ORAL CANCER DETECTION FROM HISTOPATHOLOGIC IMAGES USING DEEP NEURAL NETWORKS

Arya Anil, 2248031
Department of Statistics and Data Science
Christ (Deemed to be University), Bangalore

Royce Rose K R, 2248052
Department of Statistics and Data Science
Christ (Deemed to be University), Bangalore

***Abstract* - Oral cancer is a significant public health problem worldwide, with early detection and accurate diagnosis being critical factors for successful treatment. Oral Squamous cell carcinoma (OSCC) is a common form of oral cancer that is being reported globally. Histopathological analysis of tissue samples is considered to be the gold standard in the diagnosis of OSCC, but the process is time-consuming, labor-intensive, and requires specialized expertise. The development of machine learning techniques, including Convolution neural networks (CNNs), has shown a promising future in automating the detection and diagnosis of cancer from histopathology images. In this paper, three popular deep neural networks, SqueezeNet, GoogleNet, and ResNet50, were used to detect OSCC from histopathology images. Each of these models was trained on different learning rates and their corresponding accuracies were noted. It was found that SqueezeNet gave higher validation accuracy.**

1. INTRODUCTION:

Oral cancer is the sixth most prevalent cancer worldwide. In India, oral cancer is ubiquitous and accounts for approximately 30% of all cancer cases. The mortality rate for oral cancer is high, with about 50,000 deaths every year. Compared with other countries, about 70% of cases in India are reported in the advanced stages. According to the Indian Council of Medical Research (ICMR), early-stage oral cancer cases in India have a five-year survival rate of 30-50%, while advanced-stage patients have a survival rate of 5-20%. Late diagnosis is the primary factor contributing to the low-survival rates in India.

Oral squamous cell carcinoma (OSCC) accounts for the common type of oral cancer in India, accounting for about 90% of all cases. It is found to be more common in men than women, with a ratio of 2:1. OSCC affects the thin, flat cells that line the oral cavity, including the tongue, lips, gums, and the lining of the cheeks and throat. Risk factors include tobacco use (smoking and chewing), alcohol consumption, and human papillomavirus (HPV) infection. Early detection and treatment are critical for improving the survival rates of OSCC patients. Treatment for this is most probably a combination of surgery, radiation therapy, and chemotherapy, depending on the stage and location of the cancer. Efforts are needed to improve awareness about OSCC, its risk factors, and the importance of regular screening for early detection in India.

The histopathological analysis is fundamental in detecting oral cancer, including oral squamous cell carcinoma (OSCC). The method involves the collection of tissue samples from the affected site, which are then processed and stained for examination under a microscope. Pathologists examine the tissue samples for abnormalities in the cell structure, such as changes in size, shape, and organization, which are characteristic of cancer. Thus, histopathological analysis of tissue samples is a gold standard for diagnosing OSCC, but the process is time-consuming, labor-intensive, and requires specialized expertise.

In recent years, machine learning techniques, including convolutional neural networks (CNNs), have shown promise in automating the detection and diagnosis of cancer from histopathology images. They are a type of artificial neural network that is particularly well-suited for image recognition and analysis, thus making them well-suited for analyzing medical images.

Our study aims to compare different CNN-based models for detecting OSCC from histopathological images. For this, the models are trained on a dataset of annotated images and are evaluated using standard performance metrics such as accuracy. The result of this study may help us to conclude the model best suited for OSCC detection from histopathological images; this can contribute to improving the efficiency and accuracy of OSCC diagnosis, leading to early detection and improved survival rate.

2. LITERATURE REVIEW:

Madhusmita Das et al. [1], proposed a 10-layer convolution neural network model for the automation and early detection of OSCC—the histopathological oral cancer images were used for this study. The proposed model was then compared and analyzed with existing deep learning models like VGG16, VGG19, Alexnet, ResNet50, ResNet101, Mobile Net, and Inception Net. It was concluded that the proposed 10-layer CNN model could assist doctors in identifying oral cancer and efficient treatment planning.

