An image patch selection algorithm for the detection of Oral Squamous Cell Carcinoma using textural and morphological features

Eranjoli Nalupurakkal Subhija (subija@cusat.ac.in)
Cochin University of Science and Technology
Vaninirappuputhenpurayil Gopalan Reju
Cochin University of Science and Technology

Research Article

Keywords: Oral squamous cell carcinoma, Wavelet transform, Gray level co occurrence matrix, pretrained networks, extra trees classifier

Posted Date: June 23rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3085184/v1

Additional Declarations: No competing interests reported.

An image patch selection algorithm for the detection of Oral Squamous Cell Carcinoma using textural and morphological features

Eranjoli Nalupurakkal Subhija*
and Vaninirappuputhenpurayil Gopalan Reju

Department of Instrumentation, Cochin University of Science and Technology, Cochin, Kerala, India - 682022.
*Corresponding author. E-mail: subija@cusat.ac.in.

Abstract

A significant proportion of cancer-related deaths are caused by oral cancer, and oral squamous cell carcinoma is prevalent. Systems for computer-aided diagnosis can lower subjective errors and assist the pathologist in making a more accurate diagnosis. Feature extraction from the whole histopathology images is difficult due to the structural variation of the oral tissue images. A new patch selection algorithm can be used to create and select image patches containing nuclei-specific information and extract their textural and morphological characteristics, improving cancer diagnosis. We extract the morphological characteristics of the nucleus from the selected patches using five pretrained networks. The texture from the regions of the selected patches is also extracted from the Haar wavelet decomposed components using the gray-level co-occurrence matrix. Then, we combine the textural and morphological features to create the final feature vector, followed by feature selection using an extra trees classifier. In order to detect oral squamous cell carcinoma, we examined the feasibility of using six classifiers, including voting classifier, logistic regression, random forest, Naive Bayes, K nearest neighbor, and support vector machine. The performance of the algorithm is evaluated using accuracy, precision, sensitivity, confusion matrix, ROC curves, and AUC values of characteristic curves. Together, ResNet 50 and the voting classifier produce results with a high accuracy of 97.66 % and a precision of 98.00 %. The suggested patch-based method outperforms the image based method and is accurate and efficient for identifying oral squamous cell carcinoma and will be a reliable and precise support tool for oral pathologists.

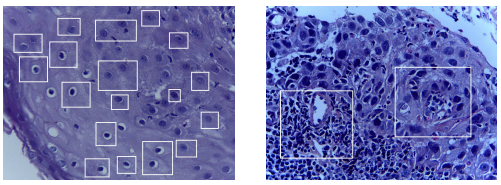
Keywords: Oral squamous cell carcinoma, Wavelet transform, Gray level co occurrence matrix, pretrained networks, extra trees classifier.

1 Introduction

Oral and pharyngeal cancer is the world's sixth most common cancer. Oral cancer is said to be responsible for about 2 % of cancer-related fatalities [1]. Out of the 0.3 million oral cancer cases reported yearly worldwide, 48 % were severe [2]. According to a study conducted in 2018 [3], there is a significantly higher chance of survival if it is diagnosed and treated early. Oral cancer is a malignant condition where an abnormal mass of tissue grows in any part of the mouth, including the lips, gums, tongue, inner cheek lining, the roof of the mouth, and the floor. Currently, the gold standard for oral cancer detection is histopathological image analysis. These images are created during a biopsy where oral tissues are prepared onto slides and stained with hematoxylin and eosin dyes, also known as H &E dyes. Every tissue component is profoundly stained by these dyes, enabling doctors to see it more clearly [4]. These stained glass slides are then converted to digital images using the whole slide imaging (WSI) method for manual or computer-assisted diagnosis and storage [5]. Pathologists can identify the malignancy from these images based on the nucleus' size and color, changes in the nuclear-cytoplasmic ratio, and the color of the stained tissue.

Oral squamous cell carcinoma (OSCC) is the most prevalent type of oral cancer. The squamous cells that constitute the lining of the oral cavity, known as the squamous epithelium, can be seen infiltrating the connective tissue as sheets, nests, and islands under a microscope. The morphological characteristics of squamous cells seen in OSCC include increased cell division, aberrant cell division, island development, and intercellular bridges [6]. In contrast, the image of a normal tissue won't show any of these abnormalities. In Fig. 1, the oral histopathology images of normal and OSCC tissue are displayed to highlight these variations.

In clinical care, oral pathologists manually examine biopsy samples under a microscope to make decisions. There are several problems with visual evaluations. First of all, it is time-consuming and challenging for pathologists to examine a large number of slides. The high cell and nucleus densities in oral histopathology images make it difficult for human eyes to distinguish these kinds of subtle microscopic characteristics [8]. Also, the manual identification of early cancer symptoms is challenging and heavily relies on the training and experience of the doctor or healthcare professional. As a result, nearly half of all cases of oral cancer are presently



(a) (b)

Fig. 1: Oral histopathology images of (a) Normal tissue, where nuclei (for ease of identification, some of the nuclei are highlighted with rectangular box) are circular and the shapes are not distorted, and (b) OSCC tissue, where the highlighted area represents islands of nuclei and the shape of nuclei are distorted [7].

discovered at an advanced stage. So, more efficient oral cancer screening programs are critically needed to address this issue.

These problems may be resolved by developing computer-aided diagnostic techniques, which are usually based on the segmentation of nuclei and concentric rings of squamous cells known as keratin pearls as well as the extraction of textural and morphological features from the images. Keratin pearls are the main indicator of oral cancer [9], but they might not always be visible in OSCC images, which could result in erroneous classification. Therefore, instead of looking for keratin pearls, textural characteristics like energy and entropy from the RGB color components of the median filtered images can be used [10]. To detect OSCC from images, they were decomposed using a 2D empirical wavelet transform, and textural features were extracted from the wavelet components [11]. The representation of an image by a wavelet transform happens in a space whose coordinate system has an interpretation closely related to the characteristics of a texture, such as frequency or size [12]. Textural characteristics were extracted from the images as well by converting them to the Gray Level Co-Occurrence Matrix (GLCM) [13] and the Local Binary Pattern (LBP) [14] because the GLCM determines an image's texture by determining how frequently pairs of pixels with specific values and spatial relationships appear in the image [15] and the local spatial patterns and the contrast in the grey scale in an image are effectively captured by LBP descriptors [16]. With the most recent advancements in machine learning, numerous deep learning-based techniques, including convolutional neural network (CNN), pre-trained deep CNN networks [17], like Alexnet, VGG 16, VGG 19, ResNet 50 [18], MobileNet [19], multimodal fusion

with CoaT (coat-lite-small), PiT (pooling based vision transformer pit-s-distilled-224), ViT (vision transformer small-patch16-384), ResNetV2 and ResNetY [20], and concatenated models of VGG 16, Inception V3 [21], have been proposed for the automated extraction of morphological features. After the feature extraction, the images were classified into normal and OSCC categories using different classifiers such as random forest [22], support vector machine (SVM) [10], extreme gradient boosting (XGBoost) with binary particle swarm optimization (BPSO) feature selection[23], K nearest neighbor (KNN) [10], duck patch optimization based deep learning method [24] and two pretrained models, ResNet 50 and DenseNet 201 [11]. However, as the number of layers of the network increases, the complexity also will increase. This will again increase if multiple models are combined.

Processing the high-resolution WSI oral histopathology images for OSCC detection involves more time, hardware, and memory. Also, when employing transfer learning on WSI images directly, for cancer detection, the input images must be resized into an appropriate input size for the network to train. However, there is a possibility of information loss when resizing the images [25]. So, we can reduce complexity and computation time by providing patches containing the relevant data for the diagnosis rather than the entire histopathology image. As we reported previously [26], combining morphological characteristics with the image's textural features can result in a powerful method for cancer detection. However, combining these features from the WSI images results in bigger feature sizes, which raises computational complexity. To address the aforementioned issues, we propose a new patch selection algorithm to create and select patches from images which hold discriminative information, followed by textural and morphological feature extraction, selection, and classification. GLCM is used to extract textural features from the wavelet decomposed patches, and pretrained networks extract morphological features directly from the patches. Then the features are fused to make a feature vector, and the best features are selected from this vector to produce the final feature vector. This final vector is then utilized to classify images into normal and OSCC. The algorithm is computationally efficient and validated using a public oral cancer dataset [7]. The experimental results show that the proposed patch-based method outperforms the other existing OSCC detection methods.

