Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Capsule network based analysis of histopathological images of oral squamous cell carcinoma

Santisudha Panigrahi [a,*], Jayshankar Das [b], Tripti Swarnkar [c]

[a] Department of Computer Science and Engineering, SOA Deemed to be University Bhubaneswar, 751030 Odisha, India
[b] Centre for Genomics and Biomedical Informatics, Institute of Medical Sciences and SUM Hospital, SOA Deemed to be University Bhubaneswar, 751030 Odisha, India
[c] Department of Computer Application, SOA Deemed to be University Bhubaneswar, 751030 Odisha, India

ARTICLE INFO

Article history:
Received 4 August 2020
Revised 23 October 2020
Accepted 4 November 2020
Available online 12 November 2020

Keywords:
Histopathology
Deep learning
Convolutional neural network
Capsule network
Oral cancer

ABSTRACT

Oral cancer is one of the most prevalent malignancy affecting oral cavity. Determining the correct type of oral cancer at the early stages is important in designing a detailed treatment plan and predicting the response of the patient to the treatment being adopted. A major challenge lies in the detection of oral cancer from histopathological images. In oral malignancy diagnosis, the main visual features are generally extracted from the architectural differences of epithelial layers and the appearance of keratin pearls. This paper proposes a new approach for classifying oral cancer using a deep learning technique known as capsule network. Dynamic routing and routing by agreement of capsule network makes it more robust for rotation and affine transformation of augmented oral dataset. This network's capability of handling pose, view and orientation makes it suitable for analysis of oral cancer histopathological images at an early stage. The performance of cross-validation indicate that the proposed methodology can efficiently classify the histopathological images of Oral Squamous Cell Carcinoma (OSCC) with 97.78% sensitivity, 96.92% specificity and 97.35% accuracy.

1. Introduction

Oral cancer is seen worldwide as the most revealed malignancy in cancer related morbidity and mortality, and is coming under the head and neck section (Kumar et al., 2016). Various image processing techniques are extensively used for the early detection of oral cancer which leads to greater treatment efficiency and increase in cancer survival rate. Determining the correct type of oral cancer at an early stage is a significantly challenging task. Therefore computer-aided applications will be very beneficial as it assists the physician to make a more detailed treatment strategy and has an enhanced prediction of patient's response to the accepted treatment.

Additionally, the classification of oral cancer by human assessment is an exceptionally time consuming procedure and vulnerable to mistake, which mainly dependent on the pathologist's knowledge and proficiency (Bui et al., 2019). With the advent of new technologies there is a demand for accurate automated diagnostic system for oral cancer using advance image processing techniques. Among the various types of imaging modalities, histopathological imaging is the more suitable for oral cancer prognosis due to the advantage of its rapid processing time, less equipment requirement, and less need for ventilation in the laboratory (Brian, 2017). Thus in this paper we focus on the histopathological imaging for the oral cancer classification.

Deep learning techniques, especially convolutional neural network (CNNs) have turn into the state-of-the-art for numerous tasks to analyze images in last few years (Krizhevsky et al., 2012). CNN of deep learning works in an ideal way for image problems by doing the simultaneous work of feature extraction and classification. In the case of archeological excavations some fragments of full object is classified to know certain features which define it then classify the complete object. A soft set decision tree method is proposed by Woźniak and Połap (2020), which classifies the fragment image depending upon the information available to the particular class. For the Bacteria shape classification an input microscopy image is segmented by region covariance model (Połap and Woźniak, 2019). Then segments are provided to CNN for recognition of

* Corresponding author.
E-mail address: santisudha.nanda@gmail.com (S. Panigrahi).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2020.11.003
1319-1578/© 2020 The Authors. Published by Elsevier B.V. on behalf of King Saud University.

visible bacteria strains. CNN's implicit feature engineering and faster learning method allows the prediction /prognosis of histopathological images in the cancer domain to be outperformed. LeNet (LeCun et al., 1989), AlexNet (Krizhevsky et al., 2012), ZFNet (Zeiler and Fergus, 2014), VGGNet (Simonyan and Zisserman, 2014), GoogleNet (Szegedy et al., 2015) are the desired model for the countless tasks of medical image analysis. The latest challenges in the area of medical imaging is to use these approaches to tackle the classification issue. Since the deep learning efficiency depends primarily on finding an architecture that meets the challenge, many researches are currently going on to design new and more complex deep networks to increase the predictable outcomes. This results in the implementation of a high number of hyperparameters, making the overall network too complex for optimization (Feurer and Hutter, 2019)

Although the CNNs demonstrate considerable versatility and efficiency in a varied array of tasks in computer vision, they come up with their own limitations. In CNNs the neurons are scalar and additive and there is no spatial relationships between neurons at a specific network layer within the kernel of the previous layer. The maxpooling of CNN losses valuable information and also does not encode relative spatial relationships between features (Bae and Kim, 2018) Because of this, CNN are not invariant to large transformations of input data. Recently, Sabour et al. (2017) proposed the concept "capsule network", where the neuron-level information is represented not as scalars but as vectors. Such vectors contain the information as follows.

Spatial relationship
Magnitude\prevalence, and
Other attribute of the extracted feature

Capsule networks are new machine learning architectures introduced to better model the hierarchical pose relationships. This comprises of capsules that are a collection of neurons which represents an object's instantiation parameters such as the pose and orientation (Hinton et al., 2011). Hence by using a dynamic routing algorithm these capsules are "routed" to capsules in the next layer that determines the agreement among these capsule vectors. It establishes a part-to-whole relationship which does not occur in regular CNNs.

