MAUN Fen Bil. Dergi., 13, 1, 26-36 Araştırma Makalesi/ Research Article
MAUN J. of Sci., 13, 1, 26-36 DOI : 10.18586/msufbd.1631254

Integrating Pretrained Deep Neural Networks with Traditional Classification Techniques for Enhanced Oral Cancer Diagnosis

Bilal ŞENOL[1]🖋, Uğur DEMİROĞLU[2]
[1] Department of Software Engineering, Faculty of Engineering, Aksaray University, Aksaray, Türkiye
[2] Department of Software Engineering, Faculty of Engineering, Architecture and Design, Kahramanmaraş İstiklal University, Kahramanmaraş, Türkiye

✉: bilal.senol@aksaray.edu.tr [1] 0000-0002-3734-8807 [2] 0000-0002-0000-8411

Received (Geliş): 01.02.2025 Revision (Düzeltme):07.04.2025 Accepted (Kabul): 22.04.2025

ABSTRACT

This research aims to introduce a hybrid method for the classification of oral cancer images. This methodology integrates conventional classification techniques, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees, with sophisticated feature extraction from pretrained deep neural networks, including GoogleNet and MobileNetV2. The suggested strategy collects information from deep learning models to construct a robust hybrid model that enhances diagnostic accuracy. The hybrid model attains a classification accuracy of 90.01% using Quadratic SVM, reflecting a 22.36% enhancement compared to individual deep learning models. Comparative assessments demonstrate the significant performance benefits attained by the hybrid approach. The findings underscore the possibility of integrating modern deep learning techniques with traditional methods to enhance the accuracy and reliability of medical image classification, notably in the diagnostic evaluation of oral cancer.

Keywords: Oral Cancer, Image Classification, Hybrid Model, GoogleNet, MobileNet-v2

Gelişmiş Ağız Kanseri Tanısı İçin Önceden Eğitilmiş Derin Sinir Ağlarının Geleneksel Sınıflandırma Teknikleriyle Entegre Edilmesi

ÖZ

Bu çalışma, ağız kanseri görüntülerinin sınıflandırılması için hibrit bir yöntem önermektedir. Bu yöntem, GoogleNet ve MobileNetV2 gibi önceden eğitilmiş derin sinir ağlarından sofistike özellik çıkarımı ile birlikte Destek Vektör Makineleri (SVM), K-En Yakın Komşu (KNN) ve Karar Ağaçları gibi geleneksel sınıflandırma tekniklerini birleştirir. Önerilen strateji, tanısal doğruluğu artıran sağlam bir hibrit model oluşturmak için derin öğrenme modellerinden bilgi toplar. Hibrit model, Quadratic SVM kullanarak %90.01 sınıflandırma doğruluğu elde eder ve bu, bireysel derin öğrenme modellerine kıyasla %22.36'lık bir iyileşmeyi yansıtır. Karşılaştırmalı değerlendirmeler, hibrit yaklaşımın sağladığı önemli performans avantajlarını göstermektedir. Bulgular, modern derin öğrenme tekniklerinin geleneksel yöntemlerle entegre edilmesinin, özellikle ağız kanserinin tanısal değerlendirilmesinde tıbbi görüntü sınıflandırmasının doğruluğunu ve güvenilirliğini artırma olasılığını vurgulamaktadır.

Anahtar Kelimeler: Ağız Kanseri, Görüntü Sınıflandırma, Hibrit Model, GoogleNet, MobileNet-v2

GİRİŞ

Oral cancer is the atypical proliferation and dissemination of cells in any region of the mouth, including the floor of the mouth, beneath the tongue, the soft palate, the hard palate, the gums, or the lips. Oral cancer impacts individuals of all ages, races, and ethnicities. Approximately fifty-four thousand Americans will receive a diagnosis of oral or oropharyngeal cancer this year, resulting in over nine thousand fatalities. Despite treatment advances, more than half of diagnosed patients die within five years. The concerning increase in early mortality rates is chiefly attributable to the absence of pain or symptoms in the initial stages of oral cancer. Consistent professional oral health assessments are essential and can identify alterations in oral health that may signify the onset of cancer prior to disease advancement. Public awareness of oral cancer is essential and may help reduce mortality. [2]. The prospects for treating oral cancer improve with early detection, and the likelihood of long-term survival is typically favorable. Educational interventions and the augmentation of public awareness and comprehension concerning early symptoms and hazards can influence early detection and final results [3]. Promoting and supporting community-based research aimed at reducing morbidity and death associated with oral cancer is essential. Alongside enabling individuals to adopt preventive measures to mitigate risk, grassroots initiatives can also counter fatalism and challenge the assumption of inevitable illness among laypersons who disregard health-related communications from authorities. This information may enhance public welfare in the long term. The subsequent content sections provide an in-depth analysis of the disease, encompassing the several forms of oral cancer, their prevalence and effects, high-risk populations, and the complex etiology involved in their development [4].

Timely identification is essential, and fostering this awareness is the principal objective of oral cancer screening initiatives. Individuals want to consistently examine the following regions for indicative indicators of oral cancer: lips, neck, floor of the mouth, ventral surface of the tongue, buccal mucosa, palate, and places associated with heightened salivary secretion. Regrettably, symptoms are often erratic and unpredictable, hence patients are highly advised to familiarize themselves with their own bodies [5]. An exhaustive strategy for the early identification and treatment of oral and oropharyngeal cancer is essential for enhancing the survival rates of patients with oral cancer. Timely identification of oral cancer significantly diminishes the physical, mental, and financial burdens of the illness, along with the mortality rates linked to the disease [6, 7]. Recently, computational methods are extensively employed to diagnosis various cancer kinds via classification algorithms.