The main contributions in this paper are summarised as follows:

- We present a new algorithm for patch selection from oral histopathological images.

- Proposed a feature fusion framework for the fusion of textural and morphological features extracted from the image patches.

- Investigated the predictive ability of six distinct classifiers to classify the images into normal and OSCC.

The rest of the paper is organized as follows. In section 2, we explained the methodology of the proposed algorithm. An explanation of the materials used for the experiment is also included in section 2. Experiment results and analysis, including comparison with the existing algorithms, are given in Section 3. Finally, Section 4 concludes the paper.

2 Materials and methods

In this section, we first present the data used for validation and the overview of our patch-based oral cancer classification model, followed by the steps and the detailed structure of the proposed algorithm.

2.1 Data description

Compared to other techniques for oral cancer detection like CT imaging and Fluorescence Visualization, oral histopathology images are especially useful in the early diagnosis of oral cancer [27]. The histopathological image database of oral cancer, with finely annotated normal and OSCC histopathology images, was used in this study [7]. There were 1224 oral histopathology images in the repository, divided into two sets with two distinct magnifications. The dataset's first group includes 439 images of OSCC and 89 oral histopathological images of the normal epithelium with a magnification factor of 100. The second group consists of 495 histopathological images of OSCC and 201 images of the normal oral tissue epithelium with a magnification factor of 400. The images were taken from 230 patients using a Leica ICC50 HD microscope. The oral tissue slides were gathered, prepared, and categorized by medical professionals.

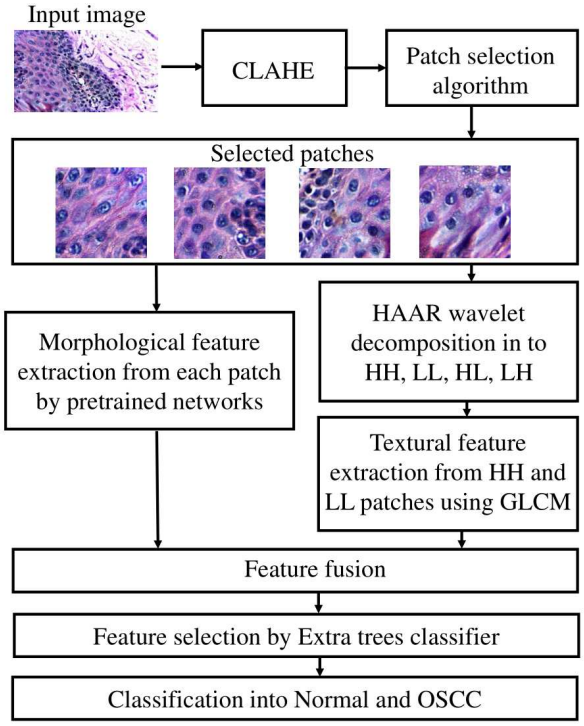


Fig. 2: Block diagram for the proposed method.

2.2 Proposed Method

The overall structure of the proposed OSCC detection method is shown in Fig. 2. The method is divided into five stages: preprocessing of the image, the extraction and selection of patches, the fusion of morphological and textural features from the chosen patches, feature selection, and classification. Following the preprocessing, patches that are 50% overlapping based on the patch's density map are extracted, and the best four patches are selected using a criteria, which is described in section 2.2.2. It is commonly accepted that the human visual system gathers significant features in the frequency domain regardless of their illumination or the presence of noise. Past investigations have shown that the visual cortex possesses orientation and band-pass filters for analyzing images [28]. So, it is more effective to analyze an image's texture in the frequency domain and extract the feature that includes all of the image information and this method is frequently used for global texture analysis [29]. In the proposed method, we transform the images into the frequency domain

using Haar wavelet transform because it improves the distinguishability of features extracted from the image [30]. After decomposing using Haar wavelets, textural features are extracted from the selected patches. However, morphological features are directly extracted from the patches using pretrained neural networks because these networks try to learn both the high-level and low-level features from the images. Then, the complete features of the histopathology image or patches are obtained by fusing the textural and morphological features. Finally, the fusing of the features is followed by feature selection and classification. A detailed explanation of the algorithm is given below.

2.2.1 Preprocessing

Many factors usually degrade histopathology images, resulting in low contrast images [31]. The preprocessing of histopathology images is the first stage in the detection of OSCC. In this research, we used contrast limited adaptive histogram equalization (CLAHE) to improve the images' local contrast [32]. The CLAHE method, which was created using a predetermined clip limit value, is used to minimize noise while enhancing the quality of histopathological images [33].

2.2.2 Patch selection

It is challenging to locate and segment the pathological area of interest in the WSI images for the diagnosis of cancer and there are difficulties in running direct deep learning models on the WSI due to their size/ high resolution. This problem can be mitigated by splitting the image into smaller patches and selecting only the relevant patches for further investigation. Hence, a patch extraction and selection method that selects the patches that contain the relevant data about the nucleus is developed in this work. Fig. 3 shows the block diagram for the patch selection algorithm.

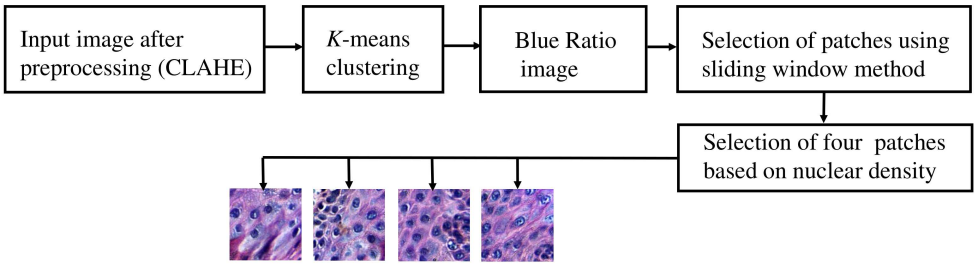


Fig. 3: Block diagram for patch selection algorithm.

In the histopathological images under test, our area of interest is the nucleus. So, after pre-processing, the K-means clustering algorithm divides the image into $K=3$ regions; nucleus, cytoplasm and background. The next step is to convert the current clustered RGB image into a blue ratio image [34], which aids in locating the image's nuclei or the most noticeable and brightest objects. The mathematical formula for converting an RGB image to a blue ratio image is given by [34]:

$$\text{Blue ratio} = \frac{100*\text{B}}{1+\text{R}+\text{G}} \times \frac{256}{1+\text{R}+\text{G}+\text{B}} \quad , \tag{1}$$

where 'R', 'B,' and 'G' are the red, blue, and green channel intensities, respectively. Finding the nuclear density is the next stage after obtaining the blue ratio image. The nuclear density, which is defined as the number of cell nuclei per unit area [35], is a crucial parameter to be noted when examining features related to tissues or cells. The nuclear density of OSCC images is typically more significant than that of normal images. For finding the nuclear density, the blue ratio image is first binary thresholded, and the image is filtered using a sliding window of an all-ones matrix of size 256 x 256 to create the nuclear density map, where the whiteness of the density map is an indication of the nuclear density level. Then, choose four image patches of size 512×512 or smaller (near the image edges) such that the patch centers are the top intensity levels in the image which satisfies the condition that the maximum overlap between any two patches is 50%. The images at different stages in a typical patch selection process are shown in Fig. 4.

