CAN WEIGHTED FEATURE ALLOCATION AND MULTI-LEVEL CLASSIFICATION IMPROVE VGG16-BASED ORAL CANCER CLASSIFICATION FROM DIGITAL IMAGES?

PATIBANDLA SANDHYA KRISHNA[1,2], SUBBA RAO PERAM[1]

[1]Department of IT and Computer Applications, School of computing and informatics, Vignan's Foundation for Science, Technology & Research (Deemed to be university), Vadlamudi, Guntur-522213, Andhra Pradesh, India.

[2]Vignan's Nirula Institute of Technology and Science for Women, Peda Palakaluru, Andhra Pradesh 522009, India

E-mail: [1]sandhyapatibandla2@gmail.com [2] drpsr_it@vignan.ac.in

ABSTRACT

Oral squamous cell carcinoma (OSCC) is a serious health problem, and early detection of cancerous tissues is crucial. While deep learning has shown great promise in the field of automated oral cancer image analysis, current methods are increasingly becoming complex, such as those involving attention, transformer, ensemble, segmentation, and multimodal architectures, with comparatively less focus on explicitly emphasizing discriminative deep features prior to classification. In this study the gap is addressed by proposing a Weighted Feature Set with Multi-Level Classification using VGG16 (WFS-MLC-VGG) for automated oral cancer image classification. This framework first extracts deep representations from VGG16 and then performs feature selection, correlation-based feature weighting and multi-level classification to mitigate the impact of redundant features and highlight discriminative features. Experimental results show that the proposed model performs with an accuracy of 98.4% for feature-weight allocation and 98.8% for classification at the maximum number of images tested and the reported feature-processing times are smaller than those of the alternative models investigated for the current work. The findings offer evidence that while there is no need for adding more to the architecture of the CNN to get explicit feature prioritization, it can still be achieved by using explicit feature prioritization. The major contribution in the principal is a feature-management approach that combines weighted feature allocation and multi-level classification, as part of a diagnostic pipeline based on VGG16. The outcome, however, can only be applied to the dataset and experimental conditions used to evaluate the model, and clinical validation is required by external factors. The study thus confirms the methodological feasibility of the weighted feature processing for classification of oral cancer images and provides key directions in the way of developing robust, scalable and clinically validated computer aided diagnosis.

Keywords: *Oral Cancer, Feature Subset, Deep Learning, Convolution Neural Network, Histopathology Classification.*

1. INTRODUCTION

Over 90% of oral squamous cell carcinomas (OSCC) begin in the oral mucosa and there is a wide range of subtypes within this cancer group. In terms of prevalence, head and neck squamous cell carcinoma (HNSCC) ranks seventh [1]. The World Health Organization estimates that over 330,000 people died and over 657,000 new cases are identified every year. The incidence of OSCC is significantly greater in South Asian countries. One's ethnic background, lifestyle, environment, and geographical location can all have a role. Early detection of OSCC is crucial for effective therapy, increased survival rates, and decreased morbidity and death [2]. The outlook for OSCC is dismal, with a median cure rate of just 50%. The definitive method for identifying OSCC is histopathological evaluation of tissue samples under a microscope [3]. The clinical utility of diagnostic pathology is diminished due to the time and error involved in histopathologists' evaluation of specimens. Because of this, it is critical to equip pathologists with accurate methods for assessing and diagnosing OSCC [4].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Recent years have seen an explosion in studies examining the potential of AI for application in medical diagnosis. The increasing availability of diagnostic imaging has given researchers the opportunity to explore the use of artificial intelligence in the analysis of medical images. There are several challenges in the field of medical image processing, but Deep Learning (DL) [5] has shown excellent effectiveness in tackling these issues, especially in the diagnosis of cancer images. DL-based Computer Aided Diagnostic (CAD) systems [6] have been suggested and implemented at scale for various forms of cancer, including breast, prostate, and lung cancers. But research indicates that DL has been underutilized in OSCC image diagnosis [7]. Oral histology photos with keratin pearls were analyzed using Convolution Neural Network (CNN). Keratin regions were segregated by the CNN model [8].

Using appearance models and stereotypes, tumours in lesion and lesion-free forms can be recognized in a variety of anatomical locations without the use of staining. Machine learning algorithms were used to make predictions about several biological models for OSCC; these models identify non-cancerous and malignant samples; these samples are then examined for oral cancer stages [9]. The predictor will determine the validity of the interplay by using three different rational test kits and a range of cancer stages. Sampling with sample justification allows for the prediction of several phases of oral cancer, including the size of tumours and the development of ulcers in the tissues [10]. The fundamental goal of this research is to develop novel approaches to estimate the development of oral cancer tumours. The front of the tongue, the interior of the cheeks and lips, the gums, and the area below the wisdom teeth are just some of the places in the mouth where cancer can grow [11]. Most people with cancer experience discomfort or even bleeding due to chronic inflammation or ulcers that fail to heal. Smoking and excessive alcohol uses are two of many risk factors for developing mouth cancer [12]. These two habits are highly predictive of developing oral cancer.

