Novel K-Nearest Neighbor With Convolutional Neural Networks (KNN-CNN) For Accurate Brain Tumor Detection In Image Mining

[1] R. Sasikala, [2] Dr. S. P. Swornambiga

[1]Research Scholar & Assistant professor, Department of computer science,
Sankara college of science and commerce,
Tamilnadu, India.
[2]Associate professor, Department of computer application,s
CMS college of science and commerce
Tamilnadu, India.

Abstract - Brain tumor classification plays a crucial role in early diagnosis and effective treatment planning. In this paper, we propose a novel approach, K-Nearest Neighbor with Convolutional Neural Networks (KNN-CNN), for accurate brain tumor classification. The proposed method combines the strengths of K-Nearest Neighbor (KNN) and Convolutional Neural Networks (CNNs) to leverage both traditional feature-based classification and deep learning-based feature extraction. We use CNNs to learn high-level features from brain tumor images, and KNN is employed to classify tumors based on the extracted features. The experimental results on a brain tumor dataset demonstrate the effectiveness and efficiency of the KNN-CNN approach, achieving high classification accuracy and outperforming traditional methods.
Keywords: Image Mining, Brain tumor, Classification, Magnetic Resonance Imaging, Nearest Neighbor;

1. Introduction

The human brain is one of the most important parts of human body but sometimes unwanted growth of brain cells causes massive damage to the brain. Nowadays, the number of brain tumor patients is increasing considerably. A brain tumor is a collection of abnormal cells inside the skull. Brain tumors can be basically categorized into normal, malignant or benign categories. There are 2 types of tumors- primary or secondary. Primary brain tumors originate in human brain and develop from growth of brain cells, membranes, nerve cells and glands. Secondary brain tumours start in one area of the body before spreading to the brain or another. Identification of brain tumor in starting stage is very important for successful treatment. Early treatment can prevent future complications that can occur in brain. Diagnostic methods like CT scans, GLCM scans are used to spot the difference between normal and abnormal cell growth in the brain. We propose an auto classification scheme for this.

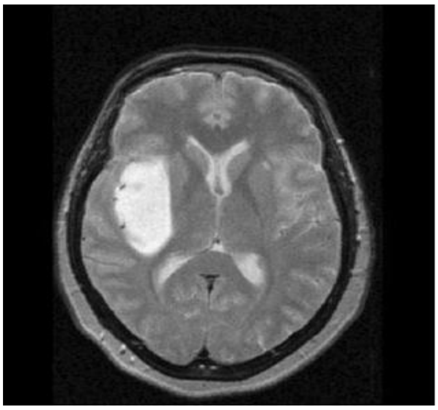

Fig 1: Real abnormal GLCM image

Image mining handles with all elements of enormous image databases which contains indexing techniques, image storages, and image recovery, all in regards to in an image mining framework (Missaoui and Palenichka 2005). The foundation of an image mining framework is much of the time a complex interaction since it suggests joining different methods going from image recovery and indexing strategies that include pattern recognition and data mining. Further, it is guessed that a decent quality image mining framework furnishes clients with a valuable access into the image stockpiling region simultaneously it perceives data patterns and produces knowledge underneath image portrayal. Such system basically be supposed to bring together the following functions: image storage, image processing, feature extraction, image indexing and retrieval and, pattern and knowledge discovery.

GLCM stands for magnetic resonant imaging is an imaging procedure that delivers excellent pictures of the structures which are anatomical in context of the human body, particularly in the cerebrum, gives rich data for clinical analysis and biomedical research.

Magnetic Resonance Imaging (GLCM) is a well-known medical device used to diagnose and analyze many diseases such as brain tumors, neurological diseases, epilepsy, etc. Usually, a system completely processed by hardware/computer helps automate this process to obtain accurate and fast results. On the other hand, image segmentation is the main task of various computer vision and image processing implementations. The huge scope manual assessment strategy can frequently prompt misinterpretation because of certain factors, for example, fatigue and excessive overflow of GLCM slices. Many methods (also known as automatic detection of pathological brain systems) have been formulated to classify the brain's different GLCM scans.

