Mouth and oral disease classification using InceptionResNetV2 method

Javed Rashid[1,2,3] · Bilal Shabbir Qaisar[3,4] · Muhammad Faheem[5] · Arslan Akram[3,6] · Riaz ul Amin[3,4] · Muhammad Hamid[7]

Abstract

Digital tools have greatly improved the detection and diagnosis of oral and dental disorders like cancer and gum disease. Lip or oral cavity cancer is more likely to develop in those with potentially malignant oral disorders. A potentially malignant disorder (PMD) and debilitating condition of the oral mucosa, oral submucous fibrosis (OSMF), can have devastating effects on one's quality of life. Incorporating deep learning into diagnosing conditions affecting the mouth and oral cavity is challenging. Mouth and Oral Diseases Classification using InceptionResNetV2 Method was established in the current study to identify diseases such as gangivostomatitis (Gum), canker sores (CaS), cold sores (CoS), oral lichen planus (OLP), oral thrush (OT), mouth cancer (MC), and oral cancer (OC). The new collection, termed "Mouth and Oral Diseases" (MOD), comprises seven distinct categories of data. Compared to state-of-the-art approaches, the proposed InceptionResNetV2 model's 99.51% accuracy is significantly higher.

✉ Muhammad Faheem
muhammad.faheem@uwasa.fi

1   Department of IT Services, University of Okara, Okara 56310, Pakistan

2   Department of CS&SE, Islamic International University, Islamabad 44000, Pakistan

3   MLC Lab, Meharban House, House # 209, Zafar Colony, Okara 56300, Pakistan

4   Department of CS, University of Okara, Okara 56310, Pakistan

5   School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland

6   Department of Computer Science, Superior University, Lahore 54000, Pakistan

7   Department of Statistics and Computer Science, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan

1 Introduction

When we grin, laugh, or pout, let's put our jaws and teeth to good use. Sound creation, food, drink consumption, and the initiation of the digestive process include the mouth and its accompanying teeth [1]. Oral and dental disorders, such as oral and pharyngeal cancer, gum disease, tooth decay, and dental fluorosis, are on the rise [2]. The bacteria that cause a mouth infection might travel to other body parts. As a result, the body secretes a protein that raises blood pressure. The increased risk of blood clot formation and decreased oxygen and nutrition delivery to the heart increases the likelihood of a heart attack [3]. Heart disease is twice as likely in people with gum and teeth disorders [4]. Globally, oral and dental problems affect more people than other health issues. Damage to nutrient absorption can be caused by the loss of alveolar bone, teeth, edentulism, or masticatory dysfunction [5]. Chronic pain, disfiguring facial abnormalities, and even death can result from untreated oral diseases, making them a major threat to public health. Most systemic diseases (90%) first appear in the patient's speech [4]. Researchers found that 91% of adults aged 20–64 had at least one cavity, with 27% having damage severe enough to be permanent [6]. The WHO predicts that by 2022, about 3.5 billion individuals will have some form of the oral disorder. The ACS (American Cancer Society) forecasted that there will be 54,010 new cases of oral cancer in the US in 2021 [4]. Gum disease will affect over 80% of Americans by 2022 [7]. According to the ADAH (American Dental Hygienists' Association), over 80% of US residents will have experienced a cavity by age 17. Twenty-five percent of adult Americans are missing at least one tooth. Ninety percent of Americans know that preventative dental care is essential [4]. Approximately once every fifteen seconds, an American seeks medical attention for an issue related to their mouth or teeth. As of 2019, the dental implant market is anticipated to be worth US$5.1 billion. However, this number will rise to US$7.97 billion by 2024 [4].

Gingivitis and severe periodontitis can be diagnosed through the use of periodontal probing depth (PPD) measurements, the presence of bleeding during probing, and radiographic evaluation of alveolar bone loss [8]. Because of the wide variety of probes, patients may experience discomfort or even agony during these operations [9, 10]. It requires considerable time and energy to carry out such examinations. Although 72% of patients looking for a new dentist rely on online reviews [4], many dentists still have the opportunity to improve their skillset. Of individuals between the ages of 20 and 64, 91% had at least one cavity, and 27% of adults had decay so severe that they lost a tooth [4]. Pakistan has a severe lack of dental care, with almost half the population never visiting a dentist [11]. To aid in diagnosing oral, dental, and related disorders accurately by medical professionals and patients, it is essential to create a state-of-the-art centralized system based on deep learning. This will lead to a historic low in the number of practicing dentists.

Computational diagnostics receive helpful assistance from the artificial intelligence subfield of machine learning (ML). Machine learning employs historical data to train algorithms and allows them to "infer" (make predictions about) data they have never seen before. Deep learning, a machine learning technology, has found applications in numerous fields [12–15]. The medical and agricultural industries, along with facial expression recognition, intelligent equipment, and language translation are just a few examples. Applying deep learning via Convolutional Neural Networks (CNN) to medical image processing has led to significant advancements in the detection of skin cancer [16], breast cancer [17], and diabetic retinopathy [18]. When used with periapical radiographs, CNNs accurately identify periodontal bone loss, apical lesions, and caries lesions [19].

The early detection and diagnosis of oral, dental, and mouth disorders are where the dental industry can profit from using machine learning techniques. These studies have limited applicability outside of a clinical setting because they rely on data collected from microscopic pictures of plaque [6], radiographic images [20], and fluorescence imaging [21]. Intraoral imaging is useful for the early diagnosis of periodontitis as well. While these studies provide useful information, they do have certain drawbacks, such as a bias towards people in more advanced stages of diseases such periodontitis [22]. Liu et al. [23] developed the MASK R-CNN method after compiling information on dental caries, plaque, osteoporosis, and periodontal disease. Li et al. [24] use a deep learning approach to identify teeth such as incisors, canines, premolars, and molars. To detect white spots and fluorotic and other non-fluorotic lesions with an accuracy of 0.84, Askar et al. [25] compiled a dataset comprising 434 pictures. The sensitivity might be anything from 86 to 97%.