This work aims at using these capsule networks which mainly based on the dynamic routing algorithm to perform the classification task. These capsules can be used effectively for very precise and efficient classification (Patrick et al., 2019). To show the efficiency of the capsules we have taken the histopathological images of oral cancer.

2. Related work

The advancement of the deep learning techniques along with computation capability have provided for automated image classification systems. An attempt has already been made to apply deep learning methods to histopathological images of oral cancer. A deep learning model called convolutional neural network (CNN) was found to be of significant use in task (Dev Kumar et al., 2018) enabling to achieve 96.88% accuracy for the different layers (epithelial, subepithelial, and keratin and keratin pearl) of oral mucosa by segmentation.

The accuracies obtained by convolutional neural network for various image analysis tasks are quite high, because it does not need explicit feature extraction approaches before the classifier is trained. This is the advantage that deep learning algorithms have been the winner in various image processing competitions. Major studies using machine learning techniques are also carried out on histological image analysis of oral cancer. Krishnan et al. (2012a) used models of SVM and Gaussian mixture for the segmentation of handcrafted features and classification of oral mucosa. Krishnan et al. (2012b) used different classifiers such as fuzzy, Decision tree, Gaussian mixture model, KNN for the classification task and found that fuzzy gives promising performance because it can manage imprecise non-linear parameters when supplied with a mixtures of texture features. Again Krishnan et al. (2011) used SVM classifier for the combination of different features of wavelet family which gives better accuracy compared to the individual features. Belvin et al. (2013) proposed a model for more illustrative and patient specific approach with ANN-based multi-class classification of oral cancer. All these models are based on machine learning algorithms, which requires extraction of hand crafted features and unable to handle complex queries. So CNN of deep learning approach was used to overcome these issues. But these CNN are not able to capture the pose, view and orientation of the input images. We have opted for capsule network which can overcome the limitations of CNN and give better prediction accuracy.

The literature study provides significant examples of what can be achieved with machine learning/deep learning for oral cancer. Within the medical domain several capsnet models are used for various cancer classification tasks. For brain tumor classification (Afshar et al., 2018) the authors have taken Magnetic Resonance Imaging (MRI) images and proved the efficiency of capsnet model in comparison to CNN model by achieving accuracy of 86.56% with one convolutional layer and 64 feature maps. Iesmantas and Alzbutas (2018) have applied the convolutional capsnet architecture on breast cancer by using a total of 400 images to classify four categories and obtained accuracy of 87%. For lung cancer screening (Mobiny and Van Nguyen, 2018), the authors have proposed a consistent dynamic routing mechanism which improves the speed by 3times of the original capsnet. Wang et al. (2019) have proposed a capsnet model, "protein post-translational modification site prediction" which outperformed the baseline convolutional neural network with small training data.

The literature reveals the importance of Capsnet model used for various disease prediction tasks. In this study a categorization of binary problem is considered: benign and malignant of oral histopathological images. In the section Materials and Methods, we address the dataset, pre-processing steps and a comprehensive demonstration of capsule network (CapsNet)–a different kind of deep learning network.

3. Materials and methods

The current study has been conducted for the classification of oral biopsy images into benign or malignant. Fig. 1 represents the diagrammatic illustration of the suggested method carried out in five stages. Stage I consists of pre-processing of the image to reduce the noise. Stage II and stage III consists of image segmentation and image augmentation. Data partitioning is carried out for k-fold cross validation in the stage 4. Stage 5 is the final stage which involves the classification of the sample images into benign and malignant by the capsule network. The whole slide image of 82 malignant and 68 benign are collected from **GDC portal**[1], which are from various oral cancer sites such as oesophagus, larynx, lower and upper lips, vocal cord, floor of mouth and floor of tongue etc. For the training and testing of the Capsule Network model, the region of interest (ROI) of size 256 × 256 pixel patches were extracted from each whole slide image for precise segmentation. Non-overlapping image patches were considered. Prediction was done for the image

[1] https://portal.gdc.cancer.gov/repository.

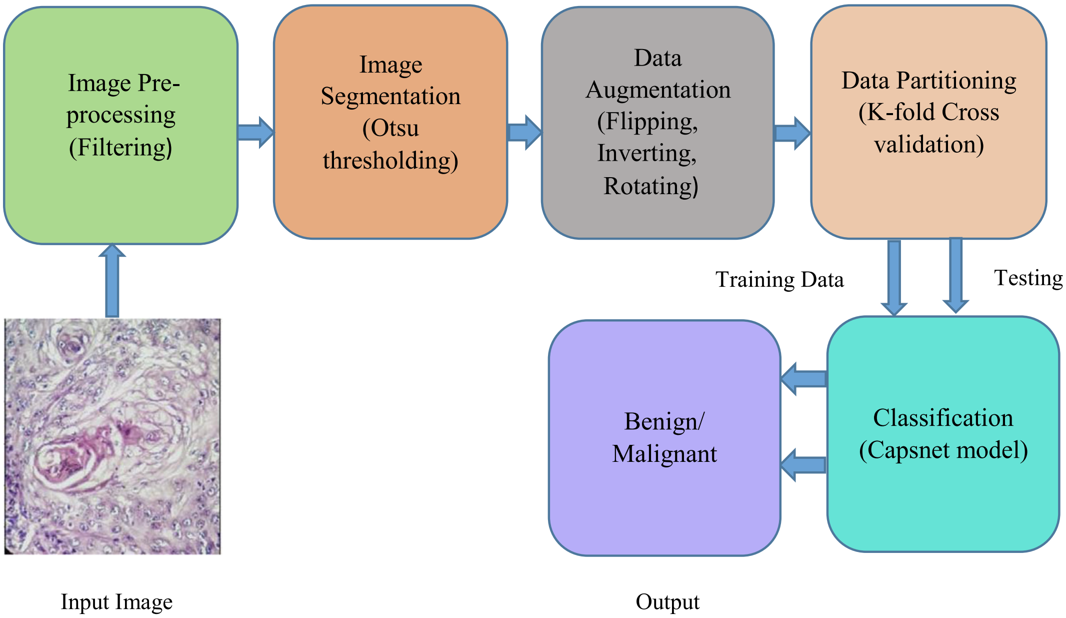
Image Pre-processing (Filtering)
Image Segmentation (Otsu thresholding)
Data Augmentation (Flipping, Inverting, Rotating)
Data Partitioning (K-fold Cross validation)
Training Data
Testing Data
Benign/ Malignant
Classification (Capsnet model)
Input Image
Output