Image classification techniques have a large application in the healthcare industry. For effective treatment, it is necessary to provide a fast and accurate diagnosis. Most often, medical imaging analysis is opted for diagnosing any disease. With the introduction of big data technologies in healthcare, a huge number of medical images are recorded on a daily basis [8]. Therefore, efficient algorithms are necessary for the analysis and classification of medical images. Medical image analysis includes various methods that analyze and detect the structure within the images, while medical image classification is used for labeling and healthcare informatics. The reason for classifying the medical images is that classification is the first step in the examination process [9]. Therefore, we need to understand what is available in a current dataset. If the current image has a diagnosis of some sort, then it is easier to look for similar datasets for research. Also, having a label on the image makes it easier to look for different ways to serve a diagnosis and can be integrated with health informatics to form an automated diagnosing model [10]. With the evolution of the area of deep learning-based computer vision, there has been significant progress in building automatic classification models for a wide array of applications in engineering, computer science, and the health care field [11-13]. Advancements in healthcare have benefited significantly from the adoption of machine learning and deep learning approaches for image classification. Traditional machine learning analysis of structured and small datasets does not fully harness the complex and high-dimensional information embedded in images. With deep learning, however, features are extracted on a learnable basis rather than predetermined, which results in better performance when dealing with high-dimensional data prevalent in imagery [14]. Especially in the healthcare field, convolutional neural networks have been effectively used to process and classify medical images and vital signals. Several researchers are currently exploring the possibilities of using convolutional neural networks to diagnose oral cancer by examining and classifying oral image data [15].

This paper aims to categorize oral cancer images utilizing pretrained deep neural networks (DNNs) with traditional classification techniques. Consequently, the benefits of DNN algorithms are integrated with the efficacy of classical approaches. The deployed pretrained deep neural networks are GoogleNet and MobileNet-v2. The feature extraction phase is conducted using these DNNs, followed by classification with several traditional classification methods. Their performances have been evaluated as the conclusion. GoogleNet, also known as Inception v1, is a convolutional neural network (CNN) architecture created by Google researchers, which achieved significant acclaim by winning the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The architecture incorporated the Inception module, enabling the network to record features at numerous scales within each layer by concurrently applying convolutional filters of varying sizes. This design innovation markedly enhanced accuracy and computing efficiency by optimizing parameter utilization, yielding fewer parameters than numerous deep networks of its time. The success of GoogleNet facilitated the development of succeeding models within the Inception family, each enhancing the modular design to address more intricate patterns and further augment efficiency, frequently incorporating batch normalization and other enhancements in later iterations [16].

MobileNetV2 is a streamlined and efficient deep neural network architecture created by Google in 2018, primarily designed for mobile and embedded vision applications. MobileNetV2 enhanced the original MobileNet by incorporating significant advancements that increased accuracy and computational efficiency, rendering it suitable for resource-limited settings. A significant advancement in MobileNetV2 is the implementation of inverted residuals and linear bottlenecks. Unlike conventional residual connections that transmit high-dimensional features via shortcut connections, MobileNetV2 employs a lower-dimensional (bottleneck) representation. The inverted residual structure allows MobileNetV2 to preserve extensive feature information with less parameters, hence improving both speed and accuracy. Moreover, it utilizes depthwise separable convolutions, a method that disaggregates conventional convolutions to diminish computational expenses while maintaining model efficacy. MobileNetV2 is extensively utilized in applications necessitating real-time image processing, including object detection, facial recognition, and various on-device AI tasks, owing to its equilibrium of efficiency and performance [17].

GoogleNet and MobileNetV2 were specifically chosen for their complementary characteristics: GoogleNet offers high representational power with optimized computational complexity, while MobileNetV2 is designed for efficient inference, making it ideal for resource-constrained environments such as mobile diagnostics. These models represent two ends of the efficiency-performance trade-off, aligning with our goal to develop a scalable and robust diagnostic framework.

The literature study indicates the existence of numerous valuable studies addressing oral cancer diagnosis by categorization methodologies. For instance, [18] presents an exhaustive analysis of recent progress in dental, oral, and craniofacial imaging, highlighting the imperative of differentiating tumor-associated tissues from imaging data. The authors emphasize the application of several deep convolutional neural network (CNN) models for the automation of histological lesion grading, acknowledging that despite their potential, obstacles including insufficient training datasets impede wider implementation. Similarly, [19] highlight the urgent necessity for efficient early detection of oral cancer, especially in marginalized areas. Their research demonstrates the capability of smartphone-derived photos evaluated by deep learning to enable prompt identification of oral squamous cell carcinoma (OSCC), which represents the predominant form of oral cancer. The work in [20] elaborates on this concept by examining diverse machine learning methodologies that complement noninvasive diagnosis of oral precancer and cancer. Their findings indicate that the incorporation of AI techniques can markedly decrease diagnostic delays and enhance the precision of identifying potentially cancerous conditions. Offering an alternative perspective, [21] enhances this discourse by examining the utilization of image recognition algorithms in the detection of oral cancer. The discourse centers on the progress in computer image processing, highlighting the capacity of artificial neural networks to improve diagnostic precision while alleviating the burden on healthcare practitioners. A systematic evaluation and meta-analysis of automated classification approaches for oral possibly malignant and malignant illnesses is provided in [22]. They emphasize the capability of machine learning algorithms to function as efficient screening instruments, particularly in resource-limited environments where expert analysis may be scarce. Likewise, [23] elaborates on the significance of machine learning and deep learning in cancer diagnosis, highlighting the rapid increase in studies related to computer-aided diagnosis. The author underscores the pressing necessity for swift and precise diagnostic instruments, especially for early cancer identification, which markedly enhances survival rates. Ultimately, [24] and [25] showcase recent progress in the identification of OSCC utilizing diverse deep learning frameworks. Their research illustrates the effectiveness of these models in precisely detecting malignant lesions from histopathological pictures and various imaging modalities, underscoring the promise of deep learning-based methods to transform oral cancer diagnostics. Table 1 lists the comparisons of the related studies.

