Volume 66, Issue 3, 2022

Journal of Scientific Research

of

The Banaras Hindu University

Hybrid Feature Extraction Based Ensemble Classification Model to Diagnose Oral Carcinoma Using Histopathological Images

Rachit Kumar Gupta*[1], Jatinder Manhas[2], and Mandeep Kour[3]

*[1]Department of Computer Science and IT, University of Jammu, 180006, Jammu, India, gupta.rachit1990@gmail.com
[2]Department of Computer Science and IT, Bhaderwah Campus, University of Jammu, 180006, Jammu, India, manhas.jatinder@gmail.com
[3]Indira Gandhi Govt. Dental College and Hospital, Jammu, India, dr_mandeep_kaur@yahoo.com

Abstract: **Detection and classification of cancerous tissue from histopathologic images is quite a challenging task for pathologists and computer assisted medical diagnosis systems because of the complexity of the histopathology image. For a good diagnostic system, feature extraction from the medical images plays a crucial role for better classification of images. Using inappropriate or redundant features leads to poor classification results because classification algorithm learns a lot of unimportant information from the images. We propose hybrid feature extractor using different feature extraction algorithms that can extract various types of features from histopathological image. For this study, feature fused Convolution Neural Network, Gray Level Cooccurrence Matrix, and Local Binary Pattern algorithms are used. The texture and deep features obtained from these methods are used as input vector to classifiers: Support Vector Machine, K-Nearest Neighbor, Naïve Bayes and Boosted Tree. Prediction results of these classifiers are combined using soft majority voting algorithm to predict final output. Proposed method achieved an accuracy of 98.71%, which is quite high as compared to previous similar research works. Proposed method was capable of identifying most of cancerous histopathology images. The combination of deep and textural features can be potentially used for creating computer assisted medical imaging diagnosis system that can detect cancer from histopathology images timely and accurately.**

Index Terms: **Oral cancer, histopathology, convolution networks, feature extraction, fused features, texture features, classification.**

I. Introduction

Oral cancer is the sixth biggest cancer in the world and the second most threatening cancer in India. According to a report oral cancer contributes to about 0.3 million deaths annually. Due to such high death rate researchers and doctors have identified some high risk factors that co-relate to the occurrence of oral cancer and oral pre-cancer. Highest ranked risk factors include prolonged tobacco usage and daily alcohol drinking habits; Second highest risk factor among Indians includes beetle nut chewing and unhealthy eating habits [1]. Oral cancer has been growing at a very rapid rate almost double occurrence rates after every three years; thus, its timely and accurate diagnosis can increase the survival period of patients.

According to the available statistics on oral cancer, the survival rates of oral cancer patients are not very good, only 30% survival rate. Oral cancer becomes more dangerous because after timely diagnosis of oral cancer there is not very significant improvement in survival rates, just an increase of 10% to 15 %. Squamous cell carcinomas are the most commonly oral cancer diagnosed accounting to almost more than 90% of all cases among oral malignancies. The symptoms for early stages of oral cancer may be whitish or reddish areas in mouth cavity, any injury, sore, blister or unusual development which does not go away after 2 week time, pain when swallowing, The tissue changes that occur in oral cancer are: loss of polarity, maturation of cells from basal to squamous cells in an unordered manner, unusual increased cellular density, premature keratinization and keratin pearls in deep epithelium layer etc. [2].

In recent years a significant increase in the cancer cases has attracted the attention of many researchers to automatic histopathological image analysis [3]. With the new and advanced medical imaging technologies data capturing in the medical domain has become so much easy. The data produced by these technologies is still very complex even experts of the domain need to invest a considerable time to give final diagnosis results. Such huge amount of data needs very huge time and effort to cleanse and filter because, not all data is important we only need relevant information and excluding any irrelevant data from this huge data. Thus, the need for robust feature extraction and selection methods has been a field of great opportunities for researchers working in medical image analysis domain. Many researchers have proposed some novel methods and some have used traditional features extraction and selection algorithms for the computer diagnostic systems depending upon the type of medical imaging data. Some researchers have proposed use of

[*] Corresponding Author

DOI: 10.37398/JSR.2022.660327

textural features, color based features, shape features, wavelet based features, geometrical features etc. on medical images. These features provide useful information that helps in accurate classification of these images. The most important task in feature extraction task is to decide the optimal feature that can enhance the performance of classification algorithm by giving it most relevant information. In this research work, a hybrid feature extraction and selection technique are proposed that will enhance the classification accuracy of histopathology images. The histopathology images take quite a long time to be examined by experts. Machine learning methods along with image processing methods are being used by many researchers to create a robust computer aided medical image diagnosis system. These methods help the expert to timely and accurate analysis of the image [4]. One more issue with medical image diagnosis is that different feature extraction algorithms can affect the classification results for the same image [5]. In this research article classification of normal tissue and cancerous tissue of oral cavity has been done with the help of hybrid features. The tissue images from histopathology slides were captured and labeled with the help of the expert. Feature extraction algorithms such as CNN, GLCM and LBP feature extraction algorithms have been used in this research article. Features extracted with each of these algorithms are applied to classification algorithms SVM, KNN, Naïve Bayes and Boo7sted Tree. Combinations of these feature extraction algorithms are also used with classification algorithms to check its effect on performance of classification algorithms.