Fig. 1. Workflow of proposed methodology.

i.e. accuracies were analysed only for the whole image not for the individual patches (Wei et al., 2019). The label of the images were determined by majority voting.

3.1. Image pre-processing

Image pre-processing has been done for noise reduction. The collected images varied in quality and dimension. Some images have a perfect, unvarying background where as others have clustered background. Pre-processing of images is required to remove the impurities present in it and make it appropriate for further process such as segmentation, feature extraction etc. For this study image pre-processing involves applying Gaussian blur to filter the noise present in the image and smoothens the edges then converted to gray-scale image. Thus, Gaussian blurring efficiently eliminates the noise existing in the low SNR (Signal-to-noise ratio) images (Gedraite and Hadad, 2011).

3.2. Image segmentation

Precise segmentation carries the actual information which lead to the accurate classification of histopathological images in an automatic diagnostic system. Automatic segmentation for the proposed study consists of the separation of relatively homogeneous artifacts such as nucleus and keratin pearl area, epithelium as foreground feature and other cellular structure such as cytoplasm and other connective tissue as background. In such problem, a binary segmentation task is required which can use a single threshold to distinguish the two classes shown in Fig. 2 (c). For our analysis, we implemented Otsu's thresholding as technique for image segmentation depending on the pixel values color distribution. This method is widely used due to the simplicity and effectiveness (Qu and Zhang, 2010). The goal of Otsu's thresholding is to determine a threshold value by iteration of all reasonable threshold values to characterize pixels falling either as background (red pixels) or foreground (blue pixels) (Otsu,1979). Fig. 2(d), shows that Otsu's approach achieves segmentation by reducing intra-class variance (among blue and any intensity of blue pixels) and optimizing the inter-class variance (among any intensity of red and blue pixels) (Das et al., 2020)

3.3. Data augmentation

Adequate amount of training data should be available to train a deep learning model. If small quantities of training data are used, there may be overfitting. To prevent the problem of overfitting due to small amount of data and the class category inequality, we applied data augmentation. Usually, deep learning models generalize well for balanced and large dataset. This study has been done for comparatively small dataset. The publicly available oral cancer dataset is small and for both malignant and benign categories, samples are not balanced. Thus, with limited augmentation, a small balanced dataset has been generated for this study.

The images acquired by the microscope are in-variant direction and the sharpness of the cell structure differs for each image based on the position of microscope's focal plane. This augmentation is the image manipulation technique done by flipping, inverting and rotating the original image. Thus, our Capsnet model takes 1000 oral cancer images of two types (Benign and Malignant). Table1 abbreviates the number of images from each class.

3.4. Data partitioning

In this study, a 10-fold cross-validation approach was used to estimate the performance of the prediction models. This method involves splitting the training data into approximately equal size k groups, or folds. Thereafter, training of the algorithm has been carried out on the k-1 folds and the remaining one fold is taken as the validation fold for evaluation of the algorithm. The method is repeated for all k folds until the algorithm is validated. Experimental studies have shown that 10 tends to be an optimum number of folds which reduces the bias and variance related with the validation process (Kohavi, 1995). In 10-fold cross-validation the whole dataset is split into 10 mutually exclusive subsets or folds. Every fold is used once to assess the model's prediction performance, generated from the combined data of the remaining nine folds, resulting in 10 independent performance estimates. This approach allocate all patients to a validation fold exactly once, thus can be used for cross-validation prediction.

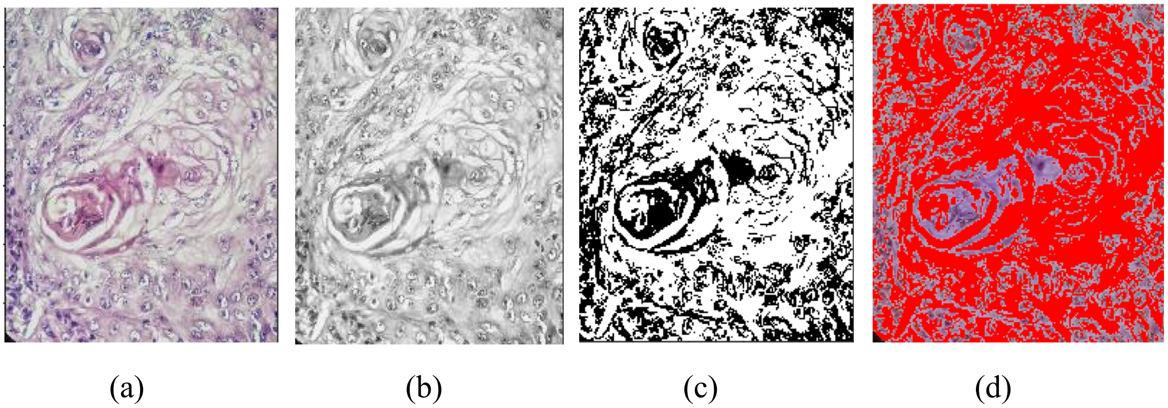

Fig. 2. (a) Original image; (b) Filtered gray scale image; (c) Threshold image; (d) Otsu's Segmented image to separate color pixels.