Many obstacles have been identified when using deep learning to diagnose oral and dental problems. Because previous studies had not concentrated on problems affecting the entire mouth simultaneously, deep learning approaches were not employed to diagnose CaS, CoS, Gum, MC, OC, OLP, and OT. CaS, CoS, Gum, MC, OC, OLP, and OT are all common mouth and oral ailments, yet no datasets exist. The third problem was the requirement to enhance the precision of the existing models. Our study's primary goal was to address these deficiencies.

In this study, we propose the deep transfer learning model InceptionResNetV2 for the classification of CaS, CoS, Gum, MC, OC, OLP, and OT affecting the mouth and oral cavity. When compared to previous models, the suggested InceptionResNetV2 achieves better results. CaS, CoS, Gum, MC, OC, OLP, and OT are all part of the self-created Mouth and Oral Disease (MOD) dataset used to train the InceptionResNetV2's transfer learning architecture. The most important results of this study are:

1. Mouth and Oral Diseases Classification using the InceptionResNetV2 method to recognizes the CaS, CoS, Gum, MC, OC, OLP, and OT diseases.
2. The performance of the proposed method is enhanced in terms of accuracy than the existing models.
3. A new dataset of Mouth and Oral Diseases (MOD) such as CaS, CoS, Gum, MC, OC, OLP, and OT has been developed.

2 Related work

Studies have been conducted on dental and oral problems, such as gingivitis, oral cancer, mouth cancer, dental caries, dental fluorosis, periodontitis, tooth damage, dental calculus, plaque, and tooth loss. Researchers recommended adopting MASK R-CNN methods for oral disease diagnosis using their dataset [23]. SqueezeNet (carious, hypo mineralized) was also proposed for detecting anomalies like white spots, fluorescence, and others. The accuracy was 87% [25] on 2781 annotated teeth.

Dental periapical radiographs are indicated for a variety of disorders, including caries, the auditory brainstem response, and infusion-related responses [26]. This article uses panoramic dental X-rays to show how periodontitis can be automatically staged using a deep-learning hybrid approach. Using deep learning, the radiographic bone level (or CEJ level) was identified in 360-degree images of the entire hand. A unique hybrid framework was hypothesized to be able to automatically detect and categorize periodontal bone loss

in each tooth. This situation calls for the use of both traditional CAD processing and deep learning for detection [15].

Dental pathologies include radicular cysts, nasopalatine duct cysts, dentigerous cysts, odontogenic kera tocysts, ameloblastomas, glandular odontogenic cysts, myxomas myxo fibromas, and adenomatoid odontogenic tumours can be identified with the help of a deep learning approach, as described in [31]. Using the periapical radiography dataset, molar and premolar detection (caries, restorative crowns) were suggested [20] using a deep CNN technique implemented in the Keras framework in Python. After training on a dataset, the authors constructed the proposed model to attain 91% accuracy. A faster R-CNN approach was developed to recognize teeth, such as upper right, upper left, lower left, and lower right, with an accuracy of 91% using self-generated dental X-ray pictures [32]. The overall accuracy of this approach was 82.8%.

The intraoral imaging data set was also made available, separating participants into those with and without inflammation. Models for both tooth recognition and inflammation detection were developed using ResNet-50. In a study [27] that compared 305 inflammatory photos to 499 non-inflammatory images, the accuracy of an inflammation recognition model was 77.12%. Dental restorations in panoramic X-rays were classified by Abdalla-Aslan et al. [33] using the Cubic Support Vector Machine (SVM) method. Differentiating between amalgam fillings (AF), composite fillings (CF), crowns (CRW), dental implants (DI), root canal treatments (RCT), and cores (CO) was successful 93.6% of the time, according to the study. The information in Table 1 allows us to draw the following conclusion.

3 Proposed method

Millions of images are utilized for training transfer learning algorithms. Our studies primarily utilized models based on transfer learning. These models are learned through transfer learning. These models' adaptability to new information results from an extensive training on a large corpus. The features gained from a larger dataset can be highly helpful when applied to an issue with a smaller dataset. Training a model from scratch is avoided more easily. The proposed procedure is shown in a flowchart format in Fig. 1.

3.1 Mouth and Oral Disease (MOD) dataset

Access to a high-quality dataset is crucial to developing deep learning models. A new MOD dataset comprises images from dental clinics in Okara, Punjab, Pakistan, and other locations (dental websites, etc.). Table 2 summarizes the 517 samples that comprise the raw Mouth and Oral Disease (MOD) dataset. The total number of samples and a total number of labels for each category are also listed in the table. The collection includes class labels for seven mouth and oral cavity diseases, including canker sores, cold sores, gingivostomatitis, mouth cancer, oral cancer, oral lichen planus, and oral thrush. Before, during, or after the photo shoot, no one revealed their age, gender, or height. Expert dental practitioners contributed to the labeling of the MOD dataset.

Due to the limited number of samples provided for each class label in the MOD dataset, classifying mouth and oral diseases is difficult. There are 62 samples of oral thrush, 78 samples of canker sores, 79 samples of cold sores, 61 samples of gingivostomatitis, 93 samples of oral lichen planus, and 90 samples of mouth cancer in this data collection. This

Table 1 Comparative summary of major state-of-the-art methods on mouth and oral disease classification

Ref	Method	Disease	Database	Limitations	Performance measures
[20]	Deep CNN	Premolars, Molars	Periapical radiographs	Used only 2 classes of teeth	Acc: 82.8%
[25]	SqueezeNet	Fluorotic lesions, White spot lesions, other lesions	Self-created	Used only the lesions of mouth	Acc: 87%
[27]	ResNet50	Inflamed, Non-inflamed	Intraoral image	Used only 2 classes of teeth	Acc: 77.12%
[28]	AlexNet, VGG-16, Detect-Net	Maxillary incisor region	Periapical radiographic	Used only mouth classes	Acc: 90% (AlexNet), 72% (VGG16), 96% (Detect-Net)
[26]	Faster R-CNN	Caries, ABR, IRR	Dental periapical radiographs	Used only 3 classes of teeth	Acc: 84.76%
[29]	GoogLeNet, InceptionV3, CNN network	Dental caries, Non-dental caries	Periapical radiographic images	Used only 2 classes of teeth	Acc: 89%
[30]	ResNet50, ResNet101	Normal teeth, Missing teeth, Residual roots, Dental implants	Dental Periapical radiographic	Used only 4 classes of teeth	Acc: 97.6% (ResNet50), 98% (ResNet101)
[15]	Deep Learning	Junction level (CEJL), Cementoe-namel, Periodontal bone level (PBL)	Panoramic image	Used only 3 classes of teeth	Acc: 91%