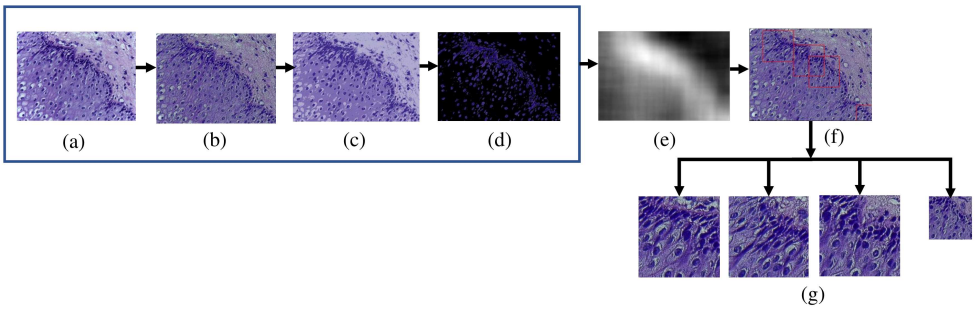


Fig. 4: Images at different stages in a typical patch selection process (a) Input image, (b) pre-processed image, (c) K-means clustered image with K=3, divides the enhanced image into three regions; the nucleus, cytoplasm and background, (d) blue ratio image, (e) nuclear density map, where nuclear density is shown by the whiteness level (f) 50% overlapping patches, highlighted by red colour bounding boxes, and (g) the selected four patches.

2.2.3 Haar wavelet decomposition of patches for textural feature extraction

Based on the image's shape, morphology, intensity, and texture, histopathological images can be classified as normal or cancerous. Extraction of information from the entire image or patches that rely on texture is crucial for the detection of OSCC. The term texture refers to specific characteristics of the internal constitution or structure of tissues, and texture-based approaches are often investigated for the examination and classification of histopathological tissues [36] [13]. Instead of extracting the textural features directly from the patches or images, it is more beneficial to split them into frequency components using wavelets like Haar, as stated in [29]. The Haar wavelet [37] transform provides excellent frequency resolution for low-frequency components, which is the mean intensity values of all image pixels and good temporal resolution for high-frequency components, which are the edges present in the image when applied to images or patches.

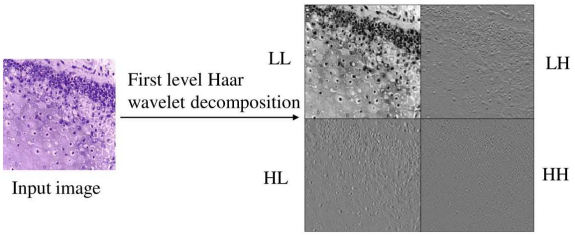


Fig. 5: First level Haar wavelet transformation with four frequency components LL, LH, HL and HH.

In our proposed approach, a Haar wavelet function was applied to each patch and the total information was split into three detailed and one approximate coefficient such as HL, LH, HH, and LL, which stand for the horizontal, vertical, diagonal detail and approximation coefficients, respectively, which are perfect for extracting the most discriminative textural characteristics [38]. This decomposition process is shown in Fig. 5. In this decomposition, the high-frequency detail components are the complementary part of the low-frequency approximation components. So, the approximation and any one detail coefficient are sufficient to describe the features of the image. The Gray Level Co-occurrence Matrix (GLCM), first suggested by Haralick in 1973 [36], is a widely used technique for extracting textural features. After converting the images to grayscale, the GLCM primarily focuses on the spatial connection of the pixel, such as angle and distance. Six texture features were chosen from a total of 28 in the GLCM, including dissimilarity, homogeneity, contrast, correlation, energy, and angular second moment (ASM). They give structural information about the cells and nuclei in each region. The GLCM features are enumerated in equations (2) to (7) and each element of the matrix is denoted by $\boldsymbol{c_{i,j}}$, where $\boldsymbol{i}$ and $\boldsymbol{j}$ denote coordinates and $\boldsymbol{N_G}$ denotes the number of grey levels in the image. The GLCM matrix's row-wise and column-wise means of all pixel intensities are denoted by $\boldsymbol{\mu_i}$ and $\boldsymbol{\mu_j}$, respectively. Whereas, $\boldsymbol{\sigma_i}$ and $\boldsymbol{\sigma_j}$, respectively, are used to represent the row-wise and column-wise variance of all pixel intensities. Dissimilarity is a measure of the distance between two elements in the GLCM matrix, whereas homogeneity is a measure of the closeness of the distribution of matrix elements in the GLCM. Contrast measures the contrast between grey levels and correlation measures how a pixel in an image or patch relates to its neighbor. Another GLCM feature, energy, is obtained after

taking the root of the sum of squared elements in the GLCM and it represents the uniformity of the image texture, with higher energy values indicating a more heterogeneous texture. The ASM is calculated as the square of energy based on the homogeneity present in an image.

$$\text{Dissimilarity} = \sum_{i,j=0}^{N_G-1} |i-j|\, c_{i,j} \tag{2}$$

$$\text{Homogeneity} = \sum_{i,j=0}^{N_G-1} \frac{c_{i,j}}{1+(i-j)^2} \tag{3}$$

$$\text{Contrast} = \sum_{i,j=0}^{N_G-1} c_{i,j}(i-j)^2 \tag{4}$$

:

$$\text{Correlation} = \sum_{i,j=0}^{N_G-1} c_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i)^2(\sigma_j)^2}} \tag{5}$$

$$\text{Energy} = \sqrt{\sum_{i,j=0}^{N_G-1} c_{i,j}^2} \tag{6}$$

$$\text{ASM} = \sum_{i,j=0}^{N_G-1} c_{i,j}^2 \tag{7}$$

Lastly, we combine all the six GLCM features from each of the four patches to create a feature vector of size 24.

2.2.4 Morphological feature extraction from the patches using Pretrained networks

Morphological features like nucleus size, shape, color, and tissue characteristics are equally important as textural features. Both morphological and textural properties will contribute to the complete set of features in the histopathology images that are used to detect cancer. The four chosen patches' morphological characteristics are extracted out and combined to create another feature vector. We used pretrained networks for the morphological feature extraction because they can extract the morphological features effectively [39]. Pretrained networks for feature extraction are deep neural networks that have been trained on large datasets such as Imagenet, CIFAR, or MNIST. Because they possess the ability to extract important and high-level features from raw input data and these networks are useful for a number of computer vision applications. There is a CNN in a pretrained network that includes hidden layers, convolutional layers, and max-pooling layers. Its weights, however, are not initially determined because they have already been initialised and optimised during the pre-training phase and some of its hyperparameters include the learning rate, batch size, and number of epochs. This work has used the pre-trained

models VGG 16, ResNet 50, Inception V3, Xception, and DensNet 121 to extract morphological features. All of the hidden layers in the VGG network use the ReLU activation function, and VGG networks take images of size 224 × 224 and here in this work they resize the 512 × 512 sized patches. Also, upsampling is used to meet the patch size requirement for VGG 16 if the patch size is less than 224 × 224. The VGG 16 has a large network with approximately 138 million total parameters and training the parameters requires more time [40]. Convolutional neural network (CNN) architecture with 50 levels is known as the ResNet 50 model. A residual neural network transforms a plain network into a residual network equivalent using skip connections. The convolution layers layered one above the other are analogous to the skip connections. Higher layers can function effectively because this link reduces the vanishing gradient issue in deeper architectures [41]. ResNet 50 uses 48 convolutional layers, one max pooling layer, and one average pooling layer. Edges, lines, and angles are morphological characteristics that can be extracted from images by ResNet 50's layers. A max-pooling layer is added, which decreases the operation's complexity while decreasing the dimension of the feature map. Xception is a 71-layer deep convolutional neural network with 36 convolutional layers forming the network's feature extraction base [42]. It is a depthwise separable convolution layer stack with residual connections. Inception V3 started as a GoogleNet module and is a convolutional neural network useful in image analysis and object detection. It is a 48-layer deep convolutional neural network and a widely used image recognition model with an accuracy greater than 78.1% on the ImageNet dataset [43]. The architecture of Inception V3 was designed to allow for deeper networks while also limiting the number of parameters to 25 million. Finally, among the DenseNet group models, the DenseNet 121 model is primarily designed for image classification [44]. Through a feed-forward connection, DenseNet CNN links every layer to every other layer. Each layer in DenseNet receives its feature maps from the layers before it, and all succeeding layers receive their feature maps as inputs. In addition, it minimizes the amount of parameters while improving feature propagation and reuse.