Deep learning is a subfield of machine learning that employs methods inspired by the brain's structure and function. Both cancerous and noncancerous lesion examples are available for use when training a neural network model [13]. The model has learned the non-linear interactions and can tell if the image is malignant or not. As a result, deep learning does away with the need for features to be extracted by domain experts [14]. In this research, CNN based VGG 16 model is employed to construct a deep learning model that can tell the difference between healthy and malignant oral tissue in oral images [15]. A CNN is an artificial neural network consisting of multiple layers. How neurons act is determined by the weights given to them. The basic CNN architecture is shown in Figure 1.

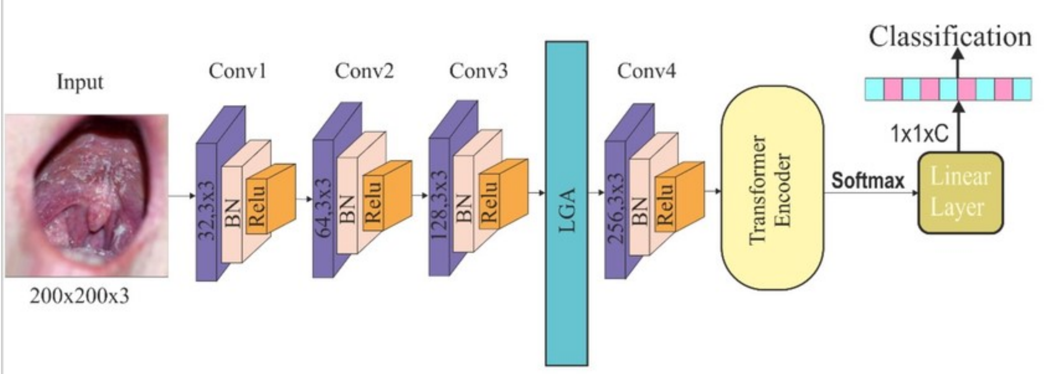

Fig 1: CNN Architecture [11]

Every typical CNN consists of an input layer, a convolution layer, a pooling layer, and a fully connected (FC) layer. By applying filters to the input image, convolution layers build feature maps [16], while pooling layers replace the network's output in those regions with a summary statistic computed from the outputs in the surrounding regions [17]. Because of the reduced area involved, less computational effort is required. A neuron in a fully connected layer has connections to all of the cells in the layer above and below it. The FC layer mediates communication between the source and destination representations [18]. A softmax layer is frequently used to classify the results of image classification. CNNs get their class knowledge from labelled training data. The feature selection process is shown in Figure 2.

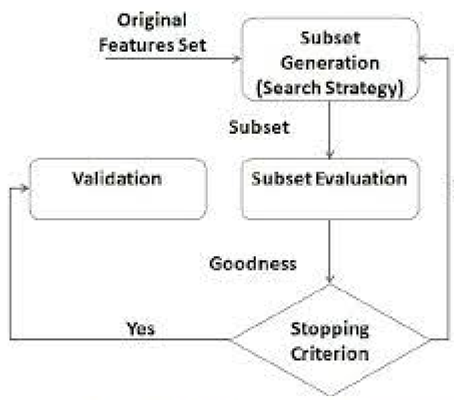

Fig 2: Feature Selection Process

In the field of vision modelling, VGG16 is widely regarded as one of the CNN architectures. Convolution layers in VGG16 use a 3x3 filter and a stride 1 in the exact same padding and maxpool layer as a 2x2 filter and a stride 2 in order to reduce the number of hyper-parameters needed [19]. The VGG16 layers are represented in Figure 3.