Table 1
Dataset of oral cancer considered for the study.

Category	Whole slide	Patches	Augmented
Malignant	82	414	1000
Benign	68	372	1000

3.5. Classification by capsule network

Capsules consists of a set of neurons whose outputs are inferred as different characteristics of the same entity, and forms the activation vector. Each capsule consists of a pose matrix which represents the presence of a specific object located at a given pixel, and an activation probability which represents the length of the vectors. The activation vector's direction collects the pose information of the object, such as position and orientation, whereas the activation vector length or magnitude measures the approximate likelihood of an object of interest being present. Upon rotating an image for example, the activation vector will also change accordingly however its length will remain the same. Fig. 1 shows the routing of data across layers by CapsNets. There may be numerous capsule layers. In our proposed architecture, we have used a primary capsule layer (reshaped and squashed output of the last convolutional layer) and CancerCaps layers (i.e. capsules signifying 2 categories of histological images: cancerous/non-cancerous). Each capsule predicts the parent capsule's output and if this prediction is consistent with the parent capsule's actual output, then the coupling co-efficient among these two capsules increases. If $\mathrm{u_i}$ is the output capsule *i*, its prediction for parent capsule *j* is determined as in equ (1).

$$\widehat{u}_{j/i} = W_{ij}\, u_j \tag{1}$$

where $\widehat{u}_{j/i}$ is the output prediction vector of the jth capsule, and $\mathrm{W_{ij}}$ is the weight matrix which should to be learned in the backward pass. The *softmax* function is used to calculate the coupling co-efficient $\mathrm{c_{ij}}$ depending upon the degree of conformation among the capsules in the layer below and the parent capsules known as "iterative dynamic routing process " as shown in Eq. (2).

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{2}$$

b_{ij} represents the log probability and set to 0, if capsule i should be coupled with capsule j initially by agreement process at the start of the routing. So the parent capsule input vector j is computed as in Eq. (3).

$$s_j = \sum_i c_{ij}\widehat{u}_{j/i} \tag{3}$$

Lastly a non-linear squashing function is used to normalize the output vectors of capsules by preventing them surpassing 1. Its length can thus be represented as the probability that a capsule will detect a given feature. Every capsule's final output is determined by its initial vector value as shown in Eq. (4).

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2}\frac{s_j}{\|s_j\|} \tag{4}$$

where s_j is the total input to capsule j and $\mathrm{v_j}$ is the output. Based on the agreement among $\mathrm{v_j}$ and $\widehat{u}_{j/i.}$ the log probabilities must be updated during routing. Therefore the updated log probability is calculated as in Eq. (5).

$$b_{ij} = b_{ij} + \widehat{u}_{j/i}.v_j \tag{5}$$

The routing co-efficient will be increased by dynamic routing mechanism to j-parent Capsule with a factor of $\widehat{u}_{j/i}.\mathrm{v_j}$. Consequently more information can be sent by a child capsule to the parent capsule whose performance $\mathrm{v_j}$ is further comparable to its prediction $\widehat{u}_{j/i}$

3.5.1. Proposed capsnet architecture

The following Fig. 3: shows the proposed Capsnet model for classifying oral cancer.

Before the primary capsule layer, one might use numerous convolutional layers if it needs. In Capsnet the max-pool layers are omitted, which was the main drawback of CNN architecture. Because of max-pool layer, CNN was losing some valuable information and the spatial relationship from the image. So Capsnet uses convolution with strides greater than 1 to minimize the dimensionality (if the stride is 2, then dimensions are decreased by the factor of 2, etc.) (Szegedy et al., 2015). The CancerCaps output is used to decide on the input image category. In Fig. 3 the entire network architecture used for this is shown. The total number of trainable parameters are 14788864.

Adam optimizer a stochastic gradient descent algorithm is used to train the whole network with parameter 0.0001 (Kingma and Ba, 2014) Adaptive methods of gradient descent are prevalent for various reasons. Firstly they adapt a learning rate for every parameter and able to learn sparse as well as more dense information. Second, they allow the learning rate to be learnt from data reducing the tuning of learning rate. Third, they tend to approach a convergence much earlier in the training scheme compared to non-adaptive methods for the same data and methods.

Each capsule in the primary capsule layer is linked to every other capsule in the CancerCaps layer. Nevertheless, instead of max-pooling an algorithm called routing-by-agreement is used which facilitates better learning (Sabour et al., 2017)