2.2.5 Feature fusion and feature selection

Concatenating the morphological features derived by the pretrained network and the textural features extracted by GLCM from the patches

creates a feature vector [45]. We employ a feature selection technique that is popular and frequently employed by researchers in order to choose the finest features from the feature vector and improve classification accuracy. As in [46], we also use extra tree classifier, also known as the extremely randomized trees classifier, for the finest feature selection and the feature vector's best features can be selected by using it [47]. Furthermore, earlier studies have shown the advantages of using extra-tree as the feature selection approach to improve classification accuracy in the white blood cell classification [47], prediction of Parkinson disease [48] and breast cancer detection [49]. In extra tree classifier, each decision tree in the extra trees forest is built from the original training sample. Then, at each test node, each tree is given a random sample of P features from the feature-set, from which each decision tree must choose the best feature to divide the data using some mathematical criterion, usually the Gini Index [50]. To perform the feature selection, each feature is ranked in descending order of its Gini Importance, and the the top P features [46] are selected.

2.2.6 Classification

In the proposed method, we present six different classification methods for the detection of OSCC. They are logistic regression (LR), random forest (RF), Naive Bayes (NB), support vector machine (SVM), K nearest neighbour (KNN) and voting classifier. The LR method is a supervised learning classification algorithm that uses just two classes to predict the probability of an output [51]. A predictive regression method called logistic regression is used to evaluate the connection between a binary dependent variable and one or more independent variables. It generates discrete results that range from 0 to 1. We choose LR as one of the classifiers because our problem is a binary classification challenge. Based on the Bayes theorem [52], naive bayes is a probabilistic classifier that can be applied to predictive modeling. The likelihood and conditional probability for each class are directly calculated from the training data. The calculations can then predict the input data using the Bayes theorem and the probability model. To calculate these probabilities efficiently, we presume the distribution is Gaussian (bell-shaped). We selected Bayes as the next classifier because it is fast and can be used to make real-time predictions. A supervised machine learning algorithm called a random forest is built on decision tree algorithms [53]. It is constructed using a combination of classifiers with a tree-like structure and decision trees,

and it is used in machine learning to handle regression and classification problems. Based on the predictions made by the decision trees, this algorithm provides an output or decision. We selected RF for our work since it works well with a large number of features and is widely used for automated cancer diagnosis using histopathology images. The SVM is a supervised learning technique that can classify data, detect outliers, and predict future results [54]. It is particularly helpful when performing classification in high-dimensional spaces, where a single hyperplane or group of hyperplanes are formed in a high or infinite-dimensional space [54]. The sample or data points that are nearest to the hyperplane are called support vectors, which act as a decision boundary between different data points. The classification feature of SVM is expanding its use in cancer diagnosis and, we selected SVM also in our work. The voting classifier, also called ensemble learning, is a type of machine learning that employs different classifier models to train and predict an output (class) based on the likelihood that the output will belong to the class with the highest probability [55]. Based on the voting criteria used by different classifiers, the result is predicted. A voting classifier can combine several weak learning classifiers to provide predictions that are more accurate than any one of the individual weak learning classifiers could produce on its own. It is an effective approach that could make an excellent option when a certain approach shows bias towards a specific component. So, along with LR, RF, NB, KNN and SVM, we selected voting classifier also in this work.

3 Results and discussion

The suggested framework is implemented in Python 3.7.6 and runs on an Intel(R) Core TM i3 -7130U 2.07 GHz CPU with 8 GB of RAM. It is essential to evaluate the precision and effectiveness of the proposed approach using a variety of performance indicators. Our proposed method was validated on a publicly available dataset for oral cancer [7]. A total of 1224 oral histopathology images were used for training and testing, including 290 normal and 934 OSCC images with a size of 2048 $\times$ 1536 in ".jpg" format. Table 1 shows the dataset's description [7].
Contrast enhancement is used as the pre-processing step to improve the input image contrast. Each input image is represented by four patches of size 512 $\times$ 512 pixels selected by the proposed patch selection algorithm based on high nuclear density. The number of images in both classes

Table 1: Description of the dataset.

Magnification	Class	No.of images
100×	Normal	89
	OSCC	439
400×	Normal	201
	OSCC	495

should be equal for accurate classification. Here, there is an imbalance between the number of normal and cancer images, and the result will lead to errors and misclassifications [46]. So, we augmented the total 1224 images using two popular data augmentation techniques, flipping and rotation, and the number increased from 1224 to 2000 for both classes. In this paper, we use images for testing and training in the ratio of 70:30.

3.1 Haar wavelet first level decomposition

For the textural feature extraction, each input image patch is decomposed by Haar wavelets to produce the four frequency components LL, HL, LH, and HH which is shown in Fig. 5.

3.2 Pretrained networks and feature selection

To extract morphological characteristics from the final four patches, we used five pretrained networks. The process of fine-tuning, in which we modify the parameters referred to as hyperparameters, is what largely determines how accurate the pretrained models are. Table 2 lists the hyperparameters, including batch size, momentum, learning rate, number of epochs, and input layer count. The optimizers used were Adam and Stochastic Gradient Descent (SGD) [56]. The study was conducted using a variety of parameters, and in this work, the optimal parameter values were selected. Table 3 shows the number of features extracted and the time taken for feature extraction by the five distinct pretrained networks. The

Table 2: Hyperparameters of pretrained networks.

	Learning rate	Momentum	No.of epochs	Batch size	No.of layers
VGG 16	0.001	0.9	200	64	16
ResNet 50	0.0001	0.9	200	64	50
Inception V3	0.05	0.9	200	64	48
Xception	0.01	0.9	200	64	36
DenseNet 121	0.1	0.9	200	64	24

least number of features are extracted by the VGG 16, and the most are extracted by ResNet 50 and Xception. In Table 4, the amount of features chosen and a comparison of feature extraction techniques are shown. In

Table 3: Comparison of different pretrained methods in terms of time for morphological feature extraction, and the number of features extracted.

	Number of extracted features	Time for feature extraction (Seconds)
VGG 16	25088	721
ResNet 50	100352	648
Inception V3	51200	143
Xception	100352	471
DenseNet 121	50176	577

Table 4: Comparison of the feature extraction methods and features selected by extra trees classifier.

Feature extraction	Number of selected features
GLCM+VGG 16	14140
GLCM+ ResNet 50	13423
GLCM+Inception V3	14128
GLCM+Xception	16530
GLCM+DenseNet 121	13090

Fig. 6, the feature maps of a patch created by five different pretrained networks are shown. The feature maps of a pretrained network capture the result of applying the filters to an input image. It is a function that maps a data vector to feature space. They represent the outputs of each layer in a neural network and capture different levels of complexity and abstraction of the input data. The purpose of visualizing a feature map for a particular input image or patch is to identify whether input features are identified or preserved in the feature maps. While feature maps near the output of the model capture more general features, those near the input of the model detect tiny or fine-grained detail [57]. Initial layers of the pretrained network identify low-level features such as colour and edges of nucleus, epithelium and tissue, while subsequent layers identify high-level features such as shape of tissue and nucleus. The final convolutional layer of the VGG 16 is widely used for visualisation because it has the tendency to capture complex visual patterns, as shown in Fig. 6(b). In contrast, ResNet 50's feature map in the final layer frequently represents high-level semantic information as illustrated in Fig. 6(c). Rich diverse visual features may be seen in Fig. 6(d) and are often captured by the feature maps from the final layers of Inception V3. A complex visual pattern and