Fig. 1 The flowchart of the
proposed method
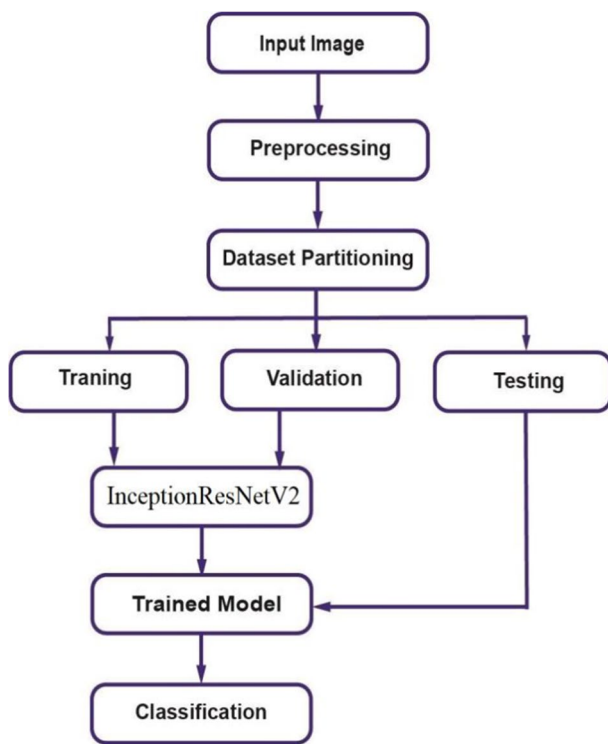
dataset can be used to train and test deep learning models for automatic oral disease diag-
nosis and categorization. Each class of MOD dataset is shown in Fig. 2. MOD dataset can
be accessed from https://drive.google.com/drive/folders/1k24VOveceyqqYS4oaBR0iWLi
DpsDEUk6?usp=drive_link.

3.2 Image resizing

The MOD dataset is expanded in Python to a size of 224 by 224. This will drastically
reduce the processing time at the expense of the model's accuracy.

Table 2 Description of the MOD
dataset (without applying data
augmentation)

Class labels	Images
Canker sores	78
Cold sores	79
Gingivostomatitis	61
Mouth cancer	90
Oral cancer	54
Oral lichen planus	93
Oral thrush	62
Total samples	517

Fig. 2 **a** CaS, **b** CoS, **c** Gum, **d** MC, **e** OC, **f** OLP and **g** OT classes of MOD dataset

3.3 Data augmentation

To prevent over-fitting and expand the training set's variety, we used the picture data generator function from the Keras package in Python. The computer's performance would improve if the variation in pixel values were reduced. By default, the parameter value (1./255) restricts pixel values to the range [0,1]. The photos were flipped to a target orientation of 25 degrees. With a width shift value of 0.1, the width shift range transformation allows for arbitrary right and left image rotation. Training images were vertically moved by 0.1 to either the top or bottom. Using a shear angle of 0.2, we achieved the desired result of preserving one image axis while lengthening the other. This was achieved by randomly increasing or decreasing the image size by using the zoom range parameter. A horizontal image was made vertically by using the flip command. We used a brightness transformation with settings ranging from 0.5 to 1.0 (where 0.0 means no brightness and 1.0 means maximum brightness) to get the results we wanted. To achieve this effect, we applied a channel shift of 0.05 points and choose the fill mode that came closest to the source material. The channel values undergo a transformation by a random number chosen from a predetermined range; hence the term "channel shift transformation." Many methods for bettering data are included in Table 3.

Table 3 Used data augmentation techniques

Transformations	Setting
Scaling	Ranged from 0 to 1
Rotation	25o
Zooming	0.2
Horizontal flipping	True
Shear	20o

3.4 Dataset splitting

The entire MOD dataset was used to generate three distinct data sets: training, validation, and testing. InceptionResNetV2 was first trained on the training dataset, and then its accuracy and efficiency were determined using the validation and test datasets. We used a 60:20:20 split for our data, allocating 60% to training and 20% each to validation and testing. As can be shown in Table 4, 5143 images were trained, validated, and tested using the MOD dataset. CaS, CoS, Gum, MC, OC, OLP, and OT labels were used on 60% of the photos in this study's model introduction. Another 40% of new images were used for testing and validation on the MOD dataset, with each stage receiving 20% of the total. The suggested InceptionResNetV2 method for classifying images of the mouth and oral disorders produced highly accurate predictions for all of the dataset's labels.

3.5 Architecture of InceptionResNetV2 model transfer learning

Deep learning's transfer learning technique has emerged as a potent tool for new jobs with insufficient labeled data. This research employs transfer learning in image classification tasks via the InceptionResNetV2 architecture. The InceptionResNetV2 model is accurate and efficient since it incorporates the best features of the Inception and ResNet architectures. In this study, we will go through the inner workings of InceptionResNetV2 [34] and show how to use it for transfer learning.