The remaining article is divided into following sections: section 2 will provide a brief description of material and methods used in this research article. In section 3 results will be presented that were recorded after the experimentation. In section 4 discussion will be presented followed by ablation study regarding the different feature sets used in experiment. Finally in section 6 conclusion of the research carried out in this research article is presented.

II. Literature review

It is a well-established fact in machine learning and image classification tasks that better features are always a guarantee better classification results and medical image classification is no exception. Since the histopathology images are very complex so it may not be not possible to capture all the details with a single feature extraction method. Also, combination of more than one types of features have shown exceptional performance in classification of medical images. We investigated the efficiency of traditional textural features in combination with state-of-the-art deep features for detection and classification of cancer from oral histopathological images. Chatterjee et al. in [6] investigated the application of different feature extraction methods and concluded that statistical and cyto-morphological features when used together can identify various pre-cancerous lesions and oral cancer from histopathology images. Sometimes only one type of features are sufficient to classify images which are simple or when the classes are few in which images are to be classified. Chodorowski et al. in [7] showed that combination of colour, size and shape features extracted from histopathology images of oral lesions can be used for classification of Leukoplakia, erythroplakia, oral sub mucous fibrosis. They emphasized on the extraction of more discriminative features from colour spaces for better classification. In [8], Hu et al. proposed an automated model that is able to classify various types of tumors in a given computed tomography image. Author used anisotropic diffusion method for removal of noise and improving image quality by enhancing borders of objects. First order algorithm was used for extracting 5 textural features (mean, standard deviation, third moment, uniformity, smoothness, and entropy) whereas gray level run length matrix also extracted 7 more features. They employed SVM as classifier for feature vector obtained and achieved an accuracy of 90.11%, with 87.5% specificity and 92.16% sensitivity. Sometimes the selection of machine learning algorithm also results in poor classification results so we can use multiple classifiers for that purpose. In [9] Das et al. proposed similar methodology where an ensemble classifier was used to classify keratin pearl from histopathology images. Author extracted textural features using Gabor filter. In [10], Krishnan et al. proposed textural feature characterization for the detection of abnormalities in oral mucous. Author used multiple methods to extract features: discrete wavelet, Fractal dimension, Brownian motion curve, Gabor filters, and local binary pattern. Author used SVM classifier to classify three different tissue types from oral mucosa. Chang et al. proposed the use of clinicopathologic and genomic data for the prognosis of oral cancer in [11]. Author extracted features by using a hybrid approach by combining Pearson correlation Coefficient, genetic algorithm, Relief-F, Pearson's correlation coefficient and genetic algorithm (CC-GA), and ReliefF and genetic algorithm (ReliefF-GA). ANFIS was used for classification with AUC of 0.90 score. Proposed method has achieved very good classification accuracy in classification of cancerous and normal tissue images. This proposed method uses both traditional handcrafted textural features from GLCM and LBP algorithms as well as new deep learning approach. Proposed algorithm can extract both minor textural differences as well as colour and shape information from histopathology images. This proposed method can be a potential CAD that can serve as a very useful second opinion to expert pathologists. Since the availability of pathologist is very limited in small towns and remote areas, this method can also be used as remote diagnostic tool that will screen the biopsy images of patients and prompt the patients whether consultation of expert pathologist is needed or not.

III. Data and Methods used

The data used in this research work is collected from Govt. Dental College and Hospital, Jammu, India. Well labeled histopathology slides of 53 patients suffering from oral malignancy were collected. The collected histopathology slides were further analyzed under camera fitted microscope so that these slides can be subjected for quality assessment of slides. Nikon NIS F 3.2 was used for capturing images from the histopathology slides. Images were captured at 40X, 100X and 400X zoom but only 4ooX images were considered for this experiment as it shows cell and other biological structures in tissues clearly. Captured images were assessed by the expert pathologist and according to the comments of expert pathologist images were included or excluded from dataset. It took a time of 8 months approximately for collection, filtering and labeling of images, A total of 1475 histopathology images were collected out of which 163 samples were very poor and hence neglected, so a total of 1312 images were used in this research. Each of these 1312 labeled images were divided into 8 patches with dimensions of 128*128 pixels. A total of 10,496 labeled images were obtained which constituted the final data set. Out of 10,496 images 5779 images were of cancerous tissue and rest 4723 images were normal tissue images. Images were preprocess with Gaussian smoothing [12], so that the unwanted noise may be reduced from images.