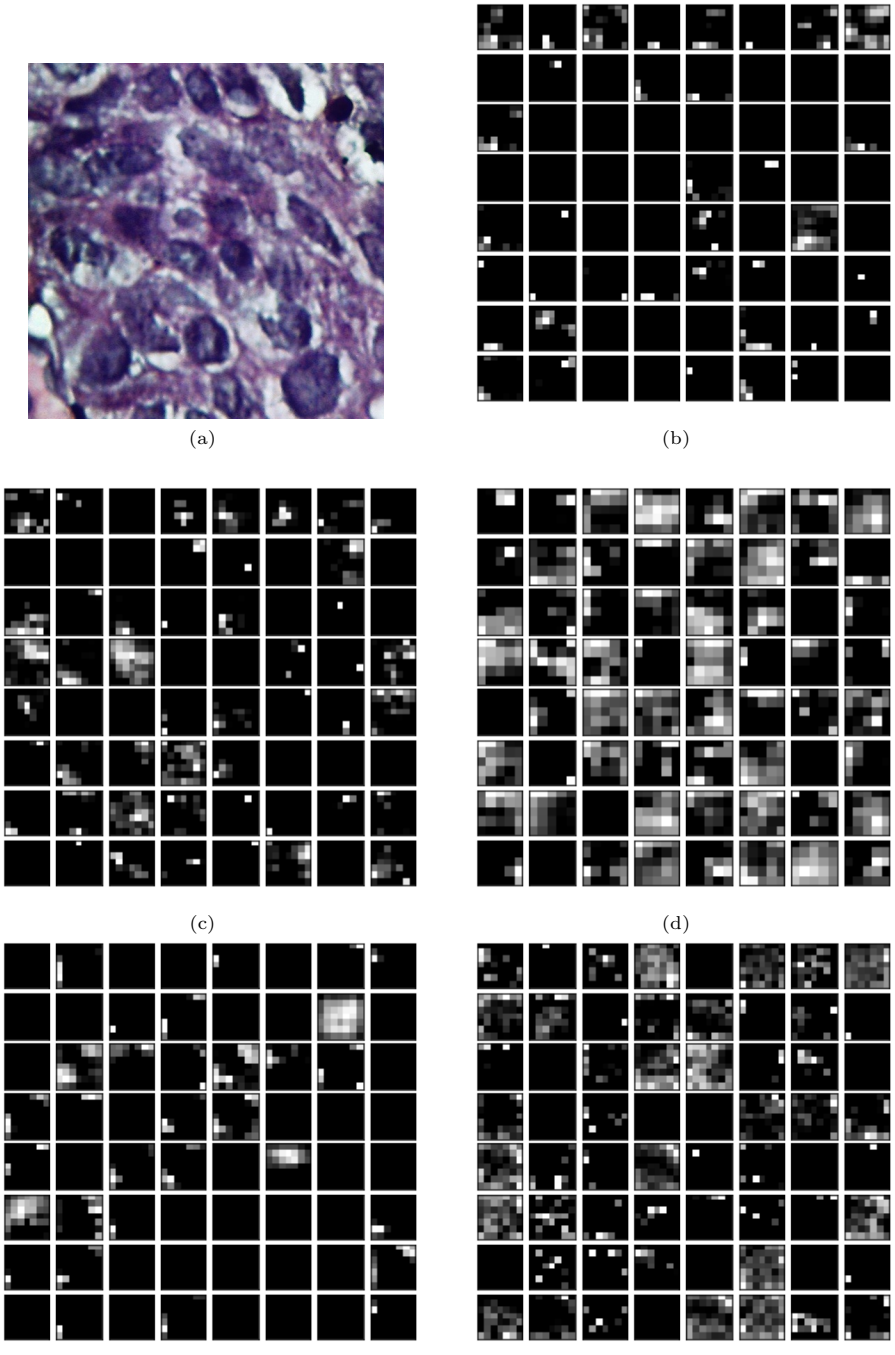


Fig. 6: 64 feature maps of an image patch using pretrained networks (a) The selected image patch (b) feature map of VGG 16, (c) feature map of ResNet 50, (d) feature map of Inception V3, (e) feature map of Xception, and (f) feature map of DenseNet 121.

semantic information can be represented using feature maps from the final layer of Xception, as seen in Fig. 6(e). It has been shown that the feature maps in the top layer of DenseNet 121 include highly abstract and integrated data, as illustrated in Fig. 6(f).

3.3 Assessment of classification schemes

In general, the confusion matrix, receiver operating characteristic (ROC) curves, the area under the ROC curve (AUC), classification accuracy, precision and sensitivity were used to assess the effectiveness of classifier models. In this work as well, we employed the same performance metrics listed in equations (8) to (10). A confusion matrix of size 2×2 can be used because the classification of oral cancer is a binary classification problem. In the confusion matrix, the number of images that were initially classified as cancerous and were subsequently predicted to be cancerous is known as the true positive (TP). True negative (TN) is the number of images predicted as normal and originally belonging to a normal class, while a false negative (FN) is the number of images predicted as normal but originally belonging to a cancerous class. The system for decision support won't be effective if there are more FNs in a confusion matrix. The classification accuracy, which is calculated as the ratio of the total number of correct predictions to the number of test images, is used to measure how well the suggested method performs. In Table 5, the confusion matrices for the six classifiers with GLCM and the five distinct pretrained feature extraction techniques are presented.
The classification accuracy can be calculated as [28],

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{8}$$

Along with the average classification accuracy, other performance metrics namely, sensitivity and precision are also used, and they are defined as follows.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad . \tag{9}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{10}$$

Table 6 summarises the six classifiers' accuracy, precision, and sensitivity. The voting classifier with ResNet 50 has fewer miss classifications than

the other five classifiers. That is, when compared to other combinations of pretrained networks and classifiers, the ResNet 50 with voting classifier model performed better in classifying oral histopathology images. The accuracy, precision, and sensitivity provided by the voting classifier with ResNet 50 in the proposed algorithm are 97.66 %, 98.00 %, and 97.35 %, respectively. ResNet 50 and the voting classifier work together to give robustness against noise and errors. ResNet's deep feature extraction capabilities helps in capturing meaningful representations and minimise the impact of noisy or irrelevant features. The voting classifier combines predictions from various models, allowing for error correction and lowering dependency on any single classifier. As a result, the classification findings are more powerful and accurate.

Table 5: Confusion matrix for different feature extraction methods and classifiers.

Feature extraction method	Classifier	TP	FN	FP	TN
GLCM+ VGG 16	LR	264	38	21	277
GLCM+ Resnet 50		270	32	20	278
GLCM+ Inception V3		265	37	45	253
GLCM+ Xception		261	41	46	252
GLCM+ DenseNet 121		273	29	34	264
GLCM+ VGG 16	RF	258	44	33	265
GLCM+ Resnet 50		247	55	40	258
GLCM+ Inception V3		243	59	45	253
GLCM+ Xception		248	54	55	243
GLCM+ DenseNet 121		256	46	40	258
GLCM+ VGG 16	NB	234	68	33	265
GLCM+ Resnet 50		247	55	40	258
GLCM+ Inception V3		182	120	47	251
GLCM+ Xception		219	83	66	232
GLCM+ DenseNet 121		275	27	109	189
GLCM+ VGG 16	SVM	263	39	44	254
GLCM+ Resnet 50		272	30	26	273
GLCM+ Inception V3		264	38	50	248
GLCM+ Xception		257	45	52	246
GLCM+ DenseNet 121		270	32	37	261
GLCM+ VGG 16	KNN	127	175	16	282
GLCM+ Resnet 50		266	36	41	257
GLCM+ Inception V3		194	108	62	236
GLCM+ Xception		224	78	77	221
GLCM+ DenseNet 121		238	64	69	229
GLCM+ VGG 16	Voting	255	47	13	285
GLCM+ Resnet 50		**294**	**8**	**6**	**292**
GLCM+ Inception V3		249	53	31	267
GLCM+ Xception		259	43	44	254
GLCM+ DenseNet 121		274	28	38	260

Table 6: Accuracy, precision and sensitivity for different classifiers LR, NB, RF, SVM, KNN and Voting classifier when GLCM and different pretrained networks are used for feature extraction.