Table 4 MOD dataset splitting after applying data augmentation techniques

Split	Classes	Label samples	Total samples
Training	CaS	480	
	CoS	540	
	Gum	360	
	MC	540	3087
	OC	324	
	OLP	540	
	OT	393	
Validation	CaS	160	
	CoS	149	
	Gum	120	
	MC	180	1028
	OC	108	
	OLP	180	
	OT	131	
Testing	CaS	160	
	CoS	149	
	Gum	120	
	MC	180	1028
	OC	108	
	OLP	180	
	OT	131	
	Total		5143

Overview of InceptionResNetV2: The InceptionResNetV2 architecture from Google Research is a deep convolutional neural network (CNN). It builds on the foundations laid by the Inception and ResNet designs, taking their best ideas and making them even more powerful. Inception modules, residual connections, and network depth set apart InceptionResNetV2. Let us take a closer look at the structure's foundational elements:

Inception Modules: InceptionResNetV2 comprises a collection of components known as "Inception modules." The model is built from parallel convolutional branches with varying filter sizes to capture details across various scales. These nodes use max pooling operations and convolutions of 1×1, 3×3, and 5×5 sizes. By combining the results of these forks, the model can accurately capture details at both the regional and global levels. The equation for convolutional layer is as follows:

$$Z^l = W^l * A^{l-1} + b^l$$

$$A^l = ReLU(Z^l)$$

where Z^l is the output of the convolutional layer, W^l is the filter weights, A^{l-1} is the input feature map from the previous layer, and b^l is the bias term.

Residual Connections: ResNet-inspired residual connections are crucial for addressing the vanishing gradient problem and making deep neural network training possible. InceptionResNetV2 adds residual connections between inception modules to improve gradient flow during backpropagation. These associations help keep important data intact and make it easier to train more complex networks.

Stem Block: The stem block is the primary gateway to the network. It comprises a series of convolutional and pooling layers that carry out the task of downsampling and initial feature extraction. By reducing the input's spatial dimensions, the stem block facilitates the model's ability to capture low-level information.

Reduction Blocks: InceptionResNetV2 has reduction blocks built strategically to scale back the spatial dimensions and add more channels. Convolutional layers followed by strides and max pooling processes usually make up these reduction blocks. They lessen the model's computing burden while allowing it to capture higher-level information.

Auxiliary Classifiers: InceptionResNetV2 includes auxiliary classifiers at intermediate nodes to enhance training and regularization. By allowing gradients to be propagated from various depths, these auxiliary classifiers help the model acquire more resilient and informative representations during training.

Global Average Pooling and Classification: A global average pooling layer transforms the feature maps into a fixed-length vector before classification. The final classification probabilities are generated by linking this vector to a fully connected layer using softmax activation. Spatial invariance is provided via global average pooling, and the number of parameters is reduced, improving the model's efficiency. Find the median of the values in each channel: Each channel's average value (c) can be calculated by adding up all the values in that channel and then dividing by the total number of places within that channel ($H \times W$):

$$avg(c) = \sum_{x=1}^{H} \sum_{y=1}^{W} \frac{f(x, y, c)}{H \times W}$$

The value of the feature map at coordinates (x, y) in channel c is denoted as f (x, y, c). An image's global context is represented as a one by one by C tensor after being processed using a global average pooling technique.

Transfer Learning with InceptionResNetV2: In InceptionResNetV2, the pre-trained model's weights are used to fine-tune the model on a new dataset, a process known as transfer learning. To use the learned representations from the pre-training, we can freeze the early layers and train only the last few levels or extra classifier layers to boost performance on the new task. This method shines when only a small amount of labeled data is available.

The InceptionResNetV2 architecture is cutting-edge because it takes the best features of both the Inception and ResNet models. It excels at picture categorization thanks to its inception modules, residual connections, and deep structure. Developers and researchers can save time and resources on image classification projects by taking advantage of transfer learning using InceptionResNetV2. It allows them to make use of pre-trained models to get excellent results. With InceptionResNetV2, researchers have shown that transfer learning may improve performance on various computer vision applications, including object detection, semantic segmentation, and medical image analysis. Using transfer learning with InceptionResNetV2, Wang et al. [35] successfully detected lung nodules in CT scans with an accuracy of 97.5%. Ronneberger et al. [36] also achieved state-of-the-art performance in semantic segmentation of biomedical pictures using transfer learning with InceptionResNetV2. In conclusion, using InceptionResNetV2 for transfer learning is a potent strategy for computer vision tasks, and its success has prompted the creation of other transfer learning systems based on other pre-trained models. Figure 3 presents the bare bones of the InceptionResNetV2 structure.

3.6 Evaluation metrics

This research uses classification accuracy, precision, recall, f1 score, error, and ROC curve as evaluation measures. These measures are used to evaluate the performance of the proposed method.

Fig. 3 Architecture of Inception-ResNetV2 model
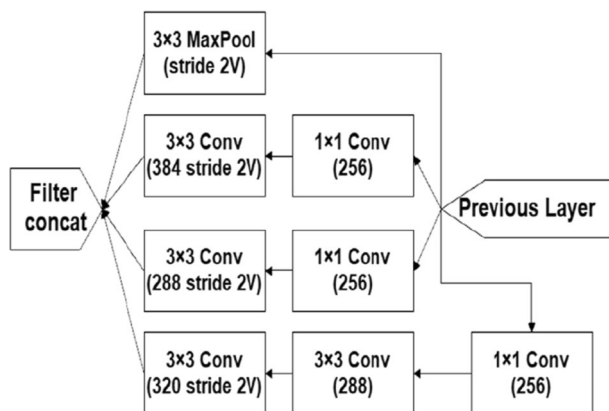
4 Results and discussion

To train and test the suggested models, we used a Google Colab [35] Pro account equipped with powerful Graphical Processing Units (GPUs) requiring no configuration. For this purpose, we employed a transfer deep learning model. The Categorical Cross entropy loss function was employed during model creation, and the Adam optimizer was used in all trials with the proposed technique at a learning rate of 0.0001. The proposed Inception-ResNetV2 used early stopping criteria with the lowest val_loss, and the best val_accuray was used as a saving option; batch size with 4 and 50 epochs was chosen to conduct all the experiments. The proposed method's performance was evaluated on the following:

- Mouth and Oral Disease Classification with a Deep Learning Model (Inception-ResNetV2) was observed using MOD dataset to assess its efficacy.
- The experiments were conducted with and without data augmentation to compare the performances for the ablation study of model.
- The comparison analysis was performed of the proposed model with existing studies.