Histopathological images are very large images and contain a lot of information. The processing of such large images is very exhaustive and time-consuming process. To avoid such exhaustive and time-consuming laborious work we divided the images into smaller patches and then the feature extraction applied on each patch. These image descriptive feature matrices obtained from feature extraction process are classified by classification algorithms. After preprocessing following feature extraction methods are used:

A. *Convolution Neural networks*

Commonly known as CNN is a state-of-the-art neural network. It has proved to be one of the best tools for image data analysis. Many researchers have used CNN and its variants for medical image data analysis. CNN can effectively extract features and classify the images with very good accuracy rates. The only thing is that it needs huge amount of data for its training and for medical imaging data the huge amount of data is not available sometimes. CNN belongs to feed forward neural networks which can be expressed as:

$$f(x) = fN(fN-1(fN-2 \ldots(f1(x)))) \quad (1)$$

Where N is number of hidden layers in network, and f(x) is the function to be carried out in the corresponding layer. In a typical CNN model, the main layers include a convolutional layer followed by activation layer followed by pooling layer and finally fully connected layers and predication layer. Convolutional layer is composed of multiple convolution kernels (K1, K2, …, K M−2,K M). Each KM represents a linear convolution function in the Mth kernel, which can be represented as:

$$K^{M}(x, y) = \sum_{p=-t}^{t} \sum_{q=-u}^{u} \sum_{r=-v}^{v} W_{M}(p, q, r) I(x-p, y-q, z-r) (2)$$

In this article CNN will be used as feature extractor only. The deep features extracted will be used for classification on different classifiers. In this research paper we have used ResNet50 for feature extraction [14]. The use of ResNet50 has been inspired by literature survey done for this research article.

B. *Feature Fusion*

We have extracted low level features from initial layers as well as high level features from final layers. As low-level features are very helpful in identifying color and simple edges information whereas high level features are very helpful in identifying complex shapes so a mix of these features will be used as final feature vector for CNN.

CNN extracts low level features from low level layers and high-level features from high level convolution layers. Low level features are generally rich in information like colour, positions, smaller edges whereas deep layers are able to extract more abstract features like complex geometric patterns and bigger shapes. CNN extracts only those deep layer features that it thinks are most relevant for classification purpose. There are many problems in which lower-level features perform better like classification of images which have complex texture contours. In some problems high level features are more important so Feature Fusion method provides a way to combine the low-level features as well as high level features in a fused feature vector that has very rich semantic information about the image. This fused feature vector that can be used for classification and can boost classification accuracy. In our Efficient Net we have used feature maps of convolution layers from all 5 blocks of ResNet 50. We have fused the features to get a feature vector of length 2048 which is same as final feature vector of ResNet 50, Since we need both original deep features and fused feature vector to evaluate the efficiency of feature fusion, we added Feature fusion layer to ResNet 50 in parallel to layer just before dense layers as shown in figure 1.

The final feature sets of both original deep features and fused deep features is then given to the dense layers of ResNet50 so that classification is carried out on these features.

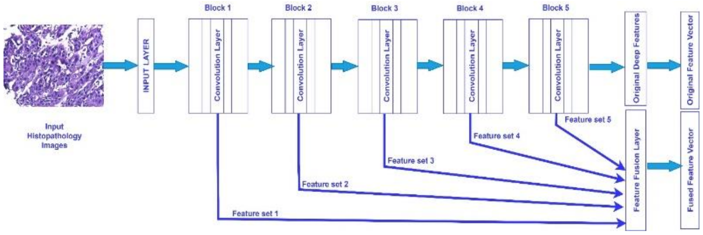

Figure 1: Showing Feature fusion in ResNet50.

C. *Gray level Co-Occurrence Matrix (GLCM)*

It is one of the most popular choice of researchers, it is a texture-based feature extraction method. The GLCM is used to determine the spatial textural relationship between pixels by performing second-order statistics operations on the images [15]. The GLCM determines the frequency of combinations of the pixel brightness values. A matrix is formed next which has exactly same number of rows and columns in it as there are number of gray values in the image. The GLCM features that we used in this study are: contrast, correlation, energy, entropy, homogeneity, sum variance, sum entropy, dissimilarity, difference entropy, inverse difference normalized. These features can be computed with following formulas:

Contrast=$\sum_{n=0}^{G_{max}-1} \ \{n^2 p_{x-y}(n)\}$

Correlation =$\frac{\sum_{i=1}^{G_{max}} \left\{\sum_{j=1}^{G_{max}} \{i.\ j.\ p_{i,j}\}\right\} - \mu_x . \mu_y}{\sigma_x . \sigma_y}$

Energy= $\sum_{i=1}^{G_{max}} \ \{[h_i]^2\}$

Entropy=$-\sum_{i=1}^{G_{max}} \ \left\{\sum_{j=1}^{G_{max}} \ \{p_{i,j} \,.\, Ln[p_{i,j}]\}\right\}$