The procedural steps of the method presented in this research are outlined as follows.

The dataset comprising photos of oral cancer was obtained from the Kaggle website [26].

• In response to "iCCP: extra compressed data" warnings received while reading certain photos in the dataset, all images were changed from JPEG to PNG format to enhance compatibility. Of the 1,231 Oral Cancer photos in the collection, 40 images (28 malignant and 12 benign) persisted in generating read errors post-conversion. The problematic photographs were eliminated from the dataset, and the remaining images advanced to the subsequent stages of the procedure.

The Oral Cancer dataset was successively trained using the GoogleNet and MobileNet-v2 deep learning models to get results. Each model was sequentially trained on the dataset, and performance parameters were documented to assess and compare their efficacy in classifying oral cancer images.

The characteristics derived from the two models were processed independently prior to the classification layer. Precisely, 1,000 features were retrieved from the loss3-classifier layer of GoogleNet and the Logits layer of MobileNet-v2. The retrieved characteristics were subsequently concatenated to create a hybrid model. This hybrid model was thereafter employed for classification using conventional classifiers, and the results were acquired to evaluate its efficacy.

The study's main contribution is the creation of a hybrid methodology that integrates the feature extraction skills of contemporary deep learning models with the classification precision of conventional machine learning techniques. Utilizing GoogleNet and MobileNetV2 as pretrained feature extractors, followed by classical classifiers such as Quadratic SVM and K-Nearest Neighbors, the method attains a notable enhancement in classification performance, with a maximum accuracy of 90.01%. The methodology exhibits computational efficiency, achieving training durations of less than three minutes for both deep learning models, and offers a thorough assessment of several classical classifiers to choose the most effective combination. The paper presents a scalable and robust diagnostic approach, appropriate for resource-constrained environments, by integrating modern deep learning technologies with conventional methodologies. This novel method increases the dependability of oral cancer diagnostics, leading to enhanced diagnostic precision and the possibility of early intervention in medical image analysis.

In this study, Section 2 details the dataset used, including its composition, preprocessing steps, and the challenges encountered during image conversion and cleaning. Section 3 focuses on the case study, describing the training and evaluation of the deep learning models (GoogleNet and MobileNetV2), feature extraction processes, and the subsequent application of classical classifiers. Detailed training parameters, computational resources, and performance metrics are presented, accompanied by visual representations such as confusion matrices and accuracy plots. The paper concludes with a discussion of the results, emphasizing the hybrid model's superior performance, potential applications in medical diagnostics, and avenues for future research. Supporting these sections are relevant references and technical details, ensuring the study's reproducibility and relevance.

Table 1. Comparisons of the method in this paper with some existing methods

Study	Methods Used	Dataset Details	Best Accuracy (%)	Key Contributions
This Study	GoogleNet, MobileNetV2 + Classical Classifiers (SVM, KNN)	1,231 images (Cancer/Non-Cancer)	90.01 (Quadratic SVM)	Hybrid approach combining DNNs and traditional classifiers
Lin et al., 2021 [19]	Smartphone images + Deep Learning	Smartphone images	86.5	Early detection of OSCC using accessible technology
Ren et al., 2021 [18]	CNNs for histological grading	Dental, oral, and craniofacial images	87.0	Automating histological grading of oral cancer lesions
Ferro et al., 2022 [22]	ML for point-of-care classification	Malignant disorders	89.5	Screening tools for low-resource settings
Albalawi et al., 2024 [24]	EfficientNet for histopathology	OSCC histopathological images	89.3	Accurate classification of histopathological images
Zhang et al., 2021 [21]	Image recognition techniques + ANN	Oral cancer cell images	88.6	Improved diagnostic accuracy with ANN-based techniques
Warin et al., 2022 [15]	Deep CNNs for lesion analysis	Oral lesion datasets	87.4	Novel deep learning framework for lesion detection
García-Pola et al., 2021 [20]	AI for early oral cancer diagnosis	Images of precancerous conditions	86.9	Review of AI applications for early detection
Sugeno et al., 2021 [10]	Transfer learning + image processing	Retinopathy images adapted for cancer	85.7	Simple and efficient preprocessing for transfer learning
Hunter et al., 2022 [7]	AI in cancer diagnostics	Comprehensive review	89.0	Overview of AI's role in early detection of oral cancer

MATERIALS AND METHODS

The dataset that was utilized in this investigation was a publically accessible version of an oral cancer image dataset that was obtained from Kaggle [26]. Sample images from the dataset can be seen in Figure 1. The dataset included 1,231 photos that were divided into two categories: cancer and non-cancer. While there are 654 photos that do not include any cancerous cells, there are 577 photographs that contain cancerous cells. The file size of the dataset is roughly 201 megabytes, and the photographs are saved in the JPEG format. The color depth of the images is 24 bits.

Ensuring a minimum resolution of 400×400 pixels provides sufficient detail for classification tasks. All of the photos were converted to the PNG format in order to resolve compatibility concerns and prevent "iCCP: extra compressed data" warnings from appearing while the processing was being done. Despite preprocessing, 40 images (28 cancerous, 12 non-cancerous) still caused read errors and were removed.

Figure 1. Sample images from the dataset

This resulted in a total of 1,191 images being available for training and testing purposes. Eighty percent of the photos in the dataset are used for training, while twenty percent are kept aside for testing. The dataset is divided into an 80/20 split. During the process of resizing each image to the usual dimensions of 224 by 224 pixels, the three RGB color channels are preserved. In addition, the images are normalized in order to guarantee homogeneity and significantly enhance the performance of the model during both the training and testing phases. This dataset is highly renowned in the field of cancer research and medical imaging investigations. It serves as an invaluable resource for the development and testing of diagnostic frameworks for oral cancer.