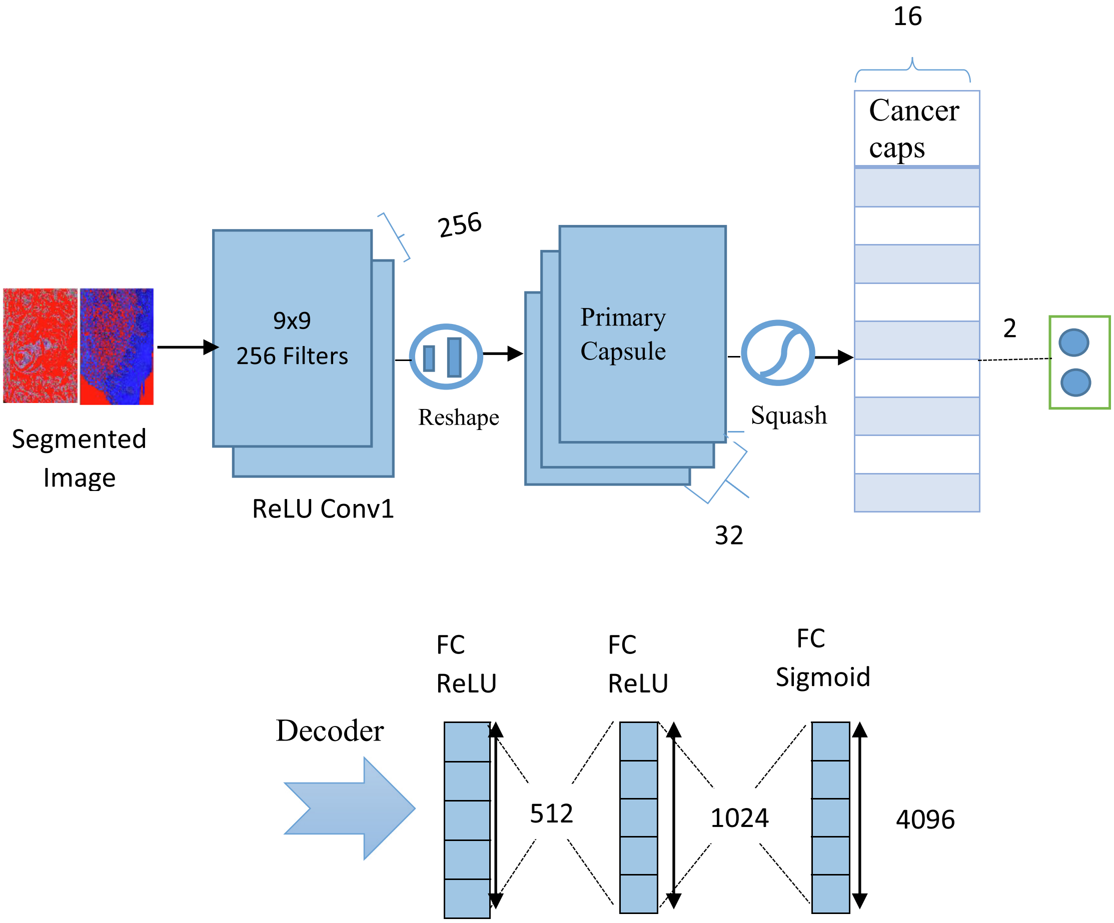

Fig. 3. Capsnet architecture used for classifying histopathological images of oral cancer.

The architecture used to diagnose oral cancer involves the input and initial convolution layers as portion of primary capsules followed by the cancer capsules.

Input Layer: This layer accepts the histopathological images of oral cancer for training the network.

Primary Capsule Layer: The input layer is first connected to a convolutional layer. It consists of 256 filters of kernel size 9 containing scalars. ReLU, the Rectified Linear Unit is used as the non-linear activation function. Output is reshaped from primary capsules to get 32 × 6 × 6 feature maps containing 8-dimentional vectors. Using a novel squashing function the output vectors must be squashed to ensure these vectors have a length between 0 and 1 (to represent a probability). This gives the output of primary capsule. During training the squash function is supplemented with a small epsilon value to evade the problem of vanishing gradient (Faizan, 2017).

Cancer Caps Layer: To determine the predicted output vectors of cancer capsules, the routing by agreement algorithm is implemented for each and every primary and cancer caps pair.

3.5.2. Reconstruction

A decoder network comprises of fully connected layers which is added to the cancer capsule network provided by tuning cancer capsnet output to reconstruct the input images. This mechanism would allow the network retain the information needed to reconstruct the cancer images throughout the network. This acts as regularization and prevents the data being over-fitted and helps to better generalize the cancer images. This decoder portion contains three fully connected layers with neurons 512, 1024 and 4096 respectively.

In the last fully connected layer, the number of neurons is the same to the pixel count in the input image, since the objective is to reduce the sum of square differences among the input images and the reconstructed images.

The reconstruction mask is applied by means of the one-hot function. During training we mask out all except the activity vector of the correct cancer capsule. The value for the target class will be one and zero for the other class.

The decoder is composed of ReLU, a non-linear activation layer followed by a sigmoid activation layer.

3.5.3. Losses

1) Margin Loss: The length of the instantiation output vector represents the likelihood of the entity of the respective capsule being present or not.

The loss incurred for each cancer capsule is k as follows:

$$L_k = T_k \max(0, m^+ - \| v_k \|)^2 + (1 - T_k) \max(0, \| v_k \| - m^-)^2 \tag{6}$$

where $T_k = 1$, if and only when there an image of class k and m+=0.9 and m−=0.1.

λ = 0.5, used for numerical stability.

This loss ensures that when the class is detected the output vectors of the digit capsules are at least m+ long and at most m− long when that class is not detected. λ = 0.5 is used to prevent the loss in the initial learning state from shrinking all vectors.

Reconstruction Loss: It is calculated as the square difference (squared Euclidean distance) of the input image and reconstructed image.

Reconstruction Loss= (Input image) ^{2}- (Reconstructed Image) ^{2}

2) Total Loss: The total loss is merely the sum of all cancer capsules losses. The reconstruction loss is scaled down by a factor λ to prevent it from dominating the margin loss (Sabour et al., 2017).

$$\text{Total Loss} = \text{Margin Loss} + \lambda(\text{Reconstruction Loss})$$

where = 0.0005.

3.6. Performance analysis

The quantitative assessment of the model's validation process is evaluated across the five statistical measures viz. precision, sensitivity or recall, specificity, accuracy and F-score. A classifier's sensitivity, recall or true positive rate (TPR) reflects the positively accurate classified images to the total number of positive images. But specificity or true negative rate (TNR) is expressed as the ratio of correctly identified negative images to the total number of negative images. Positive predictive value (PPV) or precision reflects the portion of positive images correctly categorized to the total number of positive predicted images. F-measure is also known as F-score, which represents the harmonic mean of the precision and recall. High F-measure suggests good efficiency for classification.