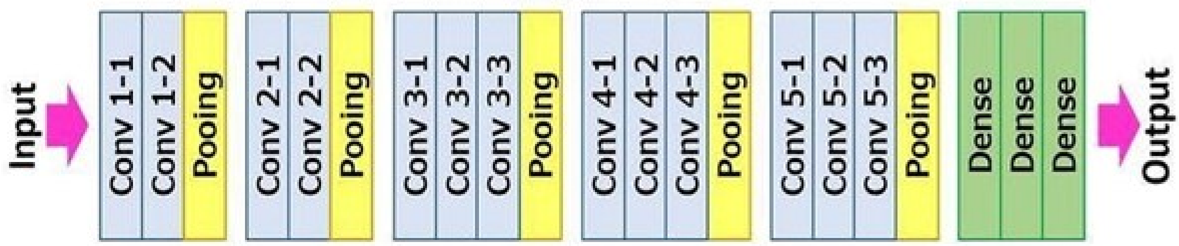

Fig 3: VGG16 Layers

The design is uniform in its use of convolution and max pool layers in this configuration. It concludes with two fully linked layers and a softmax. There are 16 weighted layers, hence the 16 in VGG16 [20]. This research develops a Weighted Feature Set model with Multi Level Classification model using VGG 16 model for accurate classification and detection of oral cancer. This network is quite big, with an estimated 138 million parameters.

While it has been shown that a deep learning model can attain good performance in oral cancer image analysis, the literature does not clearly indicate whether feature prioritization in the end-to-end deep CNN can offer an effective trade-off between discriminative representation and computational efficiency. Most recent studies cited in this paper enhance the performance by architecturally adapting the network, such as deformable convolutions, attention mechanisms, multi-scale transformers, segmentation networks, ensemble learning or multimodal data integration. In the present study, we show that discriminative capability can be improved by selecting the deep features of VGG16, weighting them, and hierarchically processing the features, thus employing another design principle, which is contrary to the previous research direction that increases the complexity of the architecture. The novelty of this study is hence the empirical evidence that proves that explicit weighted feature allocation along with multi-level classification can enhance the classification accuracy of a VGG16-based oral cancer framework to 98.8% classification accuracy and 98.4% feature-weight allocation accuracy under the experimental setting used in this study.

Recent research has shown the potential for deep learning to assist with automated oral cancer analysis, yet approaches to the problem have focused on various facets of the diagnostic task and often added architectural complexity. Based on CNN, a number of histopathological classification approaches have been developed, and recently, deformable convolution, adaptive feature fusion, transformer-based multi-scale representation, feature distillation, ensemble learning, spiking neural networks, handcrafted-feature integration, and multimodal learning have been explored. For instance, DCA-DAFFNet introduces deformable attention and adaptive feature fusion to the tumor grading task; Mul2UNet introduces multi-scale segmentation and channel disentanglement to the task of lymph-node segmentation; MedTrans uses cross-attention for multi-scale representation; and FD-Net uses feature distillation for the task of lymph-node segmentation. For the analysis of oral cancer other studies have used ensemble CNNs, Xception-spiking architectures, multimodal CNN pipelines, hybrid Gabor-deep learning features, LightGBM. The value of feature representation strategies that are becoming more complex is borne out by these studies, while it shows that there is also a continued trade-off between representation capability, computational complexity, feature redundancy, and classification efficiency.

Hence, there exists a research gap for the integration of feature relevance determination, feature weighting, and hierarchical classification in the context of a conventional and widely reproducible CNN backbone to classify oral cancer images. Most of the current methods focus on backbone architecture, segmentation quality, attention mechanisms, ensemble diversity or multimodal information integration. The present study, on the other hand, aims at finding out if the features of the diagnostic representation yielded from the VGG16 can be made more discriminative by explicitly selecting the relevant features, assigning weights based on the relationship of the selected features, and then, performing multi-level classification. This motivation is crucial, as the high dimensional deep features can include redundant or less informative information, which could lead to higher computation complexity and impact classification performance. To fill this gap, the proposed WFS-MLC-VGG (Weighted Feature Processing and Weighted Multi-layer Classification on a Deep Feature-based Network) scheme proposes to integrate the processing of weighted features into deep feature extraction method between deep feature extraction and classification sub-modules, instead of using more and more complicated network architecture. The purpose of the feature-selection and feature-weighting is aligned with the original research hypotheses of the study.

Therefore, the need of this study is not only to get a higher accuracy of classification but also to check the effectiveness of the oral cancer classification by better organization of already extracted deep features. The proposed framework is based on a hierarchical feature extraction using a VGG16 network, feature selection, weighted features by correlation-based feature weighting and multi-level classification. This design ensures that only the relevant and redundant features are minimized while discriminative features are retained. The resulting framework is an alternative research direction when compared to approaches that mainly enhance the depth of the architecture, add multiple branches of the network and/or include complex computation of attention and ensembles.