Homogeneity=$\sum_{i=1}^{G_{max}} \ \left\{\sum_{j=1}^{G_{max}} \ \left\{\frac{1}{1+(i-j)^2} . p_{i,j} \right\}\right\}$

Sum variance=$\sum_{n=2}^{2\,.\,G_{max}} \ \{\left(n - \sum_{n=2}^{2\,.\,G_{max}} \ \{n \,.\, p_{x+y}(n)\ \}^2\right)$
$. p_{x+y}(n)\}$

Sum entropy=$-\sum_{n=2}^{2\,.\,G_{max}} \ \{\, p_{x+y}(n) \,.\, Ln[p_{x+y}(n)]\}$

Dissimilarity=$\sum_{i=1}^{G_{max}} \ \left\{\sum_{j=1}^{G_{max}} \ \{|i-j| \,.\, p_{i,j}\ \}\right\}$

Difference entropy= $-\sum_{n=0}^{G_{max}-1} \ \{\, p_{x-y}(n) \,.\, Ln[p_{x-y}(n)]\}$

Inverse difference=$\sum_{i=1}^{G_{max}} \ \left\{\sum_{j=1}^{G_{max}} \ \left\{\frac{1}{1+|i-j|} . p_{i,j} \right\}\right\}$

Where G_{max} is maximum quantized value, $p_{i,j}$ are pixel locations, σ is standard deviation and μ is variance.

D. Local binary pattern (LBP)

Local binary pattern is also one of the most popular feature extraction techniques that extracts features in the form of textures from images [16]. LBP is a very simple algorithm with high discrimination power. LBP shows very little change against changes in the grayscale levels of an image. LBP works in a simple way by selecting a pixel and its neighbours. LBP then compares the weight of the central pixel by its neighbour pixels. If the weight of neighbour pixel is smaller than central pixel then assigns "0" to the neighboring pixel otherwise it assigns a "1" to the neighbour pixel. The weight of the central pixel is then calculated using Equation. (3)

$LBP_{P_x, R_d}(W_c) = \sum_{k=0}^{P_x-1} \ b(I_k - I_c).2^k$ (3)

Here Px and Rd represent the number of neighboring pixels and radius respectively, b(i) is representing the binary threshold function and is given by,

$b(i) = \{0,\ i < 0\ 1,\ x \geq$ (4)

I_k and I_c represent the intensities of the kth neighboring pixel and central pixel respectively. Fig 2 shows the process of calculating the intensity of the central pixels using neighboring pixel values.

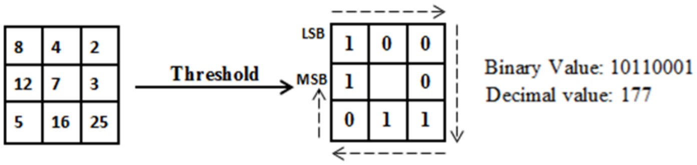

Figure 2: Binary code calculation of the central pixel in LBP

Since we are dealing with binary numbers so for each cell of dimension 3*3, the maximum number of intensity values calculated is 256, i.e., 28. Intensity values for all pixels are calculated and plotted in the form of histogram. Once all the histograms are calculated for all the cells in an image, the next step is normalization of histograms so that all values will be in a given range or we can say all values are mapped to certain range. These normalize histograms are then concatenated to create the feature vector that will be used as input to classifier.

E. SVM

This algorithm belongs to the class of supervised learning of machine learning algorithms. It is a simple but efficient classification and regression algorithm used by researchers for the last two decades. From the literature review we have seen the usefulness of SVM in medical domain. The idea behind the working of SVM is very simple and easy: just find the hyperplane which will divide the feature space into different target output classes. Generally, the hyperplane generated by SVM is for linear data but the data for classification is much more complex and non-linear in nature. So there is a need for some measure that will deal with non-linearity problem for that we have a number of kernels for SVM. Mathematically, we can state SVM as:

$f(x) = sign(w.x + b)$ (5)

here sign () = {1, -1} 1 for positive numbers and-1 for negative numbers, w is weight, x is input data and b is bias of hyperplane. The objective here is hyperplane should maximize the difference between classes so the optimization problem becomes:

$min \frac{1}{N} w^T . w$ subjected to $y_i(w.x_i + b) \geq 1, i = 1\ to\ N$ (6)

This equation is with respect to linear separable plane but the data may not be linear separable so we have to find a hyperplane in a higher dimension that will separate the classes for this we will use LaGrange's dual for above problem, we will add for each 'i' *a* slack variable δ_i.