Recall or Sensitivity = TPR=$\frac{TP}{TP+FN}$

Specificity = TNR =$\frac{TN}{TN+FP}$

Precision = PPV =$\frac{TP}{TP+FP}$

F-measure =$\frac{2\times Precision\times Recall}{precision+Recall}$

Accuracy (Acc) is the most widely used performance metrics, and is defined as the ratio of the images correctly classified to the total number of images.

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

The Error rate (ERR) or misclassification rate is the complement to the accuracy metric. This metric is the number of images misclassified by positive and negative classes.

$$\text{ERR} = 1 - \text{Acc} = \frac{FP + FN}{TP + TN + FP + FN}$$

4. Result analysis

The proposed model followed the standard protocol (Spanhol et al., 2016) which splits the dataset into training (70%) and testing (30%). The adopted protocol is applied to the dataset of oral cancer with 10-fold cross validation. It is difficult to derive the meaningful information from the networks various layers and thus it cannot be clearly understood how the network is discriminating between the class categories. However the first convolutional layer tried to identify various parts of the histopathological image.

From Fig. 4, the visualization output of the first convolutional layer shows that it identifies nuclei, cytoplasm and other units. Moving further into the network, the interoperability gets lost because of the network's complexity and the computations performed by it.

The result is discussed in terms of the following three aspects

a) Cross-validation accuracy using confusion matrix
b) Performance measure graph
c) Comparison with other state-of-the-art models

a) Cross-validation Accuracy using Confusion matrix

The result of cross-validation as a confusion matrix is discussed in Table 2. The system achieved accuracy of 97.35%. The confusion matrix shows a summary of prediction results on this classification problem. The number of correct benign and malignant predictions are 978 and 969 (diagonal) respectively. The incorrect predictions are the misclassification values such as images predicted as malignant when they are benign is 22 and images predicted as benign when actually has malignancy is 32. It gives the information about the errors being made by the classifier and the type of error. In oral diagnostics tests that produce dichotomized results (positive or negative), accuracy is assessed according to sensitivity and specificity. So these values then assess the potential of a diagnostic test in comparison to the gold standard. Table 3 displays the different performance measures of the model in %.

To further visualize the performance of the proposed binary classifier, it is represented by a receiver operating characteristic (ROC) curve (Fawcett, 2006). An ROC curve is obtained by plotting false positive rate along the X-axis and true positive rate along the Y-axis. Fig. 5. represents the ROC curve of our proposed model. The reported sensitivity is 97.78% and specificity is 96.92%. Both sensitivity and specificity values are high which is truly appropriate for a screening test. This outstanding performance is largely due to the relatively unique feature extraction nature of capsnet model which reflects the efficiency of early detection approach.

Both sensitivity and specificity values are high which shows the proposed model's ability to better capture the spatial information from the benchmark data and better discriminate between the cancerous and non-cancerous images.

b) Performance measure graph

For Capsule network, we have calculated the total loss, which is of two parts: Decoder loss and Capsnet loss. The CapsNet loss measures the misclassification error and is calculated using the Eq. (6). The decoder loss is determined by using mean square error between the input and the image being reconstructed. Our proposed model is trained with 100 epochs with 10-cross fold validation and took 6 h for training in a dedicated software package bundled in Tensorflow on "Ubuntu 16.04 and accelerated by a graphic processing unit (NVIDIA GeForce GTX 1080 Ti with 4X 32 GB of memory)".

The model accuracies and the loss curves are shown in Fig. 6. These learning curves can be used to visualize the incremental evaluation of a classifiers learning performance with epochs. This CapsNet architecture is compared with the CNN model (Panigrahi and Swarnkar, 2019) over the same dataset[1] for the classification of oral cancer. The Capsnet is found to be outperforming the CNN for the same dataset[1], which signifies the more learning capability by the Capsnet classifier. Thus, this learning curve helps in evaluating and selecting the suitable classifier. As CapsNet can handle spatial data, it provides better classification result (accuracy 97.35%) compared to CNN (Panigrahi and Swarnkar, 2019) with accuracy 96.77%. The total loss of capsnet is 0.08312 evaluated on test dataset whereas the validation loss of CNN is 0.13211. From the observation, it is shown that the capsnet provides higher accuracy and lower loss value for the same dataset, training time and number of epochs. This states that the proposed model is more appropriate for the classification task.

c) Comparison with other cutting-edge models

The comparative results for the cutting-edge approaches are shown in Table 4. Our proposed model deals with capsule network

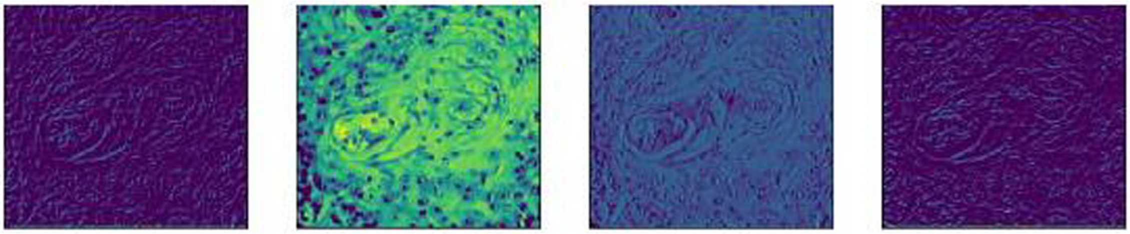

Fig. 4. Feature visualization of the first convolutional layer of the capsnet.