RESULTS AND DISCUSSION

The training parameters are listed in Table 2. The network training utilized Stochastic Gradient Descent with Momentum (SGDM) as the optimization algorithm, selected for its capacity to enhance convergence speed and stability. To improve training efficiency, a stochastic solution was employed in conjunction with SGDM, while parallel computing on a GPU facilitated 16 simultaneous workers, hence expediting processing times. The primary training parameters comprised an initial learning rate of 1e-4 and shuffling at each epoch, a measure implemented to reduce overfitting. The training environment was explicitly optimized for GPU processing, fully utilizing hardware acceleration. Table 2 presents the comprehensive characteristics of the computational setup, encompassing hardware capabilities. This optimized architecture guaranteed effective resource consumption, facilitating quicker and more stable convergence during the model's training phase.

Table 2. Specifications of the computer

Processor	12th Gen Intel(R) Core^{TM} i9-12900F 2.40 GHz
Cores, Processors	16, 24
Installed RAM	64.0 GB (63.7 GB usable)
GPU	NVIDIA RTX A4000
DirectX version	12 (FL 12.1)
GPU Memory	47.9 GB (16.0 GB Dedicated, 31.9 GB Shared)

Upon completion of the training process, the GoogleNet model achieved a training accuracy of 0.6807, with the full procedure finalized in a remarkable duration of 2 minutes and 36 seconds. The short training time reflects the efficiency of the system and the optimized training settings. Figure 2 depicts the confusion matrix generated from this training phase.

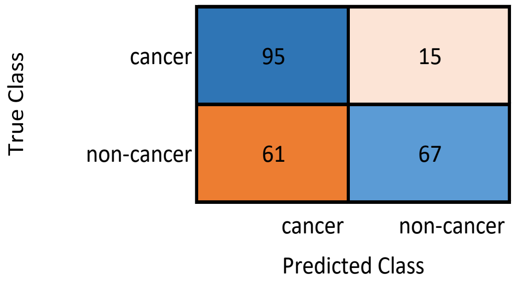

Figure 2. Confusion matrix of the training process of GoogleNet

The examination of the confusion matrix indicates that among the 238 test samples, there are 110 Cancer images and 128 Non-Cancer images. Out of the Cancer photos, 95 were correctly diagnosed, whilst 15 were erroneously labeled as Non-Cancer. In the Non-Cancer category, of the 128 photos, 67 were accurately predicted, highlighting certain categorization difficulties for the algorithm. Table 3 delineates essential aspects of the GoogleNet training process, encompassing variables such as iteration progression, duration per iteration, mini-batch efficacy, test accuracy, and error rates. This data offers insight into the model's learning progression, efficacy, and overall classification performance. Figure 3 illustrates the accuracy and loss progression of the GoogleNet training process. Figure 3 illustrates that the GoogleNet network attained a test accuracy of 68.07% during the training phase.

Table 3. The progress of training of GoogleNet

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss
1	1	00:00:12	50.00%	47.90%	16.274	11.974
1	15	00:00:20	56.25%	49.16%	10.379	10.030
1	30	00:00:27	68.75%	57.56%	0.7070	0.9143
1	45	00:00:34	75.00%	59.24%	0.9016	0.9249
1	50	00:00:36	68.75%			0.6000
2	60	00:00:41	75.00%	57.56%	0.3702	0.9833
2	75	00:00:48	93.75%	57.98%	0.3177	0.9708
2	90	00:00:55	68.75%	62.61%	0.5695	0.9064
2	100	00:00:59	81.25%			0.3176
2	105	00:01:03	68.75%	57.14%	0.5188	10.370
3	120	00:01:10	75.00%	67.65%	0.6118	0.8420
3	135	00:01:18	75.00%	58.82%	0.6285	0.9316
3	150	00:01:25	93.75%	63.87%	0.3028	0.8870
3	165	00:01:31	87.50%	65.97%	0.2769	0.8455
4	180	00:01:39	81.25%	63.87%	0.4493	0.8942
4	195	00:01:45	87.50%	70.17%	0.2318	0.7939
4	200	00:01:47	100.00%			0.0953

4	210	00:01:53	93.75%	68.91%	0.2183	0.8161
4	225	00:02:01	87.50%	60.08%	0.3801	0.9401
5	240	00:02:08	93.75%	69.75%	0.3743	0.8133
5	250	00:02:11	81.25%			0.4585
5	255	00:02:14	100.00%	69.75%	0.0912	0.7912
5	270	00:02:20	87.50%	65.13%	0.2370	0.8850
5	285	00:02:29	93.75%	69.33%	0.2168	0.8138
5	295	00:02:34	81.25%	68.07%	0.2710	0.8405

Simultaneously, the alternative element of the hybrid model, MobileNet-v2, achieved a training accuracy of 67.65%, concluding its training in 2 minutes and 58 seconds. This comparison underscores the performance and efficiency of each model inside the hybrid framework, illustrating their complimentary strengths in the categorization task. Figure 4 presents the confusion matrix for the training of MobileNet-v2. Figure 4's confusion matrix reveals that the 238 test samples comprise 110 Cancer images and 128 Non-Cancer images. Out of the Cancer pictures, 95 were accurately classified, whilst 14 were inaccurately classified. Of the Non-Cancer pictures, 65 out of 128 were accurately predicted. Table 4 presents a detailed summary of the MobileNet-v2 training process, emphasizing critical parameters including iteration progress, duration per iteration, mini-batch accuracy, test accuracy, and error rates. These results offer valuable insights into model performance and training efficiency. Figure 5 illustrates the accuracy and loss progression of the MobileNet-v2 training process. Figure 5 illustrates that the test accuracy attained during the training of the MobileNet-v2 network was 67.65%.