Pretrained model	Clasifier	Accuracy (%)	Precision (%)	Sensitivity (%)
VGG 16	LR	90.16	92.63	87.41
	RF	87.16	85.43	85.71
	NB	83.16	87.64	77.48
	SVM	86.16	85.66	87.08
	KNN	68.16	88.88	42.05
	Voting	90.00	95.14	84.43
ResNet 50	LR	91.33	93.10	89.40
	RF	84.16	86.06	81.78
	NB	84.16	86.06	81.78
	SVM	90.83	91.27	90.06
	KNN	87.16	86.64	88.07
	Voting	**97.66**	**98.00**	**97.35**
Inception V3	LR	86.33	85.48	87.74
	RF	82.66	84.37	80.46
	NB	72.16	79.47	60.26
	SVM	85.33	84.07	87.41
	KNN	71.66	75.78	64.23
	Voting	86.00	88.92	82.45
Xception	LR	85.50	85.01	86.42
	RF	81.83	81.84	82.11
	NB	75.16	76.84	72.51
	SVM	83.83	83.17	85.09
	KNN	74.16	74.41	74.17
	Voting	85.50	85.47	85.76
DenseNet 121	LR	89.50	88.92	90.39
	RF	85.66	86.48	84.76
	NB	77.33	71.61	91.05
	SVM	88.50	87.94	89.40
	KNN	77.83	77.52	78.80
	Voting	89.00	87.82	90.72

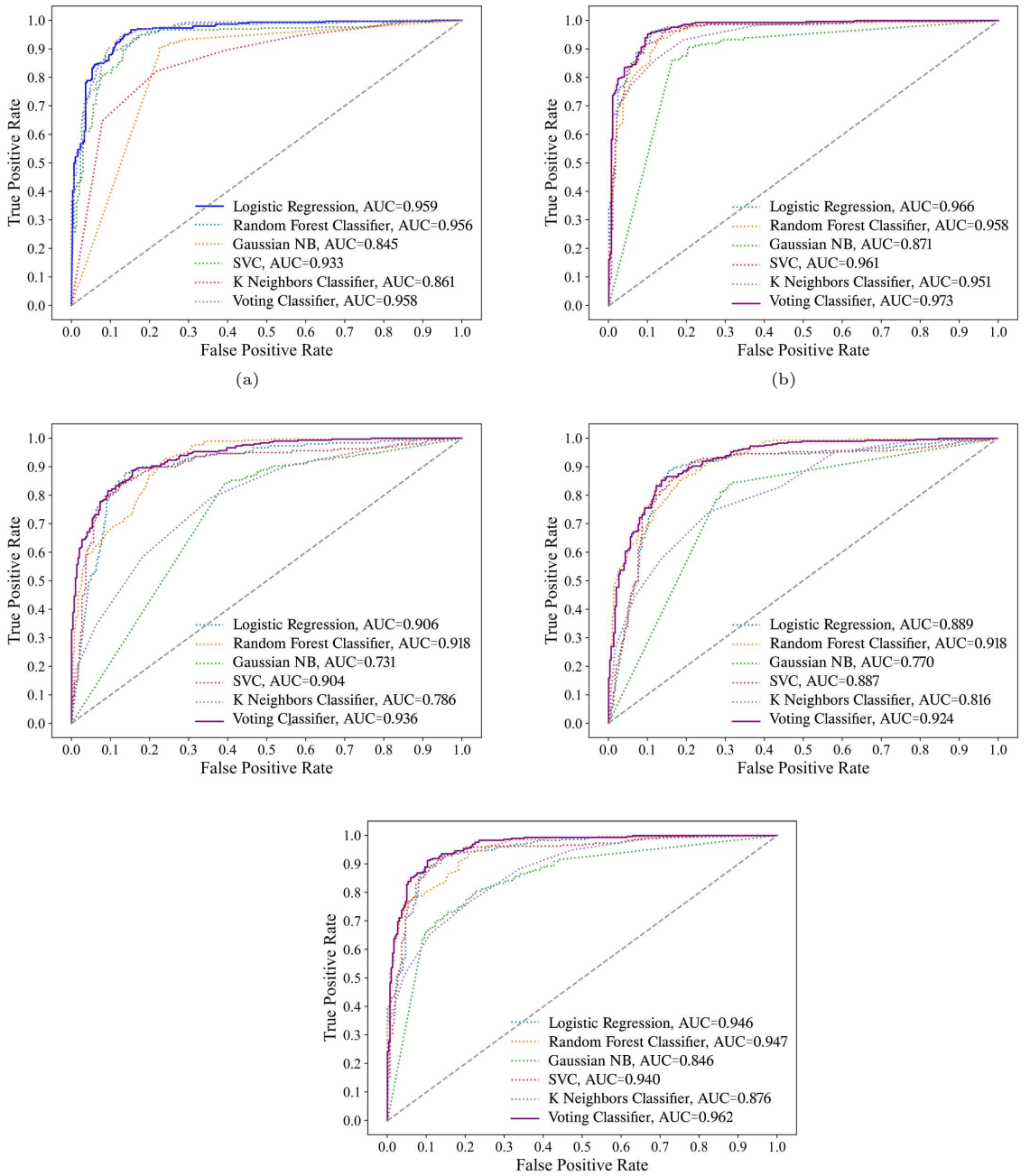


Fig. 7: ROC curves for different classifier performance (a) VGG 16, (b) ResNet 50, (c) Inception V3, (d) Xception, and (e) DenseNet 121.

3.4 ROC curves and AUC values for different classifiers

In Fig.7, the ROC curves are also plotted for the classifiers. The ROC analysis measures how well classifiers can distinguish between the two classes of normal and cancerous data [58]. The Area Under the Curve (AUC), which represents a summary of the ROC curve, is a measurement of a binary classifier's ability to distinguish between classes [58]. From the figure, it can be observed that the combination of the voting classifier with ResNet 50 outperforms the other five classifiers in terms of ROC and AUC values for oral cancer classification. The ResNet 50 use skip connections to help with deep feature extraction and to solve the vanishing gradient problem, which results in better feature representation. Additionally, the voting classifier can correct for biases to produce a more accurate and balanced classification. Therefore, the combination is appropriate to achieve high classification accuracy.

3.5 Comparison with existing algorithms

The performance metrics such as accuracy, precision and sensitivity of the proposed algorithm is compared with six other previously reported algorithms [10] [11] [18] [20] [21][22] [23][24] using the public OSCC dataset, in terms of accuracy, precision and sensitivity. The results are summarised in Table 7. Table shows that the proposed method, ResNet 50 with voting classifier outperforms the other eight algorithms in accuracy and precision for OSCC detection. However, compared to the other algorithms, the method using [23] XGBoost and BPSO has a high sensitivity of 98.90%. In contrast to sensitivity, the method's accuracy and precision are moderate. In this instance, it suggests that the feature selection process using BPSO is biased towards selecting features that are more relevant to the positive class, leading to an imbalance in the classifier's predictions.

3.6 Comparison of proposed algorithm for patches vs whole image

One of the contributions of this work is the proposed patch selection algorithm to improve computational efficiency and cancer detection accuracy. In this section, for the evaluation of the efficiency and accuracy of the proposed algorithm, instead of the image patches, the whole image is used in the algorithm and the results are compared with the case where image patches are used. For clarity, the stages of the algorithm when the whole

Table 7: Comparison of the proposed method with recent algorithms for oral cancer detection tested on oral cancer dataset [7].