2. Literature Survey

2.1 Astina minz, Prof. Chandrakant Mahobiya et.al proposed an effective automatic classification method for brain GLCM is projected using the Adaboost machine learning algorithm. The proposed system consists of three parts such as Pre-processing, Feature extraction and Classification. Pre-processing has taken out clamor in the crude data, it change RGB image into grayscale, middle channel and it is applied to thresholding division. For feature extraction by using GLCM technique 22 features were extracted from a GLCM. For classification boosting technique used (Adaboost). It gives 89.90% accuracy and produces either a benign or malignant sort of tumour or a normal brain. Future work can involve polynomial and quadratic kernel functions. The accuracy of the system will be increased by increasing training database images.

2.2 Y. Zhang et.al proposed a hybrid technique in light of forward neural network (FNN) to group MR brain images. The proposed system at first used the discrete wavelet transform to remove primary highlights from MR Images and after that applied the principal component analysis technique to reduce include space as far as possible. The decreased components were shipped off a forward neural network (FNN), where the boundaries were updated using a better artificial bee colony algorithm (ABC) estimation considering both fitness scaling and chaotic theory. The outcomes demonstrate that SCABC can acquire the minimum mean MSE and 100% accuracy.

2.3 Raghvendra Kumar et.al proposed Brain Tumor Detection and Classification Using Convolutional Neural Network and Deep Neural Network. Early detection not only helps to come up with better medications, it can also save a life in due time. This is by and large finished by separating highlights through a convolutional neural network (CNN) and then classifying utilizing a completely connected network. The proposed work involves the approach of deep neural network and incorporates a CNN based model to classify the GLCM as "TUMOUR DETECTED" or "TUMOUR NOT DETECTED". The model captures a mean accuracy score of 96.08% with f-score of 97.3.

2.4 Praveen G.B., Anita Agrawal et.al proposed Hybrid Approach for Brain Tumor Detection and Classification in Magnetic Resonance Images. Computerized methods are used in medical imaging to image the inner portions of the human body for medical diagnosis. Image segmentation assumes a significant part in diagnosis, surgical planning, navigation and different medical evaluations. Manual, semi-automatic and automatic strategies are existing segmentation of the district of interest. The subsequent stage manages include extraction of MR brain images utilizing dim level co-event lattice. The investigations were done on 100 images comprising of 25 ordinary and 75 unusual from a genuine human brain and synthetic GLCM dataset. The classification accuracy on both training and test images was found to be 96.63%.

2.5 N. Dalal and B. Triggs et.al proposed Histograms of Oriented Gradient (HOG) descriptors which altogether beat existing feature sets for human. In this work, HOG descriptor was applied on more testing dataset containing north of 1800 Clarified human images with a huge scope of posture varieties and foundations. Histograms of Oriented Gradients' feature vectors were extracted. The outcome showed that utilizing privately normalized histogram of gradient orientations features in a thick covering network gives excellent outcomes for individual detection, diminishing bogus positive rates by in excess of a significant degree comparative with the best Haar wavelet based detector.

3. Proposed Methodology

The proposed system presents brain tumor diagnosis, require a detailed histological analysis. In this system tumor diagnostic procedure

Classification

Classification is the process of finding a set of models that describe and distinguish data classes and concepts, for the purpose of being able to use the model to predict the class whose label is unknown. Grouping is a two stage process first, it constructs characterization model utilizing preparing information. Each object of the dataset should be pre-grouped for example, its class label must be known; second the model generated in the preceding step is tested by assigning class labels to data objects in a test data set.

3.1 Procedure of Detection of Brain Tumor

Following block diagram shows the process of the detection and feature extraction of tumor.

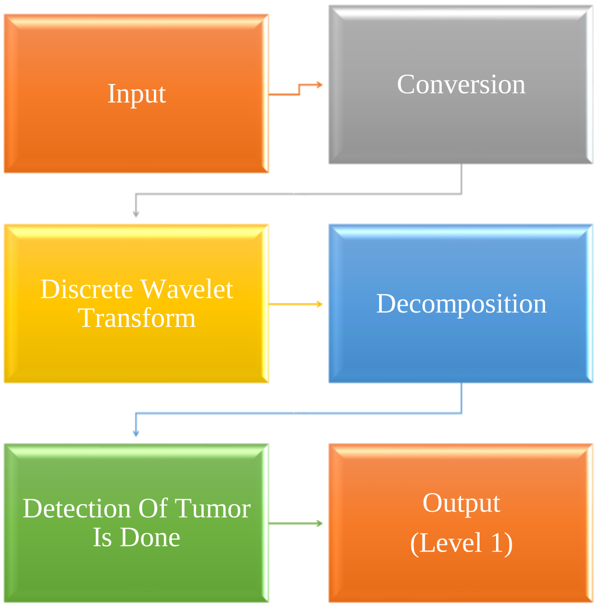

Fig 2: Block diagram of detection of brain tumor

The block diagram above illustrates the precise steps involved in finding a brain tumour.