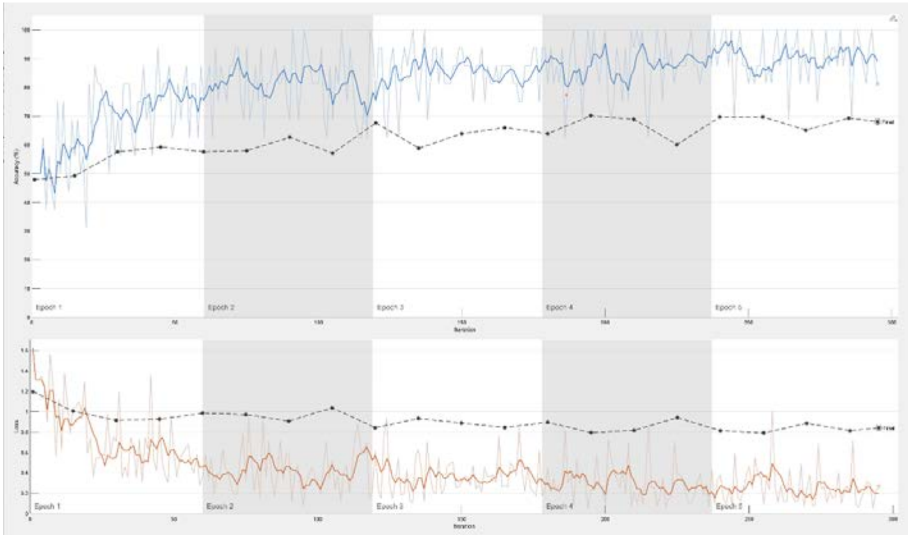

Figure 3. The accuracy and loss progress of the GoogleNet training process

Table 4. The progress of training of MobileNet-v2

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss
1	1	00:00:08	50.00%	45.80%	0.9080	0.8250
1	20	00:00:17	41.67%	47.90%	0.8768	0.7863
1	40	00:00:25	50.00%	52.94%	0.6742	0.7953
1	50	00:00:29	75.00%			0.4654
1	60	00:00:34	58.33%	56.30%	0.6729	0.7571
2	80	00:00:43	75.00%	57.14%	0.5446	0.7546
2	100	00:00:50	83.33%	58.40%	0.3013	0.7855
2	120	00:00:58	75.00%	57.14%	0.5326	0.7758
2	140	00:01:08	100.00%	60.08%	0.1639	0.7769
2	150	00:01:11	75.00%			0.3534
3	160	00:01:16	91.67%	62.18%	0.2688	0.7320
3	180	00:01:26	83.33%	58.82%	0.3174	0.7914
3	200	00:01:33	91.67%	60.08%	0.3046	0.7737
3	220	00:01:41	91.67%	59.66%	0.1606	0.7311
4	240	00:01:50	83.33%	63.45%	0.3049	0.7259

4	250	00:01:53	100.00%			0.0881
4	260	00:01:58	83.33%	63.45%	0.3204	0.7721
4	280	00:02:05	91.67%	63.45%	0.2156	0.7191
4	300	00:02:13	83.33%	65.55%	0.3288	0.7382
5	320	00:02:24	83.33%	63.03%	0.3401	0.7287
5	340	00:02:32	91.67%	65.13%	0.3390	0.7053
5	350	00:02:36	83.33%			0.4437
5	360	00:02:42	100.00%	66.39%	0.2115	0.7646
5	380	00:02:49	91.67%	62.61%	0.2080	0.7314
5	395	00:02:56	75.00%	62.61%	0.5624	0.7450

The features obtained prior to the classification layers from both GoogleNet and MobileNet-v2 were subsequently utilized for classification employing conventional classifiers, including Support Vector Machine (SVM), Neural Network, K-Nearest Neighbors (KNN), Logistic Regression, Ensemble methods, Discriminant Analysis, among others. The outcomes of these categories are displayed in Table 5. Upon evaluation of the outcomes, the Quadratic SVM model achieved an accuracy of 90.01%, representing a substantial enhancement of 21.94% relative to the baseline training accuracy of GoogleNet.
Likewise, MobileNet-v2 exhibited a 22.36% enhancement in its first training accuracy, illustrating the efficacy of employing conventional classifiers on the retrieved features for improved performance. The confusion matrix derived from the Quadratic SVM model is presented in Figure 6.

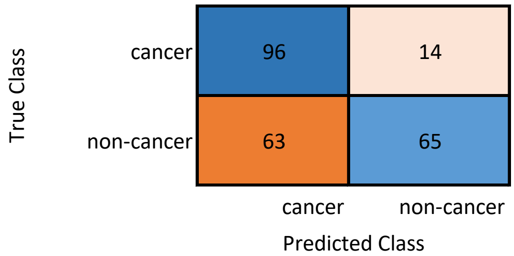

Figure 4. Confusion matrix of the training process of MobileNet-v2

An examination of the confusion matrix for the classical classifier with the best accuracy reveals that the dataset comprises 1,191 training and test samples, including 549 Cancer images and 642 Non-Cancer images.