4.1 The performance analysis of the proposed mouth and oral disease classification using InceptionResNetV2 method

The proposed Mouth and Oral Disease Classification utilizing the InceptionResNetV2 method performance was evaluated. In Fig. 4, we can see the loss and accuracy of the InceptionResNetV2 model throughout training and validation. Over time, the proposed model's training accuracy went from 96.20 to 99.01%. As an alternative, Inception-ResNetV2's model achieved 70.48% validation accuracy after the first epoch and 95.25% accuracy after the last epoch. The loss in training went down from 0.21 to 0.01 after the last epoch, while the loss in validation went down from 1.22 to 0.15.

The proposed InceptionResNetV2 model's accuracy, precision, recall, and F1 score across a variety of classes are detailed in Table 5. Oral cancer (OC), canker sores (CaS),
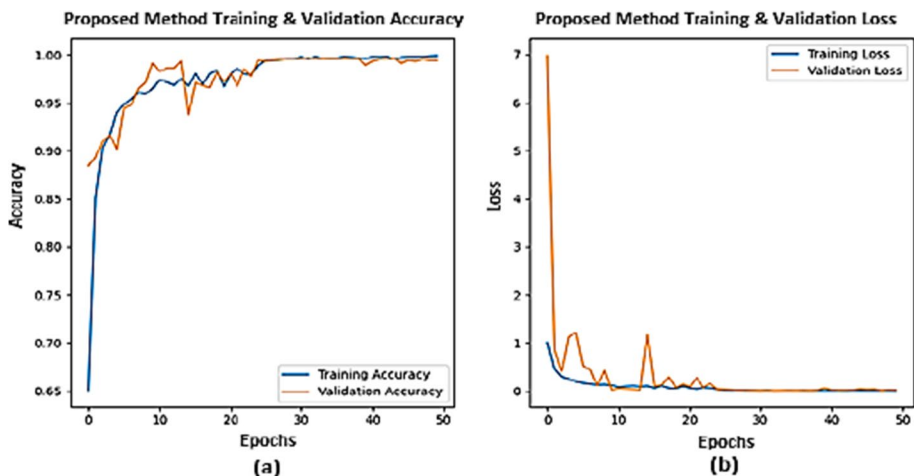

Fig. 4 The proposed InceptionResNetV2 base model: **a** accuracy **b** loss graph

Table 5 Precision, recall, f1 score, and accuracy of the proposed InceptionResNetV2 model

Class	Precision	Recall	F1 Score	Accuracy
OC	100%	100%	100%	100%
CaS	100%	100%	100%	100%
CoS	100%	100%	100%	100%
Gum	99%	99%	99%	100%
MC	99%	99%	99%	100%
OLP	100%	98%	98%	98.33%
OT	98%	100%	99%	100%
Average Accuracy				99.51%

cold sores (CoS), gangivostomatitis (Gum), mouth cancer (MC), oral lichen planus (OLP), and oral thrush (OT) are the categories.

The model attained 100% precision, recall, F1 score, and accuracy for the OC, CaS, CoS, Gum, MC, and OT classes. In other words, the model correctly identified and classified all cases as belonging to their respective classes.

The model achieved a precision of 100% in the OLP class, indicating that it correctly identified instances of Oral Lichen Planus with a high degree of accuracy. The model had a low number of false negatives or cases of OLP that were incorrectly recognized, as indicated by the recall and F1 score of 98%. The OLP class achieved a total accuracy of 98.33%. The suggested InceptionResNetV2 model was generally effective in correctly categorizing oral health issues, with an average accuracy across all classes of 99.51 percent.

These findings demonstrate the InceptionResNetV2 model's superior ability to correctly diagnose various oral health issues, as evidenced by its high accuracy, precision, recall, and F1 score. The results show that the model has promise for facilitating the diagnosis and treatment of oral disorders, which can lead to better patient care and outcomes.

Figure 5 displays the confusion matrix of the proposed InceptionResNetV2 model on the test set, showcasing the results of applying data augmentation techniques. The matrix illustrates the predicted and actual labels for each class in the classification task. The classes represented in the confusion matrix are Oral cancer (OC), canker sores (CaS), cold sores (CoS), gangivostomatitis (Gum), mouth cancer (MC), oral lichen planus (OLP), and oral thrush (OT). The values in the confusion matrix indicate the number of instances that were predicted to belong to a particular class, with the actual class labels represented by the rows and the predicted class labels represented by the columns.

For instance, looking at the first row, it can be observed that 160 instances that belong to the OT class were correctly predicted as OT, on to the second row, where all 149 OLP predictions were spot on.

Similarly, all 120 instances were correctly predicted for the OC class, as indicated by the diagonal element in the third row. The MC class also had accurate predictions for 179 instances, except for a single instance misclassified as GUM.

In the case of the GUM class, 107 instances were correctly predicted, with only one instance misclassified as MC. The CoS class had 177 instances correctly predicted, but three instances were misclassified as CaS.

Finally, all 131 instances of the CaS class were correctly predicted, as shown in the last column.

The InceptionResNetV2 model's ROC curve graph before any data augmentation is applied is shown in Fig. 6. Oral cancer (OC), canker sores (CaS), cold sores (CoS),

Fig. 5 The confusion matrix of the proposed InceptionResNetV2 model
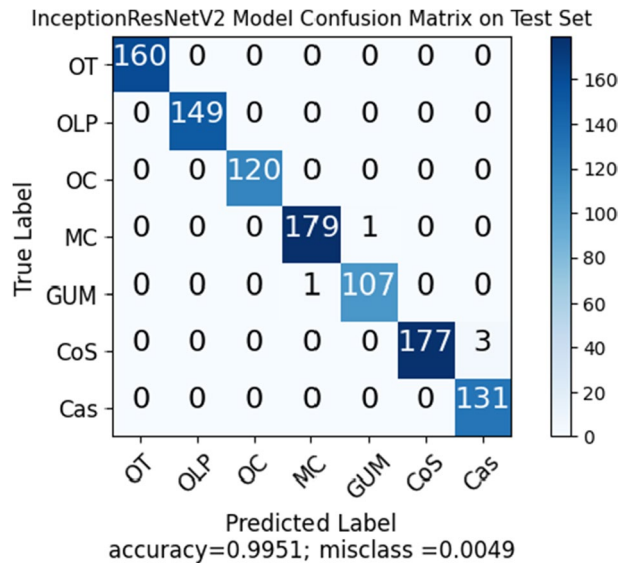
gangivostomatitis (Gum), mouth cancer (MC), oral lichen planus (OLP), and oral thrush (OT) are the classes being used to evaluate the model's ability to distinguish between the classes.