The key conclusion of this study is that the complexity of the underlying neural architecture is not always directly associated with the increase in the ability of the neural network to classify medical images. The findings suggest that the way features are extracted, prioritized and presented to the classifiers themselves can be an important factor towards improving the performance. The WFS-MLC-VGG framework is shown to be a high performance alternative feature-processing strategy for oral cancer image analysis, using a combination of off-the-shelf VGG16 representation learning and explicit feature selection, feature weighting and multi-level classification. This view may prove helpful in future research on medical-image classification, where the aim is to find a good balance between classification accuracy and efficiency, feature redundancy, computational load and architecture complexity.

1.1 Research Hypothesis

The research hypothesis are

H1: Integration of a Weighted feature set (WFS) and Multi-Level Classification (MLC) can be of much benefit in enhancing the accuracy of oral cancer detection as compared to traditional CNN-based models.

H2: Premeditating the use of feature selection and feature weighting mechanism before the classification can be used to improve the model learning as irrelevant and redundant features are removed and results in a better prediction performance.

H3: A deep learning algorithm composed of VGG16 and an optimized feature processing can be used to differentiate between histopathological images of malignant and normal oral tissues.

H4: The multi-level classification enhances the model to represent hierarchical features, which enhances the consistency and robustness of the model in classification.

H5: Deep learning-based automated image analysis can be of great use to detect oral cancer at its initial stages, thus minimizing the reliance on manual cancer diagnosis.

1.2 Research Contributions

The research contributions are

1. Designed a Oral cancer detection with Weighted Feature Set + Multi-Level Classification + VGG16.
2. Extracting features are used in the system, selection of relevant features, weight allocation to features and ranking the

features are performed in order to enhance efficiency in classification.

3. Obtained a maximum classified accuracy of 98.8% which was higher than the current models.
4. Dimensionality reduction and efficient feature processing achieved reduced prediction time than with traditional models.
5. The proposed model performs hidden layer-optimization, max-pooling and softmax-based classification.

2. LITERATURE SURVEY

The nuclei in histological images have intricate spatial relationships between them, and there are huge variations in shape and distribution among them. This makes laryngeal tumor grading a difficult assignment for computer-aided clinical diagnosis (CACD). Nevertheless, current tumor grading models that rely on CNNs are unable to adequately model the semantic information of the spatial location of nuclei and cannot adaptively represent nuclei with changeable morphology. Hence, to enhance the grading capability, J. Luo et al. [1] suggested a holistic network (DCA-DAFFNet) that combines DCAB, a deformable convolution guided attention block, with DAFF, a deep adaptive feature fusion algorithm. The adaptive representation of changeable nuclei is achieved by the use of a new DCAB that has deformable convolution receptive fields that are more adaptable. In addition, DCA-DAFFNet is built with strong global information representation capability by incorporating the vision transformer (ViT) block into attention-based CNN models, which solves the problem of nuclear spatial location's low semantic information modeling ability.

Low bone mineral density and microstructural degradation are symptoms of osteoporosis, a systemic skeletal condition that increases the risk of fractures in older individuals. Because dual-energy X-ray absorptiometry (DXA) is an expensive and inconvenient gold standard for diagnosis, researchers are looking for other ways to screen for cancer. Dental radiographs show promise as osteoporosis markers, particularly in patients with mouth cancer, due to the correlation between changes in mandibular trabecular bone and cortical breadth on radiographs and skeletal bone density. S. A. Bhat et al. [2] proposed a ResNet50 model and a bespoke CNN to divide cropped 80×80 ROIs from dental periapical X-rays into three groups: normal, osteopenia, and osteoporosis. Using DXA-derived labels, the author classified osteoporosis in postmenopausal Javanese women using the publically accessible Dataset of Dental Periapical Radiograph.

Using multi-channel data and adjusting to complicated tumor morphology are two of the biggest obstacles to automated tumor segmentation in OSCC histopathology study. Current deep learning approaches have two main problems: early fusion involves redundant spectrum signals, which leads to inefficient multi-channel use, and OSCC's complicated tumor-stroma interfaces are not well-aligned with rigid feature fusion. In response to these, L. Zhang et al. [3] presented Mul2UNet, a revamped UNet framework for OSCC segmentation that combines structure-aware multi-scale fusion with task-driven spectral disentanglement. Prior to maintaining important features, the framework optimizes channel selection in hybrid RGB-HSV space by minimizing redundant noise. Through the use of gradient-modulated convolutions, a deformable multi-scale architecture is able to capture the uneven tumor boundaries. Enhancement of nuclear atypia patterns is achieved through independent processing of specific channels and late fusion. Compared against sixteen reference models on two datasets,

When compared to CNNs, ViT scored better when it came to general image classification. This is why users investigating the viT for OSCC identification from histopathology images. In order to comprehend these types of medical images, data from various spatial resolutions is needed. There, Y. Xu et al. [4] provided a multiscale transformer to extract fine-grained and coarse-grained features from image patch tokens of varying sizes. Two branches make up the transformer model. One branch processes small-sized patch tokens, and the other processes large-sized patch tokens. Each branch uses its own specialized encoder to represent local and global context information from multiscale image patch tokens. The author also used multi-head cross-attention fusion with lateral connections to fuse information across scales. Data from the ablation demonstrates that the performance of MedTrans improves steadily with decreasing patch sizes.