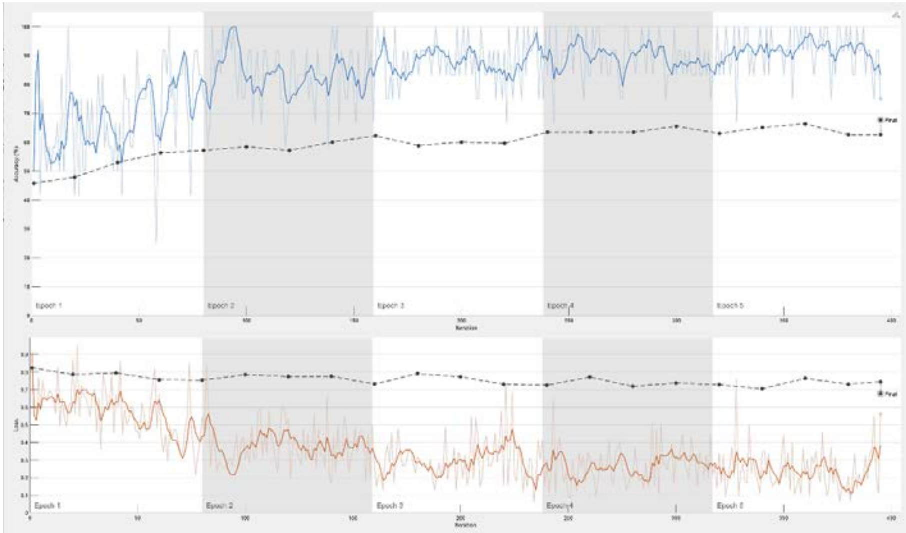

Figure 5. The accuracy and loss progress of the MobileNet-v2 training process

Table 5. Classification accuracies of the top 20 clasifiers

No	Models	Sub Models	Accuracy (%)
1	***SVM***	***Quadratic SVM***	***90,01%***
2	SVM	Cubic SVM	89,76%
3	Kernel	SVM Kernel	89,59%
4	SVM	Medium Gaussian SVM	89,50%
5	Neural Network	Wide Neural Network	88,75%
6	KNN	Weighted KNN	88,16%
7	SVM	Linear SVM	87,99%
8	Neural Network	Medium Neural Network	87,32%
9	Efficient Linear SVM	Efficient Linear SVM	86,99%

10	Kernel	Logistic Regression Kernel	86,99%
11	Neural Network	Narrow Neural Network	86,82%
12	KNN	Cubic KNN	86,40%
13	KNN	Medium KNN	86,31%
14	Ensemble	Boosted Trees	86,23%
15	Ensemble	Bagged Trees	86,23%
16	Neural Network	Bilayered Neural Network	86,06%
17	Discriminant	Linear Discriminant	85,98%
18	SVM	Coarse Gaussian SVM	85,89%
19	Ensemble	Subspace KNN	85,89%
20	KNN	Fine KNN	85,73%

Out of the Cancer pictures, 481 were accurately predicted, whilst 68 were incorrectly identified. Out of the Non-Cancer photos, 591 were correctly classified, while 51 were erroneously classed as Cancer. These results underscore the classifier's robust capacity to differentiate between Cancer and Non-Cancer images, demonstrating its efficacy in image classification.

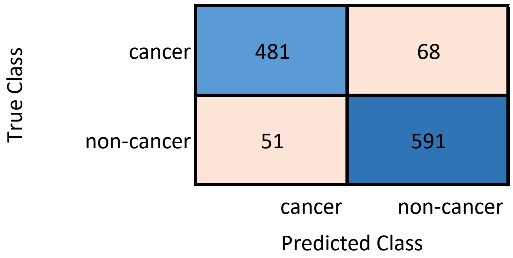

Figure 6. Confusion matrix obtained with the Quadratic SVM model

In addition, the ROC curve obtained with the Quadratic SVM model is given in figure 7 to be informative.

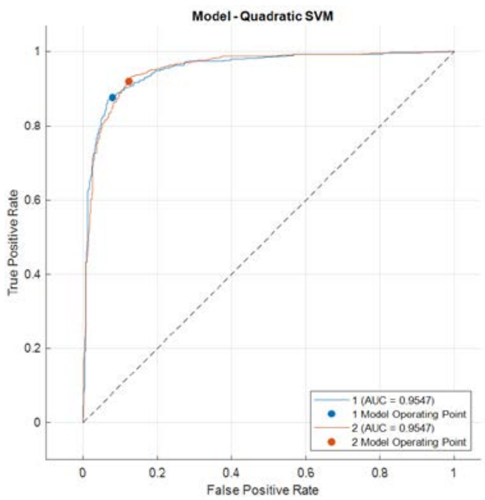

Figure 7. ROC curve obtained with the Quadratic SVM model

This study presents a hybrid methodology for classifying oral cancer photos by integrating the feature extraction proficiency of contemporary deep neural networks (GoogleNet and MobileNetV2) with the classification precision of conventional machine learning methods. The results illustrate the efficacy of this method, attaining a notable enhancement in classification accuracy relative to independent deep learning models. The Quadratic SVM classifier attained the best accuracy of 90.01%, representing an enhancement of more than 22% compared to the baseline accuracies of GoogleNet (68.07%) and MobileNetV2 (67.65%).

Conducting a statistical analysis using additional metrics that are crucial in clinical decision-making, including sensitivity (recall), specificity, precision, F1-score, and the confusion matrix-derived rates such as false positive rate (FPR) and false negative rate (FNR) should be useful.

- Sensitivity (Recall) measures the model's ability to correctly identify positive cases (i.e., cancerous images). It is particularly important in early diagnosis to minimize the risk of missed detections.
- Specificity reflects how effectively the model avoids false alarms by correctly identifying non-cancerous cases.
- Precision indicates the reliability of positive predictions, which is critical in medical screening to reduce unnecessary interventions.
- F1-score, the harmonic mean of precision and recall, provides a balanced view when dealing with imbalanced datasets or unequal costs of misclassification.
- FPR and FNR further quantify the types of classification errors, offering insight into whether the model tends to overpredict or underpredict cancer presence.

For the Quadratic SVM classifier, which achieved the highest accuracy (90.01%), the following metrics were observed:

- **Sensitivity**: 87.61%
- **Specificity**: 92.05%
- **Precision**: 90.18%
- **F1-score**: 88.88%
- **False Positive Rate (FPR)**: 7.95%
- **False Negative Rate (FNR)**: 12.39%

These results indicate that the model maintains a high balance between detecting actual cancer cases and minimizing false alarms. Notably, the low FPR and high specificity demonstrate the model's robustness in reducing misclassification of healthy patients as diseased, which is essential in screening contexts where overdiagnosis can lead to undue stress and resource burden.