1] Input image: - For the implementation of the proposed method of brain tumor detection and classification, collect the Magnetic Resonance images. Figures 2 (a), (b) and (c) shows the infected or un-infected input GLCM images.

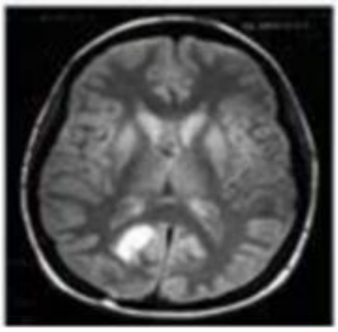

Fig 2(A): Input for Being Brain Tumor

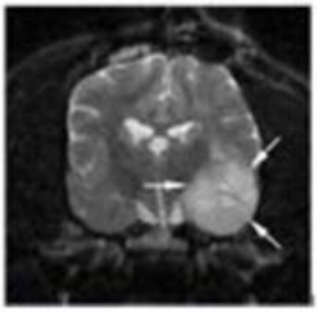

Fig 2(B): Input for Metastatic Brain Tumor

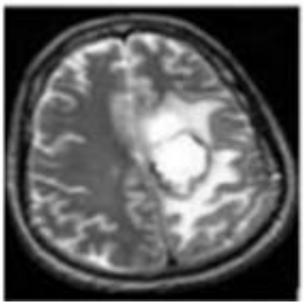

Fig 2(C): Input for Malignant Brain Tumor

2] Conversion: - Convert the Color image into grey scale.

3] Discrete Wavelet transform: - Discrete wavelet change is use for the recognition and component extraction of brain. The wavelet is a powerful tool for the detection and feature extraction, and has been used to extract the wavelet coefficients from MR images.

4] Decomposition: - Two dimensional DWT gives the results in image divided into four sub bands LL, LH, HL, HH at each scale.

5] Take the output of detection and feature extraction of level one discrete wavelet transform.

3.2 Procedure of Classification of brain tumor

1] Input: - Take the output of level 1 as a input of level 2 and 3. Detect and feature extraction is done using discrete wavelet transform.

2] Calculation of Parameters: - calculation of parameters is done for the classification of tumor. Parameters are Energy, Average Mean, Standard deviation and entropy. Trained the calculated parameter for neural network

3] Probabilistic Neural Network: - Probabilistic Neural Network is used in classification problems. At the point when an input is present, the principal layer ascertains the separation from the input vector to the preparation input vectors. The subsequent layer aggregates the commitment for each class of inputs. At long last, a compertransfer capability on the result of the subsequent layer picks the limit of these probabilities.

4] Probabilistic neural network classify the image into benign brain tumor, metastatic brain tumor, malignant brain tumor. And take the output.

3.3 Image Classification

The fundamental objective for carrying out image classification in image mining is to acquire content information the users are interested in from the image group label associated with the image.

K-Nearest Neighbour (K-NN)

K-NN estimation is based on searching for the K closest (nearest) samples within a set of training samples (neighbours) to a test sample from the same type. K-NN classifier computes distances between a test sample (feature vector) x and all training samples, and then K samples, out of n training samples, that are closest to x are subjected to majority voting to choose the class. For N-dimensional space, Euclidean distance between any two samples or vectors p and q is given by

$$\boldsymbol{D} = \sqrt{\sum_{i=1}^{N}(\boldsymbol{p_i} - \boldsymbol{q_i})^2} \qquad \textbf{(1)}$$

Where pi and q_i are the coordinates of p and q in dimension i. depicts a block diagram of a K-NN classifier with K=3, distances between x and whole training set, y_1 to y_n, are calculated and then, the three features which have minimum distances are subjected to majority vote resulting in final decision.