Method	Model/Classifier	Accuracy (%)	Precision (%)	Sensitivity (%)
Feature extraction of temporal features [10]	SVM	91.00	89.40	98.23
	KNN	80.00	77.94	94.42
Empirical wavelet transform feature extraction [11]	Ensemble of ResNet 50 and DenseNet 201	92.00	92.80	100
Concatenated model [18]	VGG 16 and Inception V3 are concatenated	96.66	95.16	98.33
Multimodal fusion and transformers neural networks [20]	ResNetV2	89.22	90.80	89.60
	ResNetY	88.72	90.80	89.80
	PiT	91.64	93.00	92.00
	ViT	89.90	91.80	91.20
	CoaT	90.34	92.00	91.40
Classification by transfer learning method [21]	Mobile Net	85.48	81.00	79.00
	Large CNN	76.61	58.00	52.00
	ResNet 50	91.13	88.00	87.00
	Inception V3	89.52	89.00	82.00
Inception V3 feature extraction [22]	Inception V3-RF	91.00	89.30	92.92
Feature selection using BPSO [23]	XGBoost	96.30	96.30	**98.90**
	RF	93.10	93.30	97.80
	ANN	94.10	94.10	97.80
Duck pack optimization with deep learning [24]	Variational autoencoder (VAE)	97.28	95.51	97.46
Fusion of textural and morphological features from patches selected (proposed)	GLCM+ ResNet 50+LR	91.33	89.40	93.10
	GLCM+ ResNet 50+RF	88.00	85.09	90.49
	GLCM+ ResNet 50+ NB	84.16	81.78	86.06
	GLCM+ ResNet 50+ SVM	92.50	92.71	92.40
	GLCM+ ResNet 50+ KNN	87.16	88.07	92.40
	GLCM+ ResNet 50+ Voting	**97.66**	**98.00**	97.35

image is used are as follows. The images are first enhanced using CLAHE method. The enhanced image is then decomposed using the Haar wavelet transform. The textural and morphological features were then extracted using GLCM (from the HH, LL components of Haar wavelet decomposition) and pre-trained networks (from the enhanced image), respectively. Then, the feature vector is formed and from which the finest features are selected using the extra trees classifier. Finally, the images were classified into normal and OSCC. The comparison results are summarised in Table 8. From the table, it can be noted that the computation time when the whole images were used was 60% more than that when the image patches were used. More importantly, the classification performance is better when image patches are used instead of whole images. So, we conclude

that the use of image patches instead of the whole image will improve both the computational accuracy and performance of the algorithm.

Table 8: Performance comparison of image based and patch selection based classification.

Input	Accuracy (%)	Precision (%)	Sensitivity (%)	Execution Time (Seconds)
Whole image	95.21	95.24	92.90	18755
Image patches	97.66	98.00	97.35	11299

4 Conclusion

A large population of people suffering from oral cancer demands an accurate automated OSCC detection system. Numerous machine-learning techniques are used in the study of automatic OSCC detection using oral histopathology images. The proposed OSCC detection algorithm selects the patches containing important information about the morphology and nuclear density of tissues using a new patch selection algorithm. Textural and morphological characteristics were then extracted from the patches and the chosen features were trained by six well-known classifiers for 2000 images. From each input image, a number of 50% overlapping patches are extracted after contrast enhancement, and the finest four patches are chosen based on high nuclear density. For accurate feature extraction from the four patches, a feature fusion-based approach is suggested. The morphological features extracted by pretrained networks and the Haar wavelet-based GLCM textural features are used to accomplish this. Finally, from these features, finest features are chosen using an extra trees classifier. Images are divided into normal and OSCC categories after the feature has been trained using five various classifiers, including LR, RF, NB, SVM, KNN, and voting. The suggested algorithm with voting classifier and ResNet 50-based feature extraction exhibited noticeably improved accuracy, precision, and sensitivity, as demonstrated in the experiment results. The algorithm used in this study is reliable and precise for assisting oral pathologysts in the detection of oral cancer. The classification of biopsy images for OSCC diagnosis described in this work would therefore assist in the development of a reliable support tool for pathologists as it serves as an effective and inexpensive model.

Declarations

Competing interests
The authors declare no conflicts of interest.

Authors' contributions
Eranjoli Nalupurakkal Subhija developed the theoretical formalism, performed the data preparation, planned the experiments and performed the numerical simulations. Both Eranjoli Nalupurakkal Subhija and Vaninirappuputhenpurayil Gopalan Reju contributed to the final version of the manuscript and Vaninirappuputhenpurayil Gopalan Reju supervised the project.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not- for-profit sectors.

Availability of data and materials
The oral cancer dataset used in this study are publicly available which can be accessed at https://doi.org/10.17632/ftmp4cvtmb.1.

References

[1] Axell, T., Pindborg, J., Smith, C., Waal, I., Oral White Lesions, A.I.C.G.: Oral white lesions with special reference to precancerous and tobacco-related lesions. Journal of oral pathology & medicine. **25**(2), 49–54 (1996)

[2] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F.: Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. International journal of cancer. **136**(5), 359–386 (2015)

[3] Dhanuthai, K., Rojanawatsirivej, S., Thosaporn, W., Kintarak, S., Subarnbhesaj, A., Darling, M., Kryshtalskyj, E., Chiang, C.-P., Shin, H.-I., Choi, S.-Y.: Oral cancer: A multicenter study. Medicina oral, patologia oral y cirugia bucal. **23**(1), 23–29 (2018)

[4] Elmore, S.A., Cardiff, R., Cesta, M.F., Gkoutos, G.V., Hoehndorf, R., Keenan, C.M., McKerlie, C., Schofield, P.N., Sundberg, J.P., Ward, J.M.: A review of current standards and the evolution of histopathology nomenclature for laboratory animals. Journal of the Institute for Laboratory Animal Research. **59**(1), 29–39 (2018)

[5] Kumar, N., Gupta, R., Gupta, S.: Whole slide imaging in pathology: current perspectives and future directions. Journal of digital imaging. **33**(4), 1034–1040 (2020)

[6] Martin, J.W., Carballido, E.M., Ahmed, A., Farhan, B., Dutta, R., Smith, C., Youssef, R.F.: Squamous cell carcinoma of the urinary bladder: Systematic review of clinical characteristics and therapeutic approaches. Arab journal of urology. **14**(3), 183–191 (2016)

[7] Rahman, T.Y., Mahanta, L.B., Das, A.K., Sarma, J.D.: Histopathological imaging database for oral cancer analysis. Data in brief. **29**, 105114 (2020)

[8] Madabhushi, A.: Digital pathology image analysis: opportunities and challenges. Imaging in medicine. **1**(1), 7–10 (2009)

[9] Das, D.K., Chakraborty, C., Sawaimoon, S., Maiti, A.K., Chatterjee, S.: Automated identification of keratinization and keratin pearl area from in situ oral histological images. Tissue and Cell. **47**(4), 349–358 (2015)

[10] Bakare, Y.B., Kumarasamy, M.: Histopathological image analysis for oral cancer classification by support vector machine. International Journal of advances and image sciences. **7**(2), 1–10 (2021)

[11] Deo, B.S., Pal, M., Panigrahi, P.K., Pradhan, C.S.H.L.P. Asima: An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification. medRxiv., 1–14 (2022)

[12] Mustafa, R.A., Saleh, K.T., Chyad, H.S.: Feature extraction based on wavelet transform and moment invariants for medical image. IJERAT. **4**(8), 80–98 (2018)

[13] Haryanto, T., Pratama, A., Suhartanto, H., Murni, A., Kusmardi, K., Pidanic, J.: Multipatch-GLCM for texture feature extraction on classification of the colon histopathology images using deep neural network with gpu acceleration. Journal of Computer Science. **16(3)**, 280–294 (2020)

[14] Ojala, T., Pietikainen, M., Harwood, D.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. Proceedings of 12th international conference on pattern recognition, IEEE. **1**, 582–585 (1994)

[15] Mohanaiah, P., Sathyanarayana, P., GuruKumar, L.: Image texture feature extraction using GLCM approach. International journal of scientific and research publications. **3**(5), 1–5 (2013)

[16] Bhagat, P., Choudhary, P., Singh, K.M.: A comparative study for brain tumor detection in MRI images using texture features. Sensors for health monitoring. **5**, 259–287 (2019)

[17] Das, N., Hussain, E., Mahanta, L.B.: Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. Neural Networks. **128**, 47–60 (2020)

[18] Amin, I., Zamir, H., Khan, F.F.: Histopathological image analysis for oral squamous cell carcinoma classification using concatenated deep learning models, cold spring harbor laboratory press. medRxiv., 1–18 (2021)