Panigrahi et al.[5], have proposed classifications of OSCC histopathology images using two approaches. Transfer learning-assisted deep convolution neural networks (DCNNs) were considered to differentiate between benign and malignant cancers. The baseline DCNN architecture was trained with ten convolution layers as proposed. In addition, a comparative study was also performed with other existing models. The results suggest that ResNet50 performs well when compared with the other DCNN models.

Fabio Alexandre Spanhol et al.[3] , proposed a method for classifying breast cancer histopathological images using deep learning techniques. The authors use a convolutional neural network (CNN) to automatically learn discriminative features from the images and classify them into four categories: normal tissue, benign lesions, in situ carcinoma, and invasive carcinoma. The proposed approach achieves high accuracy in classifying the images and outperforms traditional machine learning methods such as Decision Tree and Random Forest (RF). The findings implied that the proposed convolutional neural network model has the potential to serve as a valuable tool in aiding pathologists to diagnose breast cancer and mitigate inter-observer variability.

Panigrahi et al.[4], have suggested the use of 4-layered patches of CNN for feature extraction and classification from oral cancer dataset. The architecture that they have used comprises of four convolution layers, four pooling layers and two fully connected layers. Each Convolution layer uses ReLU to lessen gradient vanishing issue. They have also used dropout techniques to avoid overfitting. Using this architecture, they managed to achieve an accuracy of 96.77%.

Atta-ur Rahman et al.[2] proposed a transfer learning model using AlexNet in the convolutional neural network to extract rank features from oral squamous cell carcinoma biopsy images. The performance parameters like Classification Accuracy, Classification miss rate, etc. were used to analyze the proposed model. With these statistical parameters' help, the proposed model's reliability was understood.

3. METHODOLOGY:

The dataset that we have used here is the Oral Squamous Cell Carcinoma (OSCC) biopsy data retrieved from Kaggle, which has open access to everyone. Within the dataset, there are 5192 histopathological images, with 2466 depicting normal epithelium cells and 2726 depicting oral squamous cell carcinoma (OSCC). The dataset contains two labeled classes: Normal and OSCC. For the study, we have taken 200 histopathological images, with 100 images in each of these two classes.

In histopathology images, normal epithelial cells typically exhibit a uniform structure with well-defined nuclei, organized cell membranes, and evenly distributed chromatin. They are typically small in size and exhibit a regular shape and arrangement. In contrast, oral squamous cell carcinoma (OSCC) cells may display a range of morphological features that are indicative of malignancy such as Irregular shape, Nuclear atypia, Cellular disorganisation, Increased mitotic activity. With the help of machine learning, these characteristic features can be identified and used to differentiate between OSCC cells and normal epithelial cells which in turn will help us in oral cancer detection.

In recent years, researchers have been trying to automate the detection and diagnosis of cancer from histopathological images, including OSCC, by training neural networks. For the study, we have considered three CNN architectures: GoogleNet, SqueezeNet and ResNet-50.

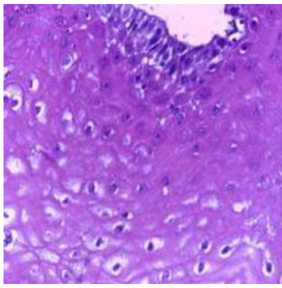


fig1.a

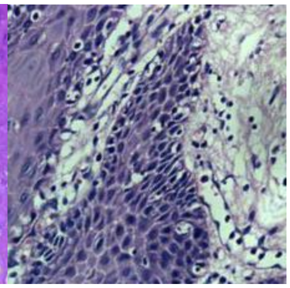


fig1.b

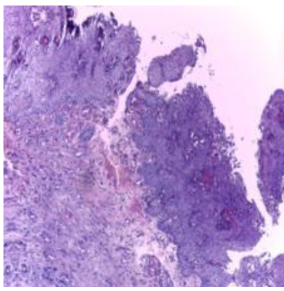


fig1.c

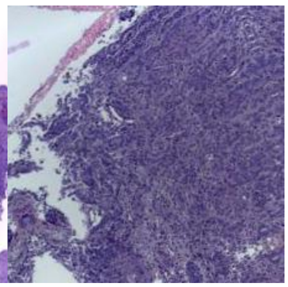


fig1.d

fig 1.a, 1.b - normal epithelial cells.

fig 1.c, 1.d - oral squamous cell carcinoma (OSCC) cells.