Table 2
The Cross-validation accuracy using Confusion Matrix (mean values).

True vs Predicted	Benign	Malignant
Benign	978	31
Malignant	22	969

Table 3
Different performance measure of the proposed model.

Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F-measure (%)
Capsnet model	97.78	96.92	97.35	96.9	97.33

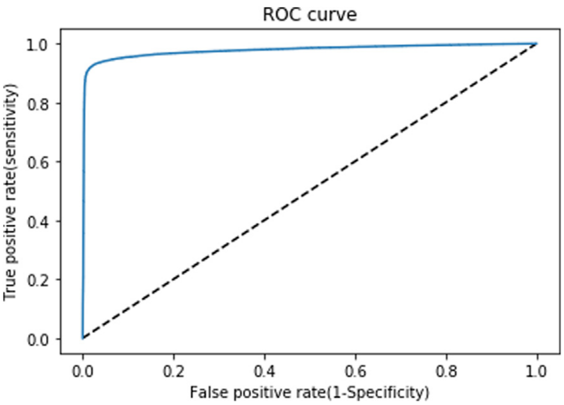

Fig. 5. ROC curve.

Table 4
Comparison of various methods of classification with proposed system

Literature	Classification methods	Accuracy in %
Krishnan et al. (2011)	SVM	88.38
Krishnan et al. (2012b)	Fuzzy	95.7
Krishnan et al. (2012a)	SVM	99.66
Belvin et al. (2013)	Back propagation based ANN	97.92
Anuradha and Sankaranarayanan (2013)	SVM	92.5
Dev Kumar et al. (2018)	Random Forest	96.88
Panigrahi and Swarnkar (2019)	CNN	96.77
Proposed method	**Capsule network**	**97.35**

based classifier for categorizing benign and malignant OSCC. The dataset considered for the study is balanced in nature, i.e. we have considered equal number of samples for both malignant and benign category. The performance of the proposed method is compared with the existing works in terms of accuracy measures. The proposed model and the CNN are deep learning architectures whereas the rest of the literature uses machine learning methods. Small amount of dataset (in the range of 10 to 100) was used in conventional machine learning algorithm which requires an expert dependant hand crafted feature extraction technique. In deep learning approach, it extracts the coarse features on its own, and the efficiency will increase with increase in the data volume. The SVM method by Krishnan et al. (2012a) and back propagation based ANN classification by Belvin et al. (2013) have exceeded the performance of our result by 2.31% and 0.57%, as the features being used for analysis are well defined hand crafted features and the dataset being considered for the purpose is very less in comparison to the proposed model. While the use of existing literature is testified on different datasets, the purpose of the study is same and all of them worked on the histopathological images of oral cancer. In this respect, we should infer, after analysing Table 4,

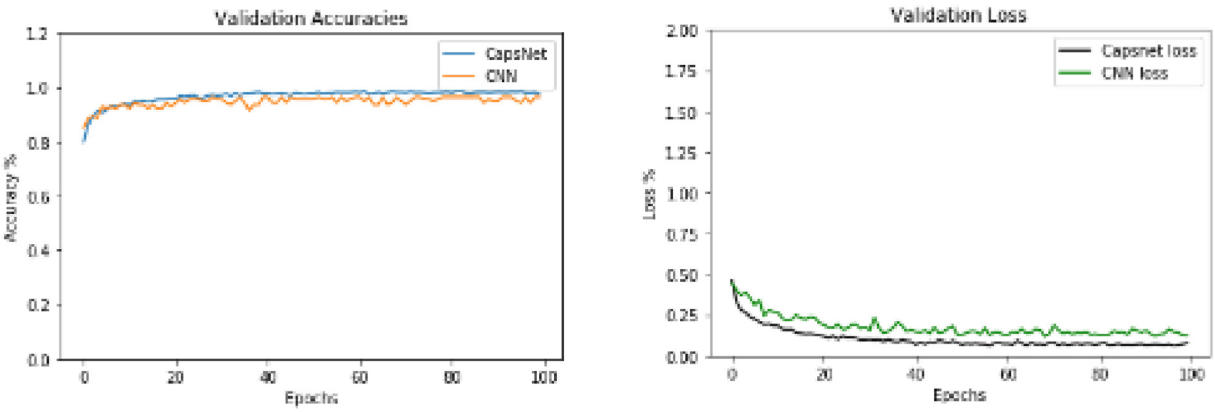

Fig. 6. Accuracy and loss curves.

that our technique is strongly comparable with recent work done as per our best knowledge in the field of OSCC, indicating the feasibility of the solution suggested. This shows that, the proposed technique for the detection of cancerous and non-cancerous tissue would also be an added advantage in the field of computer-aided diagnosis of oral squamous cell carcinoma using histopathological images due to its high throughput.

5. Conclusion

Precise screening of oral histopathological images is essential for diagnosis and treatment planning as well as providing the doctors with a reliable second opinion on the existence of oral lesion. In the current work we have proposed the use of Capsule networks for the oral cancer classification, which is an automated computer-aided method. The cross-validation accuracy was 97.35%, sensitivity and specificity were 97.78% and 96.92% respectively. Since the results of the cross-validation are very encouraging, which suggest that the capsule networks have better capabilities in capturing the pose information and spatial relationship and can better discriminate between the cancerous and non-cancerous images compared to the CNN model. Therefore this model can be used as a diagnostic tool to assist the doctors in their routine clinical screening.

The proposed system can be extended to classify the different stages of oral cancer in future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Afshar, P., Mohammadi, A., Plataniotis, K.N., 2018. Brain tumor type classification via capsule networks. In: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, pp. 3129–3133.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1106–1114.