Incorporating these diagnostic metrics provides a more comprehensive assessment of the model's performance and its clinical applicability. Future versions of the study will continue to integrate these evaluations to better align with the standards of evidence required in medical diagnostics.

This finding highlights the possibility of combining classical classifiers with features derived from pretrained deep neural networks, particularly in the realm of medical picture analysis. The hybrid method addresses several challenges common in deep learning-based medical diagnostics. Utilizing pretrained networks diminishes the necessity for big datasets, a significant constraint in medical imaging arising from privacy issues and the challenges of acquiring labeled data. The utilization of classical classifiers improves computational efficiency, as these models demand considerably fewer resources than end-to-end deep learning training. The proposed technique exhibits robust performance; however, specific restrictions must be recognized. The dataset included in this work, while balanced between Cancer and Non-Cancer categories, is rather small, potentially affecting the model's generalizability to varied and unobserved data. Enhancing the dataset with augmented or outside sourced photos may further elevate performance and resilience. Furthermore, the assessment relies exclusively on accuracy metrics; incorporating additional performance indicators such as sensitivity, specificity, and the area under the ROC curve (AUC) would yield a more thorough evaluation of the model's diagnostic efficacy. Future study may investigate the utilization of advanced architectures like Vision Transformers or EfficientNet to enhance feature extraction. Furthermore, the integration of explainability approaches such as Grad-CAM may yield significant insights into model predictions, enhancing confidence and transparency in clinical applications. Evaluating the hybrid model on additional medical imaging datasets would demonstrate its versatility and scalability. This study underscores the viability and promise of hybrid frameworks in enhancing the precision and efficacy of medical picture classification. The amalgamation of deep learning and conventional machine learning methodologies presents a promising avenue for the creation of dependable diagnostic instruments, especially in resource-limited settings. This technology could significantly enhance early diagnosis and treatment of oral cancer by addressing the stated limitations and broadening the research area. The suggested hybrid model, which combines pretrained deep neural networks (GoogleNet and MobileNetV2) with classical classifiers such Quadratic SVM, attains a classification accuracy of 90.01%, reflecting a significant enhancement of more than 22% relative to the use of deep learning models in isolation. This outcome surpasses numerous previous research in oral cancer classification. Lin et al. [19] demonstrated an accuracy of 86.5% in early oral cancer diagnosis utilizing deep learning on smartphone-based photos, whereas Ren et al. [18] attained 87.0% accuracy in histological grading employing CNN models. Ferro et al. [22] devised a machine learning classification method for malignant illnesses, attaining an accuracy of 89.5%, whereas Zhang et al. [21] utilized artificial neural networks (ANNs) for oral cancer detection, achieving an accuracy of 88.6%. In contrast to conventional methods, the hybrid technique introduced in this paper leverages deep learning for feature extraction alongside the discriminative capabilities of traditional machine learning classifiers, hence improving both accuracy and efficiency. In contrast to end-to-end deep learning models that necessitate substantial labeled datasets and significant computational resources, the proposed hybrid technique efficiently utilizes pretrained networks for feature extraction, alleviating the computational load while preserving superior diagnostic performance. Furthermore, the study's findings suggest that traditional classifiers, such as Quadratic SVM, can substantially enhance classification accuracy when utilized with features produced from deep learning, highlighting the efficacy of hybrid approaches in medical picture analysis. These findings highlight the necessity of combining contemporary deep learning techniques with conventional classification methods, presenting a promising approach to improving oral cancer diagnostics, especially in resource-limited clinical environments where computational efficiency and reliability are essential. Subsequent study may investigate the applicability of this hybrid framework to more medical imaging fields, potentially resulting in enhanced and more accessible diagnostic solutions. While this study utilized GoogleNet and MobileNetV2 for feature extraction due to their well-established performance-to-efficiency balance, it is important to acknowledge the potential of newer and more advanced deep learning architectures such as EfficientNet. EfficientNet, introduced by Tan and Le, employs a compound scaling method to balance network depth, width, and resolution, achieving state-of-the-art accuracy with significantly fewer parameters and FLOPs compared to traditional CNNs. Its scalable nature makes it particularly suitable for medical image analysis tasks, where both accuracy and computational efficiency are critical. Given these advantages, incorporating EfficientNet as a feature extractor could further enhance diagnostic performance or offer competitive results with improved resource efficiency. In future work, we plan to conduct comparative analyses involving EfficientNet, and potentially other architectures like ResNeSt and Vision Transformers, to comprehensively evaluate the robustness and adaptability of our hybrid classification framework.

Such comparisons will also help validate the generalizability of the proposed approach across different network designs and dataset characteristics.
A notable limitation of this study lies in the use of a single publicly available dataset, sourced from Kaggle, which may not fully capture the variability found in real-world clinical settings. While the dataset includes a balanced distribution of cancerous and non-cancerous images, its size and demographic diversity are limited. This constraint may affect the generalizability of the proposed hybrid model, particularly when applied to different imaging environments, devices, or patient populations with varied ethnic, age, or clinical profiles. Additionally, the dataset represents a single imaging modality and lacks metadata that could help stratify results based on anatomical location, disease stage, or imaging conditions. To enhance robustness and clinical applicability, future studies should explore multi-center datasets, incorporate cross-institutional image repositories, and test the model on different imaging modalities such as histopathological slides, intraoral camera images, or fluorescence images. Moreover, analyzing performance across distinct patient subgroups can help uncover potential biases and improve the reliability of diagnostic outputs across diverse populations. Expanding the dataset scope in these ways would strengthen the model's clinical readiness and its ability to generalize across real-world scenarios.