3.1 CNN ARCHITECTURE:

CNNs are a type of artificial neural network that can automatically learn relevant features from images, making them well-suited for analyzing medical images.

A typical CNN consists of several layers, including convolutional, pooling, and fully connected layers. Convolutional layers apply a set of filters to the input image, extracting relevant features at different spatial scales. Pooling layers downsample the output of convolutional layers and thus reduce the spatial dimensionality of the feature maps. Fully connected layers combine the learned features from the previous layers and output a classification decision.

The input to the CNN is typically a grayscale or RGB image with fixed dimensions. The output is a probability distribution over the possible classes or labels. The weights and biases are learned during training using a loss function. The loss function will measure the difference between the predicted output and the actual labels.

CNNs can have different architectures depending on the specific task and dataset. More complex architectures, such as ResNet and VGG, involve deeper and wider networks with skip connections and residual blocks that improve the flow of information and reduce the vanishing gradient problem.

3.1.1 GoogleNet

GoogleNet is a specific architecture of CNN that uses inception modules to improve feature extraction. The inception modules are designed to capture information at multiple scales by performing convolutions at different filter sizes and pooling them together. This allows GoogleNet to learn more complex features than a traditional CNN, while still maintaining a relatively small number of parameters. Also known as Inception v1, GoogleNet consists of 22 layers, including 9 Inception modules, max-pooling layers, and fully connected layers. GoogleNet also uses auxiliary classifiers, which are additional softmax classifiers inserted in the network at intermediate layers.

These classifiers encourage the network to learn more discriminative features and help to mitigate the vanishing gradient problem during training.

3.1.2 ResNet-50

ResNet, or Residual Network's deep structure and use of residual blocks allow it to effectively learn complex features and avoid the vanishing gradient problem that can occur with very deep networks.

ResNet-50 is a variant of the ResNet architecture that consists of 50 layers, including convolutional layers, batch normalization layers, ReLU activation layers, pooling layers, and fully connected layers. The architecture includes several residual blocks, with each residual block consisting of two or more convolutional layers, followed by a batch normalization layer and a ReLU activation layer. The skip connection in each residual block adds the original input to the output of the convolutional layers.

ResNet-50 also includes down-sampling layers, which reduce the spatial dimensions of the feature maps, and fully connected layers, which produce the final output. The architecture uses bottleneck blocks in which the first 1x1 convolutional layer is used to reduce the dimensionality of the input before passing it through the next 3x3 convolutional layer. This helps reduce the computational cost of the network while still allowing for the learning of complex features.

3.1.3 SqueezeNet

SqueezeNet is a deep neural network architecture that was designed for efficient processing of images on devices with limited computing resources. The main idea behind SqueezeNet is to achieve high accuracy on image classification tasks while reducing the number of parameters and computation required compared to other architectures like VGG and ResNet.

SqueezeNet is composed of three types of layers: Fire modules, pooling layers, and convolutional layers. The Fire module is the core building block of SqueezeNet, which consists of a squeeze layer followed by an expand layer. The squeeze layer uses 1x1 convolutions to reduce the number of input channels, while the expand layer uses both 1x1 and 3x3 convolutions to increase the number of output channels.

In addition to the Fire modules, SqueezeNet also uses pooling layers to reduce the spatial dimensions of the feature maps, it also uses convolutional layers to learn features from the input image. The architecture also includes skip connections that enable gradient flow across multiple layers, which helps to address the vanishing gradient problem.

4. RESULTS & ANALYSIS:

For classification, we have used the Deep Network Designer toolbox from Matlab. Initially, we divided the dataset into training and testing sets using an 80:20 split ratio. For each of the models, the training was done on three learning rates, and the corresponding accuracies were noted.

