitting. And this represented by size with the standard of nodes of the tree. When every node is used to be in classification of binary pattern, then their size has been maximum as $2d+1-1$, here d is denoted as the depth of the tree.

Steps for KNN-DA

Step 1: Choose K number of samples from the training set to generate initial population ($P1$).

Step 2: Calculate the distance between training sets in each chromosome and testing samples, as fitness value.

Step 3: Choose the chromosome with highest fitness value store it as Global maximum ($Gmax$), for iteration value 1 to L

- a. Perform reproduction
- b. Apply the crossover operator.
- c. Perform mutation and get the new population. ($P2$)
- d. Calculate the local maximum ($Lmax$).
- e. If $Gmax < Lmax$ then Assign $Gmax = Lmax$; $P1 = P2$;
- f. Repeat

Step 4:Output – the chromosome which obtains $Gmax$ has the optimum K -neighbours and the corresponding labels are the classification results.

4 Performance analysis

Contradicting of our algorithm is done with the existing algorithms in terms of various parametric metrics like accuracy, sensitivity specificity, $f1$ score, recall are chosen to evaluate the classification performance of the KNN classifier.

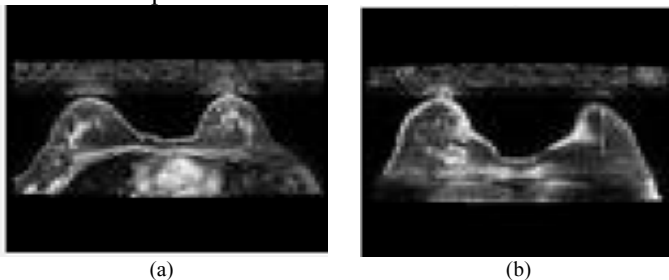


Figure 3. (a) Training input data (b) Testing input data

Figure3.(a) and (b) show that known data added with previously known decision values has been fed to classifiers. Training data classifies unknown images. A single set of 700 brain MRI images for training and also, a set of 60 images for testing are used.

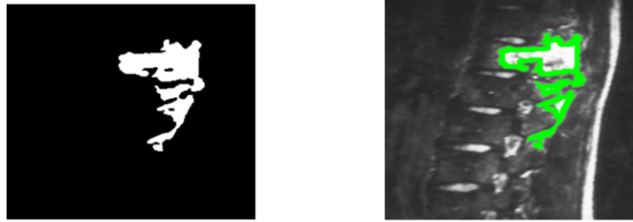


Figure. 4. (a) Extracted tumour from input image Figure.4. (b)-Detected tumour

Figure 4.(a) and (b) shows the extracted tumour feature and detected region of tumor from taken input bone image dataset. Table-1 gives overall performance comparison analysis statistically.

Table 1. Overall Performance Analysis

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)
3D-GLCM CNN	90.00	71.36	91.37	71.20
CNN with gene data	91.40	96.06	95.69	81.03
RSS-BW	97.52	98.60	96.38	83.45
KNN-decision tree	98.20	99.56	97.27	87.83

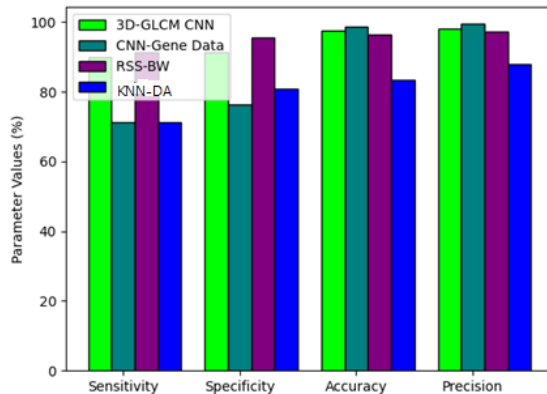


Figure 5. Overall comparative analysis of existing and proposed algorithm

The figure 5 compares the values achieved for the parameters. X axis and Y axis shows parameters considered for analysis and parameter values obtained in percentage respectively. The proposed algorithm achieves 98.2% of sensitivity, 99.56% of specificity, 97.27% of accuracy and 87.83% of precision. Finally, when compared to prevailing 3D-GLCM CNN, CNN with gene data, and RSS-BW proposed KNN-DA algorithm, shows better results.

5. Conclusion

The proposed technique has been done for detecting the tumor in bone by the extraction of the features and classification. This technique is used for identifying the presence and non-presence of tumor in input MRI which has been classified as benign or malignant class after this detection. Bone cancer cause unnatural growth of cells as to grow out of control in bone, which growth rather destroys normal bone tissue and start to spread in other body parts. Detecting cancer tissue is a vital issue for pathologists to know and recognize potential lesion tissue. This paper proposes the feature extraction of bone tumour GLCM and classification using KNN-DA. The accuracy obtained by the proposed technique is enhanced and improved. The aim of this work is to implement the design in differentiating malignant bone cancer from benign cells and this gives effects on clinical analysis. The future work can be developed with classification for various types of bone tumor like chondroma, Ewing sarcoma, and chondrosarcoma etc.

References

- [1] Yang, A., Yang, X., Wu, W., Liu, H., & Zhuansun, Y. Research on feature extraction of tumor image based on convolutional neural network, 7, 24204-24213 (2019).
- [2] Goswami, S., & Bhaiya, L. K. P. A hybrid neuro-fuzzy approach for brain abnormality detection using GLCM based feature extraction, In 2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA) 1-7 (2013).
- [3] Chen, X., Li, J., Zhang, Y., Lu, Y., & Liu, S. Automatic feature extraction in X-ray image based on deep learning approach for determination of bone age, Future Generation Computer Systems, 110, 795-801 (2020).
- [4] Nakajima, K., Nakajima, Y., Horikoshi, H., Ueno, M., Wakabayashi, H., Shiga, T., & Edenbrandt, L., Enhanced diagnostic accuracy for quantitative bone scan using an artificial neural network system, a Japanese multi-center database project. EJNMMI research, 3, 1-9 (2013).
- [5] Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier, Neural Computing and Applications, 1-14 (2020).
- [6] Reddy, C. K. K., Anisha, P. R., & Raju, G. V. S. A novel approach for detecting the tumor size and bone cancer stage using region growing algorithm, In 2015 International Conference on Computational Intelligence and Communication Networks (CICN) 228-233 (2015).
- [7] Mistry, K. D., & Talati, B. J. Integrated approach for bone tumor detection from mri scan imagery, In 2016 International Conference on Signal and Information Processing (IconSIP) 1-5 (2016).
- [8] Avula M, Lakkakula N.P, and Raja M.P. Bone cancer detection from mri scan imagery using mean pixel intensity, in Modelling Symposium (AMS), 2014 8th Asia. IEEE, 141-146 (2014).
- [9] Hossain .E and Rahaman.M.A. Detection & Classification of Tumor Cells from Bone MR Imagery Using Connected Component Analysis & Neural Network, 2018; International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE) 1-4 (2018).
- [10] Pandey and S. K. Shrivastava. A Survey Paper on Calcaneus Bone Tumor Detection Using different Improved Canny Edge Detector, 2018 IEEE international conference on system, computation, automation and networking (icscan) 1-5 (2018).