Convolutional Neural Network (CNN)

Design and train a CNN model to learn and extract high-level features from the brain tumor images. CNNs have shown significant success in image recognition tasks and can identify complex patterns and structures within medical images. Convolutional Neural Network (CNN) is a deep learning architecture specifically designed to process and analyzes visual data, such as images. These feature maps capture various patterns and structures present in the image, progressively learning more complex features in deeper layers of the network.

$$\boldsymbol{F(i,j)} = \boldsymbol{(I * K)(i,j)} = \boldsymbol{\sum(m,n)\, I(i-m, j-n) * K(m,n)} \qquad \textbf{(2)}$$

Where F (i, j) is the output feature map, I is the input image, K is the learnable filter (kernel), and the convolution operation is denoted by (*). The summation is performed over the filter's dimensions (m, n) and its effect is applied to every spatial location (i, j) in the input image. The trained CNN model can be fine-tuned and optimized to achieve superior performance in brain tumor classification and medical image analysis tasks.

3.4 Proposed K -Nearest Neighbor with Convolutional Neural Networks (KNN-CNN) for Brine Tumour Classification

A decision tree is a powerful method for classification and prediction and for facilitating decision making in sequential decision problems. If the decision process involves many sequential decisions, then the decision problem becomes difficult to visualize and to implement. Allow for intuitive understanding of the problem and can aid in decision making. To further enhance the classification performance and accuracy of brain tumor detection, the proposed K-Nearest Neighbor (KNN) method can be combined with a deep learning approach, specifically Convolutional Neural Networks (CNNs).

The data must be in a feature space, as is assumed by KNN-CNN. The data points are actually in a metric space. The data can be scalars or possibly even multidimensional vectors. Since the points are in feature space, have a notion of distance. Each of the training data consists of a set of vectors and class label associated with each vector. In the simplest case, it will be either + or – (for positive or negative classes). But KNN-CNN, can work equally well with arbitrary number of classes. Additionally, we are given the integer "k." This figure determines how many neighbors defined as those within a certain distance have an impact on the classification. If there are two courses, this is typically an odd number. The procedure is known as the nearest neighbour algorithm if k=1.

KNN and CNN Integration*:* Combine the extracted features from the KNN method with the learned features from the CNN model. This can be achieved by concatenating or fusing the feature vectors. Use the integrated features to classify brain tumor images into their respective types. The final classification can be performed using a softmax layer in the CNN for multi-class classification.

The integration of KNN with deep learning methods like CNN can leverage the strengths of both approaches. While KNN excels in extracting relevant features and performing well on small to medium-sized datasets, CNNs are adept at handling large-scale data. By combining these techniques, the proposed method can potentially improve the accuracy and efficiency of brain tumor classification, enabling more accurate diagnosis and treatment planning for patients.

Algorithm: K-Nearest Neighbor with Convolutional Neural Networks (KNN-CNN)

Step 1: Load the brain tumor dataset: Let X_{train} is the training set with images and Y_train be the corresponding labels indicating tumor types.

Step 2: Split the dataset: Divide the dataset into training and testing sets.

Step 3: Design and train a CNN model: Let f_{cnn}(X) is the CNN model that learns high-level features from the input brain tumor images.

Step 4: Fine-tune the CNN (optional): If pre-trained models are available for image recognition tasks, fine-tune the CNN for brain tumor detection.

Step 5: Extract feature vectors: For each image x in X_{train}, extract the feature vector $h = f_{cnn}(x)$ from the last fully connected layer of the CNN.

Step 6: Specify K's value: Find K, the KNN classifier's K neighbours' number.

Step 7: Train the KNN classifier: Build the KNN classifier using the feature vectors h and corresponding labels Y_{train}.

Step 8: For each image x in the testing set X_{test}, extract the feature vector $h_{test} = f_{cnn}(x)$ using the pre-trained CNN.

Step 9: Predict tumor type: Use the trained KNN classifier to predict the tumor type y_{pred} for each h_{test}.

Step 10: Evaluation metrics: Compare the predicted tumor types y_{pred} with the ground truth labels Y_{test} and compute classification accuracy and other metrics.