The ROC curve shows how different categorization thresholds affect the True Positive Rate (Sensitivity) and False Positive Rate (1—Specificity). The curve's points stand for different threshold values.

Area Under the Curve (AUC) is supplied as a performance measure, and the graph depicts each class individually. The AUC measures the classification accuracy of a model
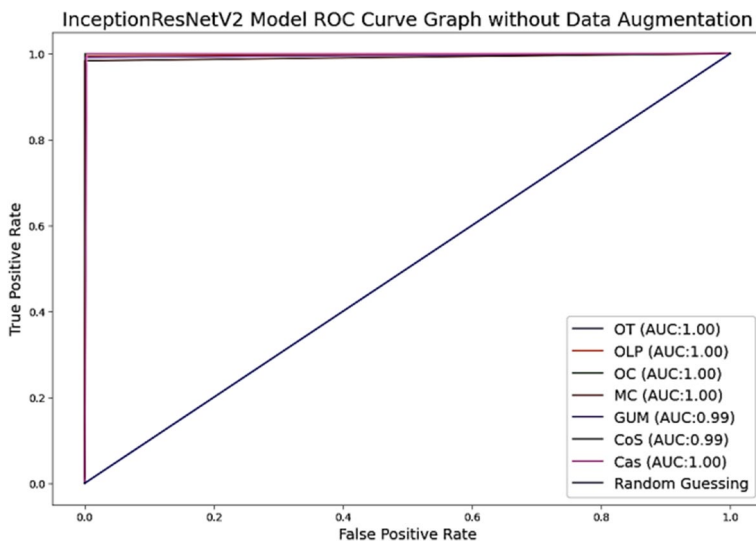

Fig. 6 The proposed InceptionResNetV2 model ROC graph on test set

and can take on values between 0 and 1. Better discriminating power can be inferred from a greater AUC.

In the graph depicting AUC values for each class, it is clear that all classes, except GUM and CoS, attained an AUC of 100%. The model maximized classification accuracy for OT, OLP, OC, MC, and CaS.

For the GUM category, the model only managed a 99% AUC. The model had a somewhat lower discriminatory power than the other classes but could still distinguish between positive and negative occurrences of Gum Disease efficiently.

A similar 99% AUC was also reached by the CoS class, suggesting a strong level of discrimination.

The ROC curve and AUC values show that the InceptionResNetV2 model correctly categorizes most oral health issues without the need for further training data. The strong AUC values indicate the model's discriminatory power, suggesting it could be useful in various diagnosis and classification applications.

All the performance measures indicate that the proposed InceptionResNetV2 model achieved exceptional results when applying data augmentation techniques *to the training set.*

4.2 The comparison of the proposed model using and without using data augmentation techniques

The proposed InceptionResNetV2 model's performance on the test set with and without data augmentation strategies is compared in Table 6. Precision, recall, F1 score, and class-specific accuracy are evaluation metrics used to assess performance.

Table 6 Comparison of the proposed inceptionresnetv2 model's results on the test set when using and not using data augmentation techniques

Performance measures	Precision	Recall	F1 Score	Accuracy
Using Data Augmentation Techniques				
OC	100%	100%	100%	100%
CaS	100%	100%	100%	100%
CoS	100%	100%	100%	100%
Gum	99%	99%	99%	100%
MC	99%	99%	99%	100%
OLP	100%	98%	98%	98.33%
OT	98%	100%	99%	100%
Average Accuracy				99.51%
Without Using Data Augmentation Techniques				
OC	62%	83%	71%	83.33%
CaS	64%	88%	74%	87.5%
CoS	100%	100%	100%	100%
Gum	78%	100%	88%	100%
MC	100%	67%	80%	66.67%
OLP	75%	30%	43%	30%
OT	50%	67%	57%	66.67%
Average Accuracy				74.07%

The model performed exceptionally well across all classes once data augmentation approaches were implemented. For the OC, CaS, and CoS courses, all measures of correctness were 100%, including precision, recall, F1 score, and accuracy. The Gum and MC groups hit 99% accuracy, 99% recall, and 99% in F1 score. The OLP category displayed significantly lower values than the other categories, with 100% precision, 98% recall, 98% F1 score, and 98.33% accuracy. Accuracy, recall, F1 score, and precision all came in at 100% for the OT category. Using data augmentation approaches, the model achieved an average accuracy of 99.51 percent.

In contrast, the model's performance was considerably worse when not supplemented with additional data. Different groups had varying precision, recall, F1, and accuracy ratings. The OC group hit an accuracy of 83.33 percent, a recall of 62 percent, an F1 score of 71 percent, and a precision of 62 percent. CaS also improved its accuracy to 87.5%, recall to 88%, F1 score to 74%, and precision to 64%. CoS consistently scored 100% on all parameters, demonstrating its superior performance. The F1 score for the Gum class was 88%, and they were 100% accurate, with a recall rate of 78% and a precision rate of 100%. The MC group hit an F1 score of 80%, 100% precision, 67% recall, and 100% accuracy. The F1 score for the OLP group was 43%, and its precision and recall were both 30%. When all was said and done, the OT group had an accuracy rate of 66.67%, a recall rate of 67%, an F1 score of 57%, and a precision rate of 50%. Without any more data, the average accuracy was determined to be 74.07%.

The comparison reveals how much of an effect data augmentation approaches have on the model's efficiency. The model maintained high levels of accuracy, precision, recall, and F1 score across all classes when data augmentation was used. The model's performance, however, was highly variable without data augmentation, with large drops in accuracy and other metrics across several classes.