Early regional lymph node metastases is a hallmark of OSCC. When OSCC spreads to the lymph nodes in the neck, the prognosis and survival prospects for the patient are dismal. In order to swiftly assess the cervical lymph metastatic status of OSCC patients and devise suitable treatment strategies, a feasible screening tool is essential. X. Zhang et al. [5] proposed a new diagnostic approach for detecting OSCC lymph node metastases using hyperspectral imaging in conjunction with the advantages of the commonly utilized pathological sections stained with hematoxylin-eosin (H&E). With a focus on cancer and non-cancer nuclei, the method's learning and decision-making stages progressively complete the segmentation of the lesions from coarse to fine, and achieve high accuracy. To distinguish between malignant and noncancerous nuclei, a suggested feature distillation-Net (FD-Net) network is trained during the learning phase. The post-processing of the segmentation findings allows for effective leson distinction depending on the previous during the decision-making step.

The detection and categorization of various cancers have found widespread use in medical image analysis using deep learning algorithms. M. Das et al. [6] applied the proposed framework to the analysis of oral cell histopathology images in order to programmatically detect OSCC. There are two stages to the proposed model's use of transfer and ensemble learning. The first step in detecting OSCC is to examine several CNN models using transfer learning techniques. The second phase involves building the ensemble model using the top two pre-trained CNNs from the first phase. The proposed classifier is compared with leading-edge models like Alexnet, Resnet50, Resnet101, Inception net, Xception net, and InceptionresnetV2. In order to prove that the proposed framework is effective, the results are examined. The study takes a three-stage comparative approach.

The most common reasons for OSCC include smoking, betel nut use, and chewing tobacco, which mostly impact the regions of the throat, nose, and mouth, as well as viral infections. In order to accurately treat OSCC and determine its severity, early detection is essential. The precise detection of OSCC using multiple screening and detection techniques is crucial for lowering mortality and morbidity. Here, S. Ramya et al. [7] provided Xception Spiking Fractional Neural Network (XSFN-Net), a new deep learning model for OSCC image classification using histopathological pictures. Starting with the histopathological images, the Medav filter is used to improve them. Using the improved image's custom loss function and the Parallel Reverse Attention Network (PraNet), pixel-wise image segmentation is executed. The next step is to use XSFN-Net to classify the OSCC based on the features that were recovered from the segmented output.

I. U. Haq et al. [8] investigated the revolutionary possibilities of AI in OSCC diagnosis as a means to overcome these obstacles and improve diagnostic results. In order to harness the potential of AI, three separate approaches were employed: Gabor + CatBoost, ResNet50 + CatBoost, and Gabor+ ResNet50 + CatBoost. With the use of principal component analysis (PCA), the research uses 32 Gabor Filter low-level features and 100,532 ResNet50 high-level features to reduce the risk of overfitting while keeping the top 4096 components. Prior to concatenation and picture classification, the retrieved features were subjected to individual CatBoost classification. Surprisingly, the best result was shown by the third technique, which combined ResNet50 feature extraction with Gabor filtering and CatBoost classification.

With around 10 million fatalities recorded in 2020, cancer ranks among the top causes of mortality globally. When looking at cancers worldwide, oral cancer is the sixth most common. Although screening for or treating in the pre-cancerous phases helps prevent the disease, the majority of cases are detected at later stages, which is why it is fatal. This leads to a substantial drop in mortality rates. B. Goswami et al. [9] presented a strategy for accurately identifying pre-cancerous lesions in the mouth and distinguishing them from benign ones. The process begins with a search through five separate color spaces, where color and texture features are extracted and subsequently identified using a Light Gradient Boosting Machine (LightGBM) machine learning methodology.

Early identification is key to effective treatment and better survival rates for OSCC, a critical health concern. While prior research has looked at oral lesion categorization using common smartphone images, it has mostly relied on pictures alone, ignoring the advantages of using other modalities. G. A. I. Devindi et al. [10] suggested a multimodal deep-learning pipeline that takes cues from physicians' methods for early oral cancer diagnosis and uses a wide variety of data sources, including patient metadata. In order to distinguish between benign and possibly malignant