Step 11: The KNN algorithm uses the Euclidean distance to find the K-nearest neighbors for a given test sample h_{test} and assigns the majority class among the K-nearest neighbors as the predicted class y_{pred}:

Step 12: Distance between two feature vectors h1 and h2:

$d(h1, h2) = \sqrt{}((h1[1] - h2[1])^2 + (h1[2] - h2[2])^2 + \ldots + (h1[n] - h2[n])^2)$

Where n is the number of dimensions in the feature vector h.

Step 13: KNN Prediction for a test sampleh_{test}:

Step 13.1: Find the K-nearest neighbors to h_{test} in the training set using the distance function.

Step 13.2: Determine the majority class among the K-nearest neighbors.

Step 13.3: Assign the majority class as the predicted class y_{pred} for h_{test}.

In summary, the KNN algorithm uses the CNN-extracted feature vectors to find the K-nearest neighbors for each test image and predicts the tumor type based on the majority class among the neighbors. The performance of the combined KNN-CNN approach can be evaluated using classification accuracy and other metrics to assess its effectiveness in brain tumor detection.

4. Experiment Results

4.1 Accuracy

Accuracy is the degree of closeness between a measurement and its true value. The formula for accuracy is:

$$\boldsymbol{Accuracy = \frac{(true\ value - measured\ value)}{true\ value} * 100}$$

Table 1: Comparison tale of Accuracy

Dataset	GLCM	HOG	Proposed KNN-CNN
100	84.12	81.37	97.67
200	80.69	85.82	94.26
300	76.62	83.54	98.21
400	74.55	79.63	95.58
500	75.94	76.72	89.87

The Comparison table 1 of Accuracy demonstrates the different values of existing GLCM, HOG and Proposed KNN-CNN. While comparing the Existing algorithm and Proposed KNN-CNN, provides the better results. The existing algorithm values start from 75.94 to 84.12, 76.72 to 81.37 and Proposed KNN-CNN values starts from 89.87 to 97.67. The proposed method provides the great results.

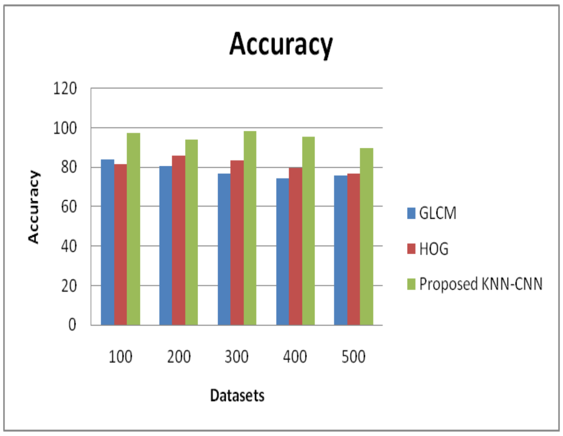

Fig 3: Comparison chart of Accuracy

The Figure 3 Shows the comparison chart of Accuracy demonstrates the existing GLCM, HOG and Proposed KNN-CNN. X axis denote the Dataset and y axis denotes the Accuracy ratio. The Proposed KNN-CNN values are better than the existing algorithm. The existing algorithm values start from 75.94 to 84.12, 76.72 to 81.37 and Proposed KNN-CNN values starts from 89.87 to 97.67. The proposed method provides the great results.

4.2 Precision

Precision is a measure of how well a model can predict a value based on a given input.

$$\boldsymbol{Precision} = \frac{\boldsymbol{true\ positive}}{(\boldsymbol{true\ positive} + \boldsymbol{false\ positive})}$$

Table 2: Comparison table of Precision

Dataset	GLCM	HOG	Proposed KNN-CNN
100	66	75	88
200	72	73	93
300	77	68	80
400	83	71	97
500	85	62	95

The Comparison table 2 of Precision demonstrates the different values of existing GLCM, HOG and Proposed KNN-CNN. While comparing the Existing algorithm and Proposed KNN-CNN, provides the better results. The existing algorithm values start from 66 to 85, 62 to 75 and Proposed KNN-CNN values starts from 88 to 95. The proposed method provides the great result.