When applied to the problem of oral health condition classification using the InceptionResNetV2 architecture, our findings highlight the significance of data augmentation strategies in improving deep learning models' performance and generalization capabilities. This data suggests that overfitting occurred when the recommended InceptionResNetV2 method was utilized for training and testing with fewer data.

4.3 Comparison with state-of-the-art studies

Table 7 compares the proposed InceptionResNetV2 model with the state-of-the-art mouth and oral disease classification models. A CNN-based approach to periodontal diseases employing a dental calculus database is described by Park et al. [38], which was published in 2023. The approach has been claimed to have a 95% success rate. In 2023, Gomes et al. [39] presented a CNN-based method for categorizing Papule/nodu, Vacule/spot, Vesicle/ bullous, Erosion, Ulcer, and Plaque diseases. The technique uses a custom-built database to get a 95.09% success rate. Singha et al. [40] offer an ensemble method combining ResNet50 and DenseNet201 models to diagnose lung, breast, oral, and other medical disorders. The technique is applied to the problem of oral cancer, and it shows an accuracy of 92%. The classification of dental wear, periapical conditions, periodontitis, tooth decay, missing teeth, and impacted teeth using EfficientNetB0 is shown by Jaiswal et al. [41]. The method utilizes a custom-built database to produce a 93.2% success rate. Classification tasks involving oral health issues or dental radiographs are addressed in a wide range of studies [25, 27, 28, 33, 37, 42, 43] published between 2019 and 2021. These studies employ diverse architectures, including ResNet50, SqueezeNet, and Deep CNN. Accuracy

Table 7 Comparison of the proposed InceptionResNetV2 model with state-of-the-art methods

Ref. Year	Method	Disease	Database	Accuracy
[38] 2023	CNN	Periodontal diseases	Dental calculus	95%
[39] 2023	CNN	Papule/nodu, Vacule/spot, Vesicle/bullous, Erosion, Ulcer and Plaque	Self–created	95.09%
[40] 2022	Ensemble of Resnet50, Densenet201	Lung cancer, Breast cancer, Oral cancer, and various medical diagnosis	Oral cancer dataset	92%
[41] 2022	EfficientNetB0	Tooth wear, Periapical, Periodontitis, Tooth decay, Missing tooth, and Impacted tooth	Self–created	93.2%
[27] 2020	ResNet50	Inflamed, Non-inflamed	Intraoral image	77.12%
[25] 2021	SqueezeNet	Fluorotic lesions, White spot lesions, other lesions	Self–created	87%
[42] 2021	Deep CNN	Decay, Periapical Periodontitis, and Periodontitis	Self–created	96%
[43] 2021	CNN	Caries, Non-caries	Self–created	71.43%
[37] 2019	CNN	Incisors, Canines, Premolars, Molars	Periapical radiographs	96%
[28] 2021	AlexNet, VGG-16, Detect- Net	Maxillary incisor region	Periapical radiographic	90%,72%, 96%
[33] 2020	Cubic SVM	Amalgam filling, Dental implant, Composite filling, Crown, Root canal treatment, Core	Periapical radiographs	93.6%
Proposed Method		CaS, CoS, Gum, MC, OC, OLP, OT	MOD	99.51%

claims for these approaches range from 71.43% to 96%; they are applied to various diseases and use databases and custom-built datasets. The proposed InceptionResNetV2 model is specifically developed for classifying CaS, CoS, Gum, MC, OC, OLP, and OT in the context of MOD. The proposed model was reported to have a test set accuracy of 99.51%.

Table 7 depicts that the proposed InceptionResNetV2 model fares against the state-of-the-art models. It demonstrates the superior performance of the proposed model above many other existing methods in the area, demonstrating its high accuracy in identifying diverse oral health issues. These findings highlight the success of the proposed model and its potential for furthering disease classification in the field of oral health.

5 Conclusion and future work

In conclusion, it is heartening to see that deep learning models successfully detect CaS, CoS, Gum, MC, OC, OLP, and OT. With such a limited sample size for training and testing, it's no surprise that some models perform poorly in instability. The Mouth and Oral Disorders (MOD) dataset was developed to address this issue and is split into seven subcategories. Compared to other models, the 99.51% accuracy of our novel InceptionResNetV2 technique for diagnosing mouth and oral disorders is significantly higher. A wide variety of data augmentation techniques were employed to further improve accuracy. Suppose we want to improve our ability to detect and diagnose oral and dental disorders early on. In that case, future research should widen the scope to cover additional diseases on a wider and more generalizable dataset.

Data availability The data will be available upon reasonable request to the corresponding author.