Anuradha, K., Sankaranarayanan, K., 2013. Comparison of feature extraction techniques to classify oral cancers using image processing. Int. J. Appl. Innov. Eng. Manage. (IJAIEM) 2 (6), 456–462.

Bae, Jaesung, Kim, Dae-Shik, 2018. End-to-end speech command recognition with capsule network. In: Interspeech, pp. 776–780.

Brian, K., 2017. Biopsy. Retrived from https://www.healthline.com/health/biopsy.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern recognition, pp. 1–9.

Dev Kumar, D., Surajit, B., Ashok Kumar, M., Bhaskar, M., Gopeswar, M., Pranab Kumar, D., 2018, Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. Tissue Cell 53, 111–119.

Fawcett, Tom, 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27 (8), 861–874.

Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L., 2016. Breast cancer histopathological image classification using Convolutional Neural Networks. In: In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 2560–2567.

Feurer, Matthias, Hutter, Frank, 2019. Hyperparameter optimization. In: Automated Machine Learning. Springer, Cham, pp. 3–33.

Gedraite, Estevão S., Hadad, Murielle, 2011. Investigation on the effect of a Gaussian Blur in image filtering and segmentation. In: Proceedings ELMAR-2011, IEEE, pp. 393–396.

Hinton, Geoffrey E., Krizhevsky, Alex, Wang, Sida D., 2011. Transforming auto-encoders. In: International conference on artificial neural networks, Springer, Berlin, Heidelberg, pp. 44–51.

Iesmantas, Tomas, Alzbutas, Robertas, 2018. Convolutional capsule network for classification of breast cancer histology images. In: International Conference Image Analysis and Recognition. Springer, Cham, pp. 853–860.

Kingma, Diederik P., Ba, Jimmy, 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Krishnan, M. Muthu Rama., Shah, P., Choudhary, A., Chakraborty, C., Paul, R.R. ,Ray, A.K., 2011. Textural characterization of histopathological images for oral submucous fibrosis detection. Tissue Cell 43, 318–330.

Krishnan, M. Muthu Rama., Chakraborty, C., Paul, R.R., Ray, A.K., 2012a. Hybrid segmentation, characterization and classificationof basel cell nuclei from histopathological images of normal oral mucosa and submucous fibrosis. Expert Syst. Appl. 39,1062–1077.

Krishnan, M. Muthu Rama, Venkatraghavan, V., Acharya, U.R., Pal, M., Paul, R.R., Min, L.C., Ray, A.K., Chatterjee, J., Chakraborty, C., 2012b. Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm. Micron 73(1), 352–364.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

Kumar, Malay, Nanavati, Ronak, Modi, Tapan G., Dobariya, Chintan, 2016. Oral cancer: etiology and risk factors: a review. J. Cancer Res. Ther. 12 (2), 458.

Bui, Marilyn M., Asa, Sylvia L., Pantanowitz, Liron, Parwani, Anil, van der Laak, Jeroen, Ung, Christopher, Balis, Ulysses, Isaacs, Mike, Glassy, Eric, Manning, Lisa, 2019. Digital and computational pathology: Bring the future into focus. J. Pathol. Inf. 10.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision, Zurich, pp. 818–833.

Mobiny, A., Van Nguyen, H., 2018. In: Fast capsnet for lung cancer screening. Springer, Cham, pp. 741–749.

Das, Navarun, Hussain, Elima, Mahanta, Lipi B., 2020. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. Neural Networks, in press.

Otsu, Nobuyuki, 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybernet. 9 (1), 62–66.

Patrick, Mensah Kwabena, Adekoya, Adebayo Felix, Mighty, Ayidzoe Abra, Edward, Baagyire Y., 2019. Capsule networks–a survey. J. King Saud Univ.-Comput. Inf. Sci., in press.

Połap, D., Woźniak, M., 2019. In: July. Bacteria shape classification by the use of region covariance and convolutional neural network. IEEE, pp. 1–7.

Qu, Zhong, Zhang, Li, 2010. Research on image segmentation based on the improved Otsu algorithm. In: 2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics, vol. 2, IEEE, pp. 228–231.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: The Proceedings of the 14th International Conference on AI (IJCAI), Morgan Kaufmann, San Mateo, CA, pp. 1137–1145.

Faizan, S., 2017, Retrived from https://www.analyticsvidya.com/blog/2017/03/F Introduction-to-gradient-descent-algorithm-along-its's-variants/

Sabour, S., Frosst, F., Hinton, G.E., 2017. Dynamic routing between capsules. In: Advances in Neural Information Processing Systems, pp. 3859–3869.

Panigrahi, Santisudha, Swarnkar, Tripti, 2019. In: Automated Classification of Oral Cancer Histopathology images using Convolutional Neural Network. IEEE, pp. 1232–1234.

Belvin, T., Vinod, K., Sunil, S., 2013. In: Texture Analysis Based Segmentation and Classification of Oral CancerLesions in Color Images Using ANN. IEEE, pp. 1–5.

Wang, Duolin, Liang, Yanchun, Dong, Xu., 2019. Capsule network for protein post-translational modification site prediction. Bioinformatics 35 (14), 2386–2394.

Wei, Jason W., Tafe, Laura J., Linnik, Yevgeniy A., Vaickus, Louis J., Tomita, Naofumi, Hassanpour, Saeed, 2019. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. Sci. Rep. 9 (1), 1–8.

Woźniak, M., Połap, D., 2020. Soft trees with neural components as image-processing technique for archeological excavations. Pers. Ubiquit. Comput., 1–13

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L. D., 1989. Backprop- agation applied to handwritten zip code recognition. Neural Comput. 1, 541–551.