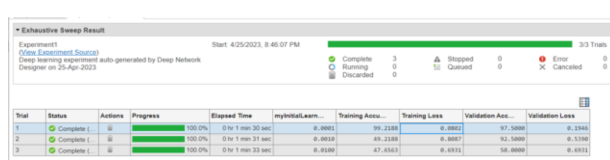
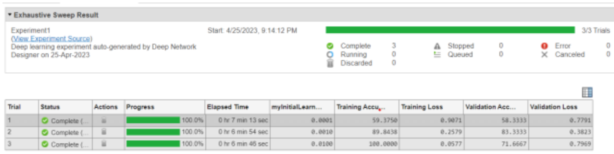
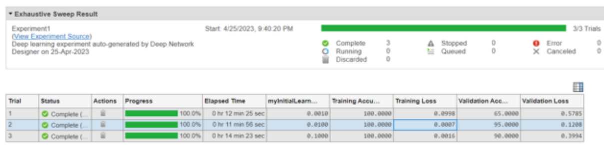


fig2.a SqueezeNet trials with different learning parameters.



Trial	Status	Actions	Progress	Elapsed Time	myInitialLearn...	Training Accu...	Training Loss	Validation Acc...	Validation Loss
1	Complete (...		100.0%	0 hr 7 min 13 sec	0.0001	59.3750	0.9071	58.3333	0.7791
2	Complete (...		100.0%	0 hr 6 min 54 sec	0.0010	89.8438	0.2579	83.3333	0.3823
3	Complete (...		100.0%	0 hr 6 min 46 sec	0.0100	100.0000	0.0577	71.6667	0.7969

fig2.b GoogleNet trials with different learning parameters.



Trial	Status	Actions	Progress	Elapsed Time	myInitialLearn...	Training Accu...	Training Loss	Validation Acc...	Validation Loss
1	Complete (...		100.0%	0 hr 12 min 25 sec	0.0010	100.0000	0.0998	65.0000	0.5785
2	Complete (...		100.0%	0 hr 11 min 56 sec	0.0100	100.0000	0.0007	95.0000	0.1208
3	Complete (...		100.0%	0 hr 14 min 23 sec	0.1000	100.0000	0.0016	90.0000	0.3994

fig2.c .ResNet trials with different learning parameters.

From fig 1.a it is observed that SqueezeNet gave higher accuracies for a learning rate of 0.0001. Similarly from fig 1.b it is clear that GoogleNet gave higher accuracy for a learning rate of 0.0010. For ResNet, higher accuracy was achieved for a learning rate of 0.0100. Hence, we can conclude that depending on the architecture, the learning rate has to be tuned for better model performance.

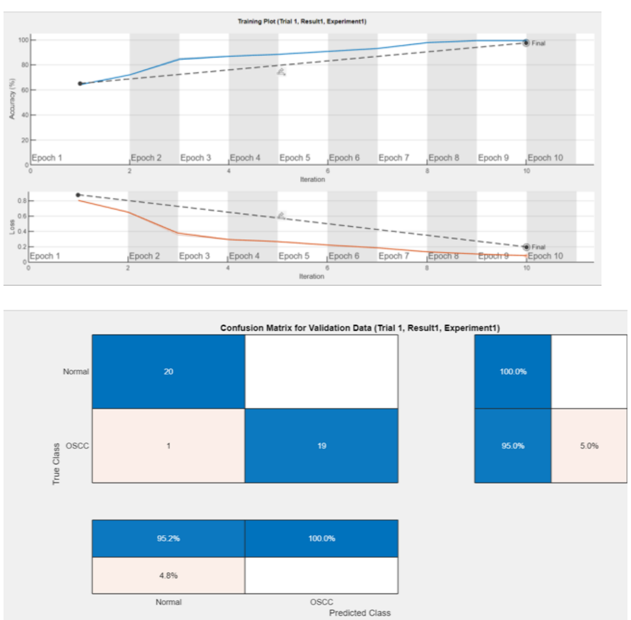


fig3.a Training plot and Confusion Matrix for SqueezeNet.

fig3.b Training plot and Confusion Matrix for GoogleNet.