[19] Abbas, A., Mahmood, S.N.: Hyperparameter tuning bidirectional gated recurrent unit model for oral cancer classification. Computers, Materials & Continua. **73(3)**, 4541–4557 (2022)

[20] Lima, L.M., Assis, M.C.F.R., Soares, J.P., GroVelloso, T.R., Barros, L.A.P., Camisasca, D.R., Krohling, R.A.: On the importance of complementary data to histopathological image analysis of oral leukoplakia and carcinoma using deep neural networks. Intelligent Medicine. (2023)

[21] Palaskar, R., Vyas, R., Khedekar, V., Palaskar, S., Sahu, P.: Transfer learning for oral cancer detection using microscopic images. arXiv., 1–8 (2021)

[22] Abdul Rauf, A.R., Mohd Isa, W.H., Khairuddin, I.M., Mohd Razman, M.A., Arzmi, M.H., PP Abdul Majeed, A.: The classification of oral squamous cell carcinoma (OSCC) by means of transfer learning. Robot Intelligence Technology and Applications. **6**, 386–391 (2022)

[23] Deif, M.A., Attar, H., Amer, A., Elhaty, I.A., Khosravi, M.R., Solyman, A.A., et al.: Diagnosis of oral squamous cell carcinoma using deep neural networks and binary particle swarm optimization on histopathological images: An artificial intelligence of medial things (AIoMT) approach. Computational Intelligence and Neuroscience. **2022** (2022)

[24] Shetty, S.K., Patil, A.P.: Duck pack optimization with deep transfer learning-enabled oral squamous cell carcinoma classification on histopathological images. International Journal of Grid and High Performance Computing (IJGHPC). **15**(2), 1–21 (2023)

[25] Ahmad, N., Asghar, S., Gillani, S.A.: Transfer learning-assisted multi-resolution breast cancer histopathological images classification. The Visual Computer. **38**(8), 2751–2770 (2022)

[26] Subhija E, N., Reju, V.G.: A feature fusion based detection of oral squamous cell carcinoma with morphological and textural features in Haar wavelet transform space. International Conference on Next Generation Intelligent Systems (ICNGIS). (2022)

[27] Morikawa, T., Shibahara, T., Nomura, T., Katakura, A., Takano, M.: Non-invasive early detection of oral cancers using fluorescence visualization with

optical instruments. Cancers. **12**(10), 2771 (2020)

[28] Di Lillo, A., Motta, G., Storer, J.A.: Texture classification based on discriminative features extracted in the frequency domain. IEEE International Conference on Image Processing. **2**, 11–53 (2007)

[29] Bhattacharjee, S., Kim, C.-H., Park, H.-G., Prakash, D., Madusanka, N., Cho, N.-H., Choi, H.-K.: Multi-features classification of prostate carcinoma observed in histological sections: analysis of wavelet-based texture and colour features. Cancers. **11**(12), 1937 (2019)

[30] Kiaee, N., Hashemizadeh, E., Zarrinpanjeh, N.: Using GLCM features in Haar wavelet transformed space for moving object classification. IET Intelligent Transport Systems. **13**(7), 1148–1153 (2019)

[31] Rao, K., Bansal, M., Kaur, G.: Retinex-centered contrast enhancement method for histopathology images with weighted clahe. Arabian Journal for Science and Engineering. **47(11)**, 1–18 (2022)

[32] Sengoz, N., Yigit, T., Ozmen, O., Isik, A.H.: Importance of preprocessing in histopathology image classification using deep convolutional neural network. Advances in Artificial Intelligence Research. **2**(1), 1–6 (2022)

[33] Amsini, P., Rani, R.U.: Enhanced type 2 triangular intuitionistic Fuzzy C means clustering algorithm for breast cancer histopathology images. Fourth international conference on computing methodologies and communication (ICCMC)., 589–594 (2020). IEEE

[34] Saha, M., Chakraborty, C., Racoceanu, D.: Efficient deep learning model for mitosis detection using breast histopathology images. Computerized Medical Imaging and Graphics. **64**, 29–40 (2018)

[35] Nagase, A., Takahashi, M., Nakano, M.: Automatic calculation and visualization of nuclear density in whole slide images of hepatic histological sections. Bio-Medical Materials and Engineering. **26**, 1335–1344 (2015)

[36] Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. IEEE Transactions on systems, man, and cybernetics. (6), 610–621 (1973)

[37] Stankovic, R.S., Falkowski, B.J.: The Haar wavelet transform: its status and achievements. Computers & Electrical Engineering. **29**(1), 25–44 (2003)

[38] Gupta, D., Choubey, S.: Discrete wavelet transform for image processing. International Journal of Emerging Technology and Advanced Engineering. **4**(3), 598–602 (2015)

[39] Kensert, A., Harrison, P.J., Spjuth, O.: Transfer learning with deep convolutional neural networks for classifying cellular morphological changes. SLAS Discovery: Advancing Life Sciences R&D. **24**(4), 466–475 (2019)

[40] Rezende, E., Ruppert, G., Carvalho, T., Theophilo, A., Ramos, F., Geus, P.d.: Malicious software classification using VGG16 deep neural network's bottleneck features. Information Technology-New Generations: 15th International Conference on Information Technology., 51–59 (2018)

[41] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition., 770–778 (2016)

[42] Chollet, F.: Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition., 1251–1258 (2017)

[43] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition., 1–9 (2015)

[44] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition., 4700–4708 (2017)

[45] Theodorakopoulos, I., Kastaniotis, D., Economou, G., Fotopoulos, S.: Hep-2 cells classification via fusion of morphological and textural features. 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)., 689–694 (2012). IEEE

[46] Desir, C., Petitjean, C., Heutte, L., Salaun, M., Thiberville, L.: Classification of endomicroscopic images of the lung based on random subwindows and extra-trees. IEEE Transactions on Biomedical Engineering. **59**(9), 2677–2683 (2012)

[47] Baby, D., Devaraj, S.J., Hemanth, J., *et al.*: Leukocyte classification based on feature selection using extra trees classifier: Atransfer learning approach. Turkish Journal of Electrical Engineering and Computer Sciences. **29**(8), 2742–2757 (2021)

[48] Nahar, N., Ara, F., Neloy, M.A.I., Biswas, A., Hossain, M.S., Andersson, K.: Feature selection based machine learning to improve prediction of parkinson disease. Brain Informatics: 14th International Conference, BI 2021, Virtual Event, Proceedings 14., 496–508 (2021)

[49] Sharma, D., Kumar, R., Jain, A.: Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. Measurement: Sensors. **24**, 100560 (2022)

[50] Lovász, L., Naor, M., Newman, I., Wigderson, A.: Search problems in the decision tree model. SIAM Journal on Discrete Mathematics. **8**(1), 119–132 (1995)

[51] Peng, C.-Y.J.: An introduction to logistic regression analysis and reporting. Educational Research. **3**(10), 3–14 (2002)

[52] Berrar, D.: Bayes' theorem and Naive Bayes classifier. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics. **403**, 412 (2018)

[53] Liu, Y., Wang, Y., Zhang, J.: New machine learning algorithm: Random forest. International Conference on Information Computing and Applications., 246–252 (2012)

[54] Xue, H., Yang, Q., Chen, S.: Svm: Support vector machines. The top ten algorithms in data mining., 51–74 (2009)

[55] Dietterich, T.G.: Ensemble methods in machine learning. International workshop on multiple classifier systems. **1857**, 1–15 (2000)

[56] Hung, T.V., Viet, V.Q., Van Thuat, D.: A deep learning-based procedure for estimation of ultimate load carrying of steel trusses using advanced analysis. Journal of Science and Technology in Civil Engineering (STCE)-HUCE . **13(13)**, 113–123 (2019)

[57] Chu, J.L., Krzyżak, A.: Analysis of feature maps selection in supervised learning using convolutional neural networks. Advances in Artificial Intelligence: 27th Canadian Conference on Artificial Intelligence. **27**, 59–70 (2014)

[58] Hajian-Tilaki, K.: Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian journal of internal medicine. **4**(2), 627–635 (2013)