Declarations

Conflict of interest The authors declare no known competing interests financial interests or personal relationships that could have appeared to influence the work reported in this study.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Larissa Hirsch M (2018) KidsHealth / Parents / Mouth and TeethMouth and Teeth
2. Gezgin O, Botsali M (2018) Evaluation of teeth development in unilateral cleft lip and palate patients in mixed dentition by using medical image control systems. Niger J Clin Pract 21:156–162
3. Catharine (2010) "Medial news today," Dec. 2010. [Online]. Available: https://www.medicalnewstoday.com/articles/200132#1
4. Seirawan S (2021) "unnus," Jun. 2021. [Online]. Available: https://unnus.com/medical/
5. Martins JN, Marques D, Francisco H, Caramês J (2018) Gender influence on the number of roots and root canal system configuration in human permanent teeth of a Portuguese subpopulation. Quintessence Int 49:103–111
6. Aberin STA, de Goma JC (2018) Detecting periodontal disease using convolutional neural networks. 2018 IEEE 10th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management (HNICEM). IEEE, pp 1–6
7. Mcdowell D (2022) "Wayzata dental," 2022. [Online]. Available: https://www.wayzatadental.com/dental-statistics/
8. Preshaw PM (2015) Detection and diagnosis of periodontal conditions amenable to prevention. BMC Oral Health 15:1–11
9. Hefti AF (1997) Periodontal probing. Crit Rev Oral Biol Med 8:336–356
10. Osbom JB, Stoltenberg JL, Huso BA, Aeppli DM, Pihlstrom BL (1992) Comparison of measurement variability in subjects with moderate periodontitis using a conventional and constant force periodontal probe. J Periodontol 63:283–289
11. Khan AA (2015) Dentistry in Pakistan-standing at the crossroads facing numerous challenges with an uncertain future. JPDA 25(1):1–2
12. Rashid J, Khan I, Ali G, Rehman, Shafiq U, Alturise F, Alkhalifah T (2023) Real-time multiple guava leaf disease detection from a single leaf using hybrid deep learning technique. Comput Mater Contin 41:1235–1257
13. Rashid J, Khan I, Ali G, Almotiri SH, AlGhamdi MA, Masood K (2021) Multi-level deep learning model for potato leaf disease recognition. Electronics 10:2064
14. Sohail M, Ali G, Rashid J, Ahmad I, Almotiri SH, AlGhamdi MA, Nagra AA, Masood K (2021) Racial identity-aware facial expression recognition using deep convolutional neural networks. Appl Sci 12:88
15. Chang HJ, Lee SJ, Yong TH, Shin NY, Jang BG, Kim JE, Huh KH, Lee SS, Heo MS, Choi SC et al (2020) Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. Sci Rep 10:1–8
16. Ali G, Dastgir A, Iqbal MW, Anwar M, Faheem M (2023) A hybrid convolutional neural network model for automatic diabetic retinopathy classification from fundus images. IEEE J Transl Eng Health Med 11(6):012116
17. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A (2017) Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Invest Radiol 52:434–440
18. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama 316:2402–2410
19. Schwendicke F, Golla T, Dreher M, Krois J (2019) Convolutional neural networks for dental image diagnostics: a scoping review. J Dent 91:103226
20. Lee JH, Kim Dh, Jeong SN, Choi SH (2018) Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. J Periodontal Implant Sci 48:114–123
21. Rana A, Yauney G, Wong LC, Gupta O, Muftu A, Shah P (2017) Automated segmentation of gingival diseases from oral images. 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT). IEEE, pp 144–147
22. Balaei AT, de Chazal P, Eberhard J, Domnisch H, Spahr A, Ruiz K (2017) Automatic detection of periodontitis using intra-oral images. 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 3906–3909
23. Liu L, Xu J, Huan Y, Zou Z, Yeh SC, Zheng LR (2019) A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal. IEEE J Biomed Health Inform 24:898–906

24. Li Z, Wang SH, Fan RR, Cao G, Zhang YD, Guo T (2019) Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling. Int J Imaging Syst Technol 29:577–583

25. Askar H, Krois J, Rohrer C, Mertens S, Elhennawy K, Ottolenghi L, Mazur M, Paris S, Schwendicke F (2021) Detecting white spot lesions on dental photography using deep learning: a pilot study. J Dent 107:103615

26. Khan HA, Haider MA, Ansari HA, Ishaq H, Kiyani A, Sohail K, Muhammad M, Khurram SA (2021) Automated feature detection in dental periapical radiographs by using deep learning. Oral Surg Oral Med Oral Pathol Oral Radiol 131:711–720

27. Alalharith DM, Alharthi HM, Alghamdi WM, Alsenbel YM, Aslam N, Khan IU, Shahin SY, Dianišková S, Alhareky MS, Barouch KK (2020) A deep learning-based approach for the detection of early signs of gingivitis in orthodontic patients using faster region-based convolutional neural networks. Int J Environ Res Public Health 17:8447

28. Kuwana R, Ariji Y, Fukuda M, Kuwada C, Muramatsu C, Katsumata A, Fujita H, Ariji E (2021) Performance of deep learning object detection technology in the detection and diagnosis of maxillary sinus lesions on panoramic radiographs. Dentomaxillofac Radiol 50:20200171

29. Lee JH, Kim DH, Jeong SN, Choi SH (2018) Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent 77:106–111

30. Mahdi FP, Motoki K, Kobashi S (2020) Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs. Sci Rep 10:19261

31. Watanabe H, Ariji Y, Fukuda M, Kuwada C, Kise Y, Nozawa M, Sugita Y, Ariji E (2021) Deep learning object detection of maxillary cyst-like lesions on panoramic radiographs: preliminary study. Oral Radiol 37:487–493

32. Chen H, Zhang K, Lyu P, Li H, Zhang L, Wu J, Lee CH (2019) A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Sci Rep 9:1–11

33. Abdalla-Aslan R, Yeshua T, Kabla D, Leichter I, Nadler C (2020) An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. Oral Surg Oral Med Oral Pathol Oral Radiol 130:593–602

34. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826

35. Xie H, Yang D, Sun N, Chen Z, Zhang Y (2019) Automated pulmonary nodule detection in CT images using deep convolutional neural networks. Pattern Recogn 85:109–119

36. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp 234–241

37. Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, Schwendicke F (2019) Deep learning for the radiographic detection of apical lesions. J Endod 45:917–922

38. Park S, Erkinov H, Hasan MAM, Nam SH, Kim YR, Shin J, Chang WD (2023) Periodontal disease classification with color teeth images using convolutional neural networks. Electronics 12:1518

39. Gomes RFT, Schmith J, Figueiredo RMd, Freitas SA, Machado GN, Romanini J, Carrard VC (2023) Use of artificial intelligence in the classification of elementary oral lesions from clinical images. Int J Environ Res Public Health 20:3894

40. Singha Deo B, Pal M, Panigrahi PK, Pradhan A (2022) An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification. medRxiv, pp 2022–11

41. Jaiswal P, Katkar V, Bhirud SG (2022) Multi Oral Disease Classification from Panoramic Radiograph using Transfer Learning and XGBoost. J Adv Comput Sci Appl 13(12)

42. Chen H, Li H, Zhao Y, Zhao J, Wang Y (2021) Dental disease detection on periapical radiographs based on deep convolutional neural networks. Int J Comput Assist Radiol Surg 16:649–661

43. Sonavane A, Yadav R, Khamparia A (2021) Dental cavity classification of using convolutional neural network. In: IOP conference series: materials science and engineering, vol 1022, no 1. IOP Publishing, p 012116