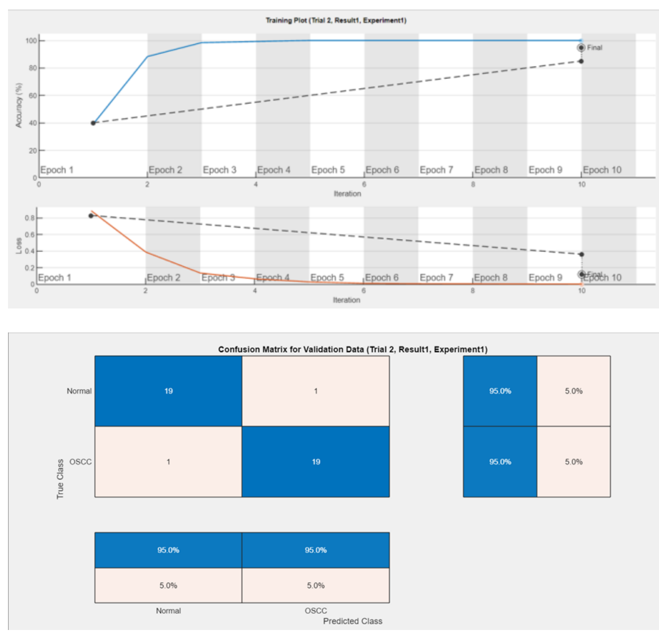


fig3.c Training plot and Confusion Matrix for ResNet.

CNN Architecture	Learning Rate	Training Accuracy	Validation Accuracy
SqueezeNet	0.0001	99.2188	97.5
GoogleNet	0.0010	89.84	83.33
ResNet	0.0100	100	95

Table1. The Performance of Various CNN architectures for Histopathology OSCC Dataset

5. CONCLUSION:

In conclusion, our study aimed to compare existing deep learning-based systems for the detection of oral cancer from histopathologic images. We implemented ResNet, GoogleNet, and SqueezeNet architectures to train and evaluate our models. Our findings showed that deep neural networks effectively detect oral cancer with high accuracy.

SqueezeNet achieved the highest performance among the three architectures, followed by ResNet and GoogleNet. This highlights the importance of selecting the appropriate architecture for the task.

The potential of our approach to improving the accuracy of diagnosis, treatment planning, and patient outcomes is significant. Future research could focus on exploring the use of larger datasets and more advanced deep-learning techniques to enhance the performance of the system further. Modifications in the existing architectures as well as the development of new architectures can contribute to the increase in accuracy.

Overall, our study demonstrates the promise of deep neural networks in detecting oral cancer from histopathologic images and highlights the need for further research in this area.

6. REFERENCES

[1]M. Das, R. Dash, and S. K. Mishra, “Automatic detection of oral squamous cell carcinoma from histopathological images of oral mucosa using deep convolutional neural network,” *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, p. 2131, 2023

[2]A.-ur Rahman, A. Alqahtani, N. Aldhafferi, M. U. Nasir, M. F. Khan, M. A. Khan, and A. Mosavi, “Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning,” *Sensors*, vol. 22, no. 10, p. 3833, 2022.

[3] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using*Conference on Bioinformatics and Biomedicine (BIBM)*, 2019.

[4] S. Panigrahi and T. Swarnkar, “Automated classification of oral cancer histopathology images using convolutional neural network,” *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019.

[5] S. Panigrahi, B. S. Nanda, R. Bhuyan, K. Kumar, S. Ghosh, and T. Swarnkar, “Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning,” *Heliyon*, vol. 9, no. 3, 2023.

[6] S. P. Angayarkanni, “Hybrid convolution neural network in classification of cancer in histopathology images,” *Journal of Digital Imaging*, vol. 35, no. 2, pp. 248–257, 2022.

[7]T. Y. RAHMAN, L. B. MAHANTA, C. CHAKRABORTY, A. K. DAS, and J. D. SARMA, “Textural pattern classification for oral squamous cell carcinoma,” *Journal of Microscopy*, vol. 269, no. 1, pp. 85–93, 2017.