CONCLUSIONS

This paper presents a novel hybrid method for oral cancer image detection that combines pretrained deep neural networks with traditional machine learning classifiers. The findings indicate that integrating the feature extraction abilities of GoogleNet and MobileNetV2 with conventional classifiers like Quadratic SVM significantly enhances diagnosis accuracy. This methodology effectively balances computational economy and performance, yielding solid findings in a resource-efficient manner. The proposed hybrid framework combines deep learning and traditional methods to provide an effective tool for medical image interpretation. Its relevance to a publically accessible oral cancer dataset underscores its potential as a reliable diagnostic instrument, especially in contexts with limited computational resources or restricted availability of labeled data. The results of this study facilitate the advancement of scalable diagnostic systems that can improve the early diagnosis of oral cancer. This approach improves classification accuracy and supports future integration of AI in clinical settings, aiming to enhance patient outcomes and diagnostic systems.

REFERENCES

[1] Shigeishi H. Association between human papillomavirus and oral cancer: a literature review, International Journal of Clinical Oncology. 28:8 982-989, 2023.
[2] Farsi S., Gardner J.R., King D., Sunde J., Moreno M., Vural E. Head and neck cancer surveillance: The value of computed tomography and clinical exam, American Journal of Otolaryngology. 45:6 104469, 2024.
[3] García-Pola M., Pons-Fuster E., Suárez-Fernández C., Seoane-Romero J., Romero-Méndez A., López-Jornet P. Role of artificial intelligence in the early diagnosis of oral cancer. A scoping review, Cancers. 13:18 4600, 2021.
[4] Elmaghraby D.A., Alshalla A.A., Alyahyan A., Altaweel M., Al ben Hamad A.M., Alhunfoosh K.M., Albahrani M.A. Public knowledge, practice, and attitude regarding cancer screening: a community-based study in Saudi Arabia, International Journal of Environmental Research and Public Health. 20:2 1114, 2023.
[5] Nagao T., Warnakulasuriya S. Screening for oral cancer: Future prospects, research and policy development for Asia, Oral Oncology. 105 104632, 2020.
[6] Borse V., Konwar A.N., Buragohain P. Oral cancer diagnosis and perspectives in India, Sensors International. 1 100046, 2020.
[7] Hunter B., Hindocha S., Lee R.W. The role of artificial intelligence in early cancer diagnosis, Cancers. 14:6 1524, 2022.
[8] Cai L., Gao J., Zhao D. A review of the application of deep learning in medical image classification and segmentation, Annals of Translational Medicine. 8:11, 2020.
[9] Yu H., Yang L.T., Zhang Q., Armstrong D., Deen M.J. Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives, Neurocomputing. 444 92-110, 2021.
[10] Sugeno A., Ishikawa Y., Ohshima T., Muramatsu R. Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning, Computers in Biology and Medicine. 137 104795, 2021.
[11] Muhammad K., Khan S., Del Ser J., De Albuquerque V.H.C. Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey, IEEE Transactions on Neural Networks and Learning Systems. 32:2 507-522, 2020.
[12] Waring J., Lindvall C., Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare, Artificial Intelligence in Medicine. 104 101822, 2020.
[13] Himeur Y., Elnour M., Fadli F., Meskin N., Petri I., Rezgui Y., Amira A. AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives, Artificial Intelligence Review. 56:6 4929-5021, 2023.
[14] Wambugu N., Chen Y., Xiao Z., Tan K., Wei M., Liu X., Li J. Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review, International Journal of Applied Earth Observation and Geoinformation. 105 102603, 2021.
[15] Warin K., Limprasert W., Suebnukarn S., Jinaporntham S., Jantana P., Vicharueang S. AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer, PLOS One. 17:8 e0273508, 2022.

[16] Chillakuru P., Madiajagan M., Prashanth K.V., Ambala S., Shaker Reddy P.C., Pavan J. Enhancing wind power monitoring through motion deblurring with modified GoogleNet algorithm, Soft Computing. 1-11, 2023.
[17] Dong K., Zhou C., Ruan Y., Li Y. MobileNetV2 model for image classification, 2nd International Conference on Information Technology and Computer Application (ITCA), IEEE. 476-480, 2020.
[18] Ren R., Luo H., Su C., Yao Y., Liao W. Machine learning in dental, oral and craniofacial imaging: a review of recent progress, PeerJ. 9 e11451, 2021.
[19] Lin H., Chen H., Weng L., Shao J., Lin J. Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis, Journal of Biomedical Optics. 26:8 086007, 2021.
[20] García-Pola M., Pons-Fuster E., Suárez-Fernández C., Seoane-Romero J., Romero-Méndez A., López-Jornet P. Role of artificial intelligence in the early diagnosis of oral cancer. A scoping review, Cancers. 13:18 4600, 2021.
[21] Zhang H., Li W., Zhang H. An Image Recognition Framework for Oral Cancer Cells, Journal of Healthcare Engineering. 2021:1 2449128, 2021.
[22] Ferro A., Kotecha S., Fan K. Machine learning in point-of-care automated classification of oral potentially malignant and malignant disorders: a systematic review and meta-analysis, Scientific Reports. 12:1 13797, 2022.
[23] Bechelli S. Computer-aided cancer diagnosis via machine learning and deep learning: a comparative review, arXiv preprint arXiv:2210.11943, 2022.
[24] Albalawi E., Thakur A., Ramakrishna M.T., Bhatia Khan S., SankaraNarayanan S., Almarri B., Hadi T.H. Oral squamous cell carcinoma detection using EfficientNet on histopathological images, Frontiers in Medicine. 10 1349336, 2024.
[25] Warin K., Suebnukarn S. Deep learning in oral cancer—a systematic review, BMC Oral Health. 24:1 212, 2024.
[26] https://www.kaggle.com/datasets/fatmaahmedabdallah/oral-cancer-dataset, Last Accessed on 30.10.2024