# Mouth and oral disease classification using InceptionResNetV2 method

## Introduction

The study highlights the growing prevalence of oral and dental diseases, including cavities, gum disease, and oral cancer, and their systemic health impacts, such as increased risks of heart disease. It discusses the challenges in diagnosing these conditions and the need for advanced, AI-driven solutions. Machine learning, particularly deep learning using CNNs, has shown promise in medical image processing for detecting dental issues. However, previous studies lacked comprehensive datasets and did not address all oral diseases simultaneously. To overcome these limitations, the researchers propose using InceptionResNetV2 for classifying seven oral diseases. Their new Mouth and Oral Disease (MOD) dataset improves classification accuracy, surpassing existing models.

## Dataset: Mouth and Oral Disease (MOD) dataset

Dataset Source & size: The MOD dataset includes images collected from dental clinics in Okara, Punjab, Pakistan, and other sources like dental websites. The dataset consists of 517 samples with class labels for 7 different mouth and oral cavity diseases.

Disease Categories and Sample Distribution:

- Oral Thrush (OT) – 62 samples
- Canker Sores (CaS) – 78 samples
- Cold Sores (CoS) – 79 samples
- Gingivostomatitis (Gum) – 61 samples
- Oral Lichen Planus (OLP) – 93 samples
- Mouth Cancer (MC) – 90 samples
- Oral Cancer (OC) – 54 samples

Data Labeling: Expert dental practitioners labeled the dataset to ensure accuracy.

Challenges:

- The dataset has a limited number of samples per class, making classification more challenging.
- No demographic information (age, gender, height) was collected before, during, or after the photoshoot.

**Preprocessing**

Resizing**:** Images resized to 224x224 pixels.

Augmentation: To prevent overfitting and increase dataset diversity, Keras image data generator was used with several augmentations. Pixel values were normalized to the range [0,1]. Images were rotated by 25 degrees, shifted horizontally and vertically by 0.1, and sheared with a 0.2 angle. The zoom range parameter randomly altered image sizes, and horizontal flips were used to create vertical variations. Brightness was adjusted between 0.5 and 1.0, and a channel shift of 0.05 was applied to modify pixel values. These transformations improved training set variety and model performance.

Dataset splitting: The MOD dataset was split into three sets: 60% for training, 20% for validation, and 20% for testing. Total samples after augmentation were 5143 images: 3087 for training, 1028 for validation and 1028 for testing with labels for CaS, CoS, Gum, MC, OC, OLP, and OT.

**Model Architecture**

The authors used **InceptionResNetV2**, a hybrid deep learning model combining Inception and ResNet architectures, to extract features at multiple scales while maintaining efficient gradient flow. Transfer learning is employed to enhance performance on the MOD dataset.

InceptionResNetV2 is a deep convolutional neural network (CNN) developed by Google Research, combining the strengths of Inception and ResNet architectures. It features Inception modules, which use parallel convolutional branches with different filter sizes (1x1, 3x3, and 5x5) to capture details at varying scales. The model also includes residual connections to address the vanishing gradient problem and improve training efficiency. The stem block reduces the spatial dimensions of input data, while reduction blocks scale down dimensions and add channels for better high-level information capture. Auxiliary classifiers are used to enhance training, and a global average pooling layer reduces parameters and provides spatial invariance before classification. The final classification is done using a softmax activation.

## Training

The models were trained and tested using a Google Colab Pro account with powerful GPUs. A transfer deep learning model was employed, utilizing the Categorical Cross-Entropy loss function and the Adam optimizer with a learning rate of 0.0001. Early stopping was implemented based on the lowest validation loss, with the best validation accuracy used for saving the model. A batch size of 4 and 50 epochs were selected for all experiments. The model's performance was evaluated on the Mouth and Oral Disease Classification using the MOD dataset, with and without data augmentation, and a comparison was made with existing models to assess its efficacy.

## Results

The proposed model achieved 99.51% accuracy, outperforming previous methods.

The confusion matrix and ROC curves showed near-perfect classification for most disease categories.

Ablation Study: Without data augmentation, accuracy dropped to 74.07%, highlighting its importance.