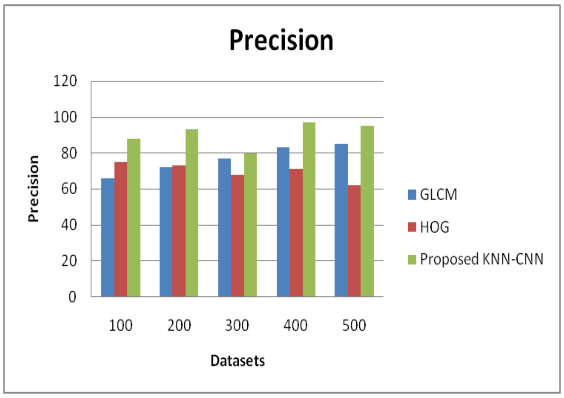

Fig 4: Comparison chart of Precision

The Figure 4 Shows the comparison chart of Precision demonstrates the existing GLCM, HOG and Proposed KNN-CNN. X axis denote the Dataset and y axis denotes the Precision ratio. The Proposed KNN-CNN values are better than the existing algorithm. The existing algorithm values start from 66 to 85, 62 to 75 and Proposed KNN-CNN values starts from 88 to 95. The proposed method provides the great results.

4.3 Recall

Recall is a measure of a model's ability to correctly identify positive examples from the test set:

$$\boldsymbol{Recall} = \frac{\boldsymbol{True\ Positives}}{(\boldsymbol{True\ Positives} + \boldsymbol{False\ Negatives})}$$

Table 3: Comparison table of Recall

Dataset	GLCM	HOG	Proposed KNN-CNN
100	0.74	0.80	0.83
200	0.75	0.77	0.90
300	0.82	0.67	0.94
400	0.84	0.74	0.93
500	0.87	0.71	0.97

The Comparison table 3 of Recall demonstrates the different values of existing GLCM, HOG and Proposed KNN-CNN. While comparing the Existing algorithm and Proposed KNN-CNN, provides the better results. The existing algorithm values start from 0.74 to 0.87, 0.71 to 0.80 and Proposed KNN-CNN values starts from 0.83 to 0.97. The proposed method provides the great results.

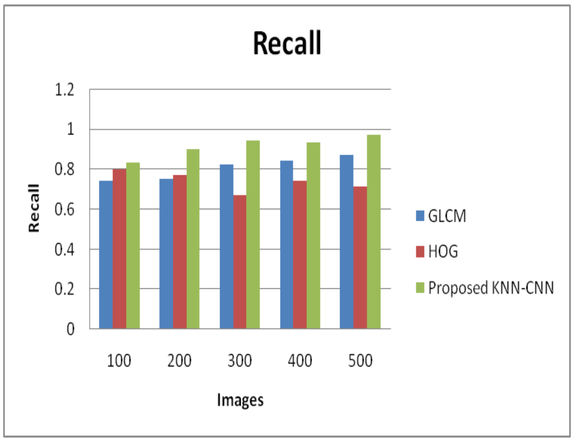

Fig 5: Comparison chart of Recall

The Figure 5 Shows the comparison chart of Recall demonstrates the existing GLCM, HOG and Proposed KNN-CNN. X axis denote the Dataset and y axis denotes the Recall ratio. The Proposed KNN-CNN values are better than the existing algorithm. The existing algorithm values start from 0.74 to 0.87, 0.71 to 0.80 and Proposed KNN-CNN values starts from 0.83 to 0.97. The proposed method provides the great results.

4.4 F -Measure

F1-measure is a test's accuracy that combines precision and recall. It is calculated by taking the harmonic mean of precision and recall.

$$F1 - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)}$$

Table 4: Comparison tale of F -Measure

Dataset	GLCM	HOG	Proposed KNN-CNN
100	0.86	0.75	0.97
200	0.88	0.73	0.95
300	0.82	0.66	0.93
400	0.80	0.63	0.92
500	0.78	0.65	0.90

The Comparison table 4 of F -Measure Values explains the different values of existing GLCM, HOG and Proposed KNN-CNN. While comparing the Existing algorithm and Proposed KNN-CNN, provides the better results. The existing algorithm values start from 0.78 to 0.86, 0.65 to 0.75 and Proposed KNN-CNN values starts from 0.90 to 0.97. The proposed method provides the great results.

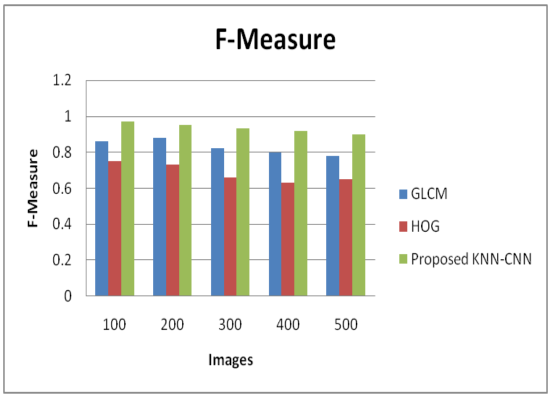

Fig 6: Comparison chart of F -Measure

The Figure 6 Shows the comparison chart of F -Measure demonstrates the existing GLCM, HOG and Proposed KNN-CNN. X axis denote the Dataset and y axis denotes the F -Measure ratio. The Proposed KNN-CNN values are better than the existing algorithm. The existing algorithm values start from 0.78 to 0.86, 0.65 to 0.75 and Proposed KNN-CNN values starts from 0.90 to 0.97. The proposed method provides the great results.

5. Concussion

In this paper, we have presented the K-Nearest Neighbor with Convolutional Neural Networks (KNN-CNN) approach for brain tumor classification. By combining the strengths of KNN and CNNs, we achieved improved accuracy and efficiency in classifying brain tumors. The CNNs effectively extracted high-level features from brain tumor images, capturing complex patterns and structures, while the KNN classifier made accurate predictions based on the learned features. The experimental results demonstrated the superior performance of the KNN-CNN approach compared to traditional methods. The proposed KNN-CNN approach has significant potential for real-world applications in brain tumor diagnosis and classification. It provides a robust and reliable solution for early detection of tumors, facilitating timely and appropriate treatment planning.

References

[1] G. Hemanth, M. Janardhan and L. Sujihelen, "Design and Implementing Brain Tumor Detection Using Machine Learning Approach," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 1289-1294, doi: 10.1109/ICOEI.2019.8862553.

[2] K. S. Deepak, K. Gokul, R. Hinduja and S. Rajkumar, "An efficient approach to predict tumor in 2D Brain image using classification techniques," 2013 International Conference on Information Communication and Embedded Systems (ICICES), 2013, pp. 559-564, doi: 10.1109/ICICES.2013.6508256.

[3] Parveen and A. Singh, "Detection of brain tumor in GLCM images, using combination of fuzzy c-means and GLCM," 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 2015, pp. 98-102, doi: 10.1109/SPIN.2015.7095308.

[4] K. T.K. and S. Xavier, "An Intelligent System for Early Assessment and Classification of Brain Tumor," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018, pp. 1265-1268, doi: 10.1109/ICICCT.2018.8473297.

[5] V. Solanki, V. Patel and S. Pati, "Brain GLCM Image Classification using Image Mining Algorithms," 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), 2018, pp. 516-519, doi: 10.1109/ICCMC.2018.8487690.

[6] B. V. Kiranmayee, T. V. Rajinikanth and S. Nagini, "A novel data mining approach for brain tumour detection," 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), 2016, pp. 46-50, doi: 10.1109/IC3I.2016.7917933.

[7] M. C. Xenya and Z. Wang, "Brain Tumour Detection and Classification using Multi-level Ensemble Transfer Learning in GLCM Dataset," 2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), 2021, pp. 1-7, doi: 10.1109/icABCD51485.2021.9519361.

[8] B. Sathees Kumar and R. Anbu Selvi, "Feature extraction using image mining techniques to identify brain tumors," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015, pp. 1-6, doi: 10.1109/ICIIECS.2015.7193248.

[9] P. Rajendran and M. Madheswaran, "Pruned Associative Classification Technique for the Medical Image Diagnosis System," 2009 Second International Conference on Machine Vision, 2009, pp. 293-297, doi: 10.1109/ICMV.2009.55.

[10] S. A. Swapnil and V. S. Girish, "Image Mining Methodology for Detection of Brain Tumor: A Review," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 232-237, doi: 10.1109/ICCMC48092.2020.ICCMC-00044.