$\delta_i = max\ (0, 1 - y_i(w.x_i + b))$ and LaGrange's multipliers $\alpha_1,\ \alpha_2, \alpha_3, \dots \dots, \alpha_n$ >=0 and solve this by using quadratic programming as follow:

$L(w, b, \alpha)$=$\frac{1}{N} w^T . w - \sum_{i=1}^{n} \ \alpha_i y_i(w.x_i + b) + \sum_{i=1}^{n} \ \alpha_i$ (7)

Finally, our optimization problem becomes:

$$L(\alpha) = \sum_{i=1}^{n} \ \alpha_i - \frac{1}{N} \sum_{i,j=1}^{n} \ \alpha_i \alpha_j y_i y_j x_i . x_j$$

$subject\ to \sum_{i=1}^{n} \ \alpha_i . y_i = 0, \alpha_i \geq 0$ (8)

Let us suppose that α_i^* is optimal solution then we can easily find the optimal values for weight and bias as:

$w^* = \sum_{i=1}^{n} \ y_i \alpha_i^* x_i$ and

$b^* = \frac{1}{N} w^*[x_p + x_n]$

Where $x_p\ and\ x_n$ are support vectors.

Now if we use the kernel in SVM the above equation 8 can be re written as:

$L(\alpha) \ = \sum_{i=1}^{n} \ \alpha_i - \frac{1}{N} \sum_{i,j=1}^{n} \ \alpha_i \alpha_j y_i y_j K(x_i . x_j)$

$subject\ to \sum_{i=1}^{n} \ \alpha_i . y_i = 0, \alpha_i \geq 0$ (9)

Where the dot product of $x_i.x_j$is replaced by $K(x_i.x_j)$, here K is kernel may be linear, polynomial or exponential. We have used Gaussian radial basis function as kernel which is an exponential function given as:

$$K(x_i,x_j)\text{=exp }(-\gamma\|x_i - x_j\|^2) \qquad (10)$$

$where\ \gamma > 0$ is parameter of kernel.

F. K-nearest Neighbors

This algorithm is a fairly simple and very useful algorithm for classification purposes. It also belongs to the class of supervised learning algorithms. It simply starts with deciding the number of neighbors (k), then calculate the Euclidean distance of K neighbors. Count the data points in each category; assign the new data to category in which data points are maximum. Euclidean distance can be found using following formula:

$$\text{Euclidean distance}=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

The problem with KNN is the selection of optimal value of K, for smaller K values the data gets noisy and effect of outlier can be seen on its performance. For bigger values of K the algorithm over fits data so we have to take a value that is neither too big nor too small, we have taken K as 5 that is also default value.

G. Naïve Bayes

This algorithm is one of the faster algorithms in machine learning algorithms. This algorithm assumes the input features are independent of each other and are contributing equally int the prediction of class. It uses byes theorem to calculate the probability of one event occurring from the probability of another event already occurred. Mathematically:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (11)$$

P(A) is priori probability of A, P(A|B) is posteriori probability. If we add the assumption of Naïve to Bayes theorem, we can write Bayes Theorem as:

$$\text{P}(\text{A}|B_1,2,B_3\ \ldots..B_n) = \frac{P(A)\ P\ (A)\ P(A)\ldots\ldots\ldots\ldots P(A)\ \ P(B)}{P(B_1)\ P(B_2)\ P(B_3)\ldots\ldots\ldots\ldots P(B_n)} \qquad (12)$$

We can write the above equation as:

$$\text{P}(\text{A}|B_1,2,B_3\ \ldots..B_n) \propto\ P(B)\prod_{i=1}^{n}\ P(A) \qquad (13)$$

Now if we select the maximum probability for different input values of B, it becomes out Naïve Bayes Classifier, which can be mathematically stated as:

$$B = argmax\ _y P(B)\prod_{i=1}^{n}\ P(A) \qquad (14)$$

H. Boosted Tree

Boosted tree algorithm is basically a collection of decision trees connected sequentially so that the next tree minimizes the error of previous tree. In this algorithm decision trees are weak learners but their combination and error residual checking on previous trees make this algorithm more efficient. Logarithmic loss function is used to compute loss.

$$H_P(q) = -\frac{1}{N}\sum_{i=1}^{N}\ y_i\ .log\big(P(y_i)\big) + (1-y_i).log(1-P(y_i))$$
(15)

We have to select learning rate (α) and number of decision trees carefully in this algorithm because it controls working of overall algorithm. It is recommended that we keep a smaller learning rate initially. Similarly, number of estimators (n_estimators) also has an effect on over fitting of algorithm. A large number of estimators make it prone to over fitting. The steps in boosting algorithm are as follow:

$f_1(X) = y$

the residual becomes y- α$f_1(X)$

$f_2(X) \approx$ y- α$f_1(X)$

the residual becomes y- α$f_1(X)$- α$f_2(X)$

$f_3(X) \approx$ y- α$f_1(X)$- α$f_2(X)$ and so.

I. Majority Voting

Majority voting is a method of combining results from a number of ensemble classifiers in machine learning. The basic idea behind the working of majority voting is that aggregate the result of each classifier and predicts the target class based on majority of voting. Mainly majority voting is of two types hard voting and soft voting. Hard voting simply takes in account the highest probability of predicted result and declares it as output. Whereas in soft voting the average of output probability is taken and the class with highest average is taken as final result. In this article we have used soft voting technique.

$$Y' = argmax\,\frac{1}{n}\sum_{i=1}^{n}\ (p_{1,}p_{2,\ldots\ldots\ldots}p_{n,}) \qquad (16)$$

Where, Y' is final output of majority voting, 'n' is number of classifiers, p_i is probability of ith classifier.

IV. Experimentation and Results

In this research paper three feature extraction algorithms are used. CNN and feature fused CNN for deep feature extraction, LBP and GLCM for extraction of textural features. Four well known classifier algorithms are used SVM, KNN, Naïve Bayes and Boosted trees. The choice of classification algorithm is based on the literature survey. All these algorithms have shown great potential in medical image diagnosis. Figure 3 gives an overview of the experiment.

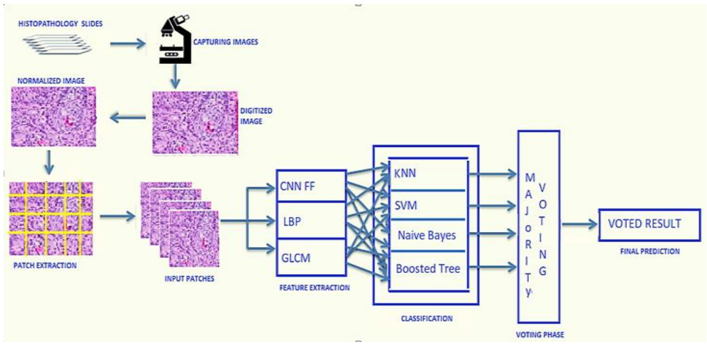

Figure 3: Overview of proposed system

Implementation of proposed system was done in Python using Tensor flow and keras on Google COLAB. 10-Fold cross validation was used to training and testing purpose of our prosed system. 20 % of data from dataset was used for testing purpose of the model. Test data was not used in during 10-fold cross validation so this data is totally new for model. Dataset was preprocessed in order to remove unwanted noise and blur that exist in the images during capturing process. For this, the image is first applied with a 2D Gaussian smoothing filter, as in Equation 16.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}} \qquad (16)$$

Where x is the distance from the origin in horizontal axis, y is the distance from the origin in vertical axis, and σ is the standard deviation of the Gaussian distribution [4]. After applying smoothing filter 2D median filtering was applied for sharpening of image so that gray level intensities at corners and edges become more apparent. Preprocessing reduces the amount of false feature information by reducing noise and other undesired information.

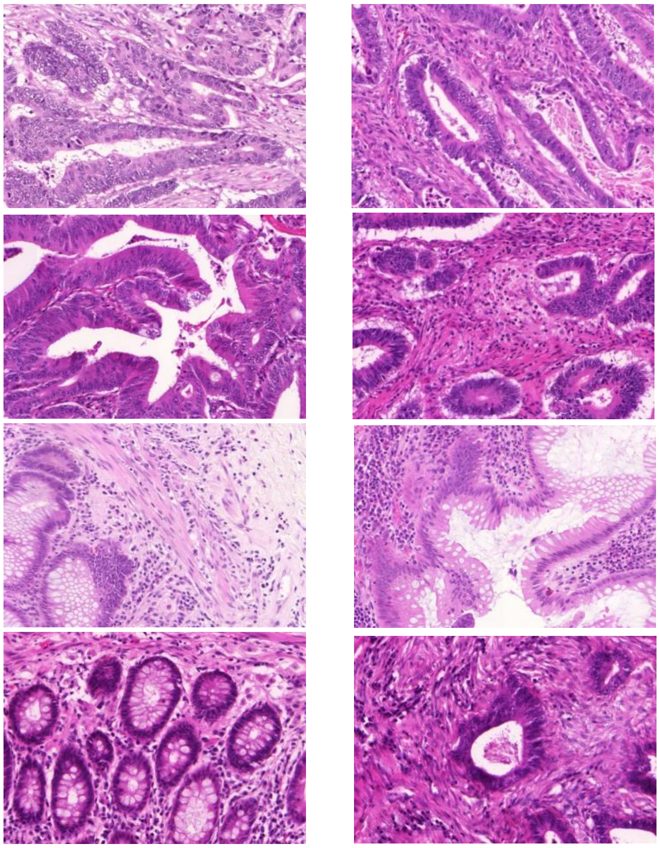

Figure 4: Sample of histopathology images in collected dataset.

LBP generated a feature vector of length 256. GLCM generated feature vector of length 22. ResNet50 generated 2048 features. We applied PCA to include only those features that can give better accuracy and removing any insignificant features that don't contribute significantly in the accuracy of the classification algorithms. After applying PCA number of features of LBP was reduced to 89. GLCM features were reduced to 13. Deep features extracted by ResNet50 were not reduced because it is a fused feature set and we wanted to keep all the features. The feature selection is based on the effect of features on accuracy of classification algorithm. We have performed rigorous experiments to see the optimal number of features at which classification algorithms performed better. Table 1 gives the accuracy of classification algorithms based on the types of feature extraction algorithm.

Table 1: Accuracy obtained by classifier algorithms on different types of features

	Deep Features	**Fused deep Features**	**LBP**	**GLCM**
SVM	90.39%	**94.16%**	89.57%	92.53%
KNN	86.17%	88.37%	81.11%	**90.35%**
Naïve Bayes	88.29%	91.82%	90.63%	**92.61%**
Boosted trees	91.63%	**95.68%**	91.47%	93.72%
The proposed hybrid ensemble method achieved highest accuracy of **98.71%**				

A comparative analysis of proposed system has been done with other similar already existing methods in the form of a table. Table 2 shows the domain in which researcher has done the similar work, features used by researcher, classification algorithms used and the accuracy achieved by their method.

Table 2: Showing the performance comparison in terms of accuracy of proposed methods with other similar methods.

Author	**Application Area**	**Features**	**Classifiers**	**Accuracy**
Chatterjee et al. [6]	Oral pre cancer/cancer detection	Morphological, intensity and color, Texture	KNN, SVM, MLP, Decision tree, Random Forest	90%
Chodorowski et al. [7]	Oral lesion classification	Color, Shape	SVM, BPM	85%
Das et al. [9]	Oral carcinoma diagnosis	Deep, Texture	CNN, random forest	98.6%
Chang et al. [11]	Oral cancer prognosis	Clinico-pathologic, Genomic	ANN, ANFIS, SVM, Logistic regression	93.81%
Proposed method	Oral cancer detection	Deep, Texture	SVM, KNN, Naïve Bayes, Boosted Tree, Ensemble	**98.71%**

From table 2, it is clear that our proposed method is performing better than other listed methods. Das et al., [9] has also achieved a comparable accuracy in his work of oral squamous cell carcinoma diagnosis. Chang et al., [11] also achieved very good accuracy of 93.81% by using clinicopathologic and genomic features. In terms of classification accuracy, we can conclude that proposed method is better in detection of oral cancer from histopathological images and can be used as a potential tool for timely and accurate diagnosis of oral malignancies.

V. Ablation Study

In this section the effect of various feature extraction methods on classification algorithms. An investigation of feature fusion mechanism on CNN will also be done. For evaluation purpose accuracy will be chosen as primary evaluation metric.

A. Effect of Fused Deep Features on classifiers

In case of SVM we can clearly see that the classification accuracy has increased significantly from 90.39% to 94.16% when we used fused deep features instead of conventional deep features. The main reason behind this improvement is that fused features are more diverse and informational as it contains both low level and high-level features.

In case of KNN the classification accuracy has increased by 2.20% by using fused features and final classification accuracy of 88.37% which is lowest of all classifiers. The reason for this is that the dimensionality of deep features and fused deep features is higher than most of the other types of features. Hence KNN suffers with large dimensional data and produces poor result.

In case of Naïve Bayes classifier, the classification accuracy increased from 88.29% to 91.82% when we used fused feature set. The reason for this increase in classification accuracy is that Naïve Bayes treats each feature as independent of other and treats them equally important, so fused feature vector is diverse than conventional deep features hence diversity of features aided to enhance the accuracy of Naïve Bayes.

In case of Boosted Tree there is significant increase from 91.63% to 95.68%. The reason is that Boosted Trees usually produce poor result when the feature used for classification are corelated, but in case of fused features the features are diverse and there is a very less of correlation among them as they are collected from different layers. Hence Boosted trees performed better than other classifiers.

B. Effect of LBP and GLCM features on classifiers

LBP and GLCM are both feature extraction methods based on texture feature extraction. LBP extracts local texture patterns whereas GLCM extracts second order statistical texture patterns. From table 1 it can be clearly seen that LBP is the worst performer among all feature extraction methods. GLCM has achieved better results on classifiers. SVM achieved 92.53% classification accuracy as compared to 89.57% with GLCM and LBP respectively. KNN achieved 90.35% classification accuracy on GLCM and 81.11% accuracy on LBP. The reason for this difference is that KNN has effect of magnitude of features, the greater the magnitude of feature the greater will be distance value of Euclidean distance in KNN. Hence performance is affected more in LBP than in GLCM because the magnitudes of LBP features are greater than GLCM features. For Naïve Bayes there is a difference of 2% accuracy in LBP and GLCM. GLCM performed better than LBP because of the local correlation between LBP features is more than GLCM. In case of Boosted Tree, GLCM performed better than other classifiers and achieved 93.72% accuracy as compared to 91.47% on LBP. The reason for this being GLCM features are more distinctive than LBP features, hence Boosted trees perform better with GLCM.

VI. Conclusion

In this research paper we attempted to study three well known feature extraction algorithms. These algorithms were applied to histopathology images of oral cavity for feature extraction. Extracted features were applied to four classifiers and their classification accuracies were calculated. These classifiers were then combined with majority voting algorithm to predict the final output. The results obtained in table conclude that deep features and GLCM features perform better than LBP features. Boosted trees achieved highest accuracy of all three four classifiers on all types of features. Feature fusion in CNN has shown great potential for classification of histopathology images. Clearly, we can see a remarkable increase in accuracy of classifiers on fused CNN features rather than basic CNN features. Fusion of all these feature extractors can be used in creation of a diagnostic tool that can aid a pathologist in early and timely detection of oral cancer from histopathology images. The proposed method could be developed as a tool that could provide a second opinion to the pathologist in case diagnosis is inconclusive.

In future we plan to study and apply more complex feature extraction algorithms on medical image classification. We will enhance the proposed model to classify multi-modality medical images like MRI and CT scans in future by making necessary changes in initial layers and fine tuning in layers where needed. Also, we attempt to create and publish a publicly available benchmark database for oral histopathology images so that researchers can work on it and get the ample amount of data to work with.

In the last conclusion remark the future direction for researchers willing to work in this domain is to explore the feature fusion and attention mechanism in case of histopathology images. Feature fusion and feature attention is also a domain that is needed to be explored because of its capability to improve the classification results in case of medical images.

Conflict of Interest: Authors declare that there is no conflict of interest in this research work with any individual researcher or any organization.

Acknowledgement: Authors would like to acknowledge the help and support shown by the medical and technical staff of

Indira Gandhi Government College and Hospital, Jammu.

Contributions of Authors:

Rachit Kumar Gupta: Data curation, writing original draft, conceptualization and methodology, Visualization.

Jatinder Manhas: software, formal analysis, validation, investigation, writing review and editing, project administration.

Mandeep Kaur: Data curation, supervision, resources, validation.

References

[1] Oral Cancer Facts (2020), retrieved 17 March, 2021 from https://www.oralcancerfoundation.org/facts/index.html

[2] Medical News Today (2020), retrieved 17 March, 2021 from https://www.medicalnewstoday.com/articles/165331

[3] Sertel, O., Lozanski, G., Shana'ah, A., Gurcan, M. N. (2010). Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation. IEEE Transactions on Biomedical Engineering, 57(10), 2613-2616.

[4] Mikhaylov, V. V., Bakhshiev, A. V. (2017). The System for Histopathology Images Analysis of Spinal Cord Slices. Procedia Computer Science, 103, 239-243.

[5] Nabizadeh, N., Kubat, M. (2015). Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features. Computers & Electrical Engineering, 45, 286-301.

[6] Chatterjee, S., Nawn, D., Mandal, M., Chatterjee, J., Mitra, S., Pal, M., Paul, R. R. (2018). Augmentation of Statistical Features in Cytopathology towards Computer Aided Diagnosis of Oral Pre-cancer and Cancer, Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII), Chennai, India, pp. 206-212.

[7] Chodorowski, A., Choudhury, C. R., Gustavsson, T. (2008). Image analysis and CAD system for mucosal lesions, 8th IEEE International Conference on BioInformatics and Bioengineering, Athens, Greece, pp. 1-4.

[8] Hu, Z., Alsadoon, A., Manoranjan, P., Prasad, P. W. C., Ali, S., Elchouemic, A. (2018). Early stage oral cavity cancer detection: Anisotropic pre-processing and fuzzy C-means segmentation, 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, pp. 714-719.

[9] Das, D. K., Bose, S., Maiti, A. K., Mitra, B., Mukherjee, G., Dutta, P. K. (2018). Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. Tissue and Cell, vol. 53, pp. 111-119.

[10] Krishnan, M. M. R., Pal, M., Bomminayuni, S. K., Chakraborty, C., Paul, R. R., Chatterjee, J., Ray, A. K. (2009) Automated classification of cells in sub-epithelial connective tissue of oral sub-mucous fibrosis — An SVM based approach, Computers in Biology and Medicine, vol. 39, no. 12, pp. 1096-1104.

[11] Chang, S. W., Kareem, S.A., Merican, A. F., Zain, R. B. (2013). Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods, BMC Bioinformatics, vol. 14, 170.

[12] Shapiro, L. G. Stockman, G. C. (2001). "Computer Vision", pp 137, 150. Prentice Hall.

[13] Gao, X.W., Hui, R., Tian, Z. (2017). Classification of CT brain images based on deep learning networks. Computer Methods and Programs in Biomedicine. 138, 49—56.

[14] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 27-30, Las Vegas, NV, USA.

[15] Albregtsen, F., Nielsen, B., Danielsen, H. E. (2000). Adaptive gray level run length features from class distance matrices. In Pattern Recognition, Proceedings. IEEE 15th International Conference on Vol. 3, pp. 738-741.

[16] Ojala, T., Pietikainen, M., Harwood, D. (2019). A comparative study of texture measures with classification based on featured distributions, Pattern Recognition, vol. 29, no. 1, pp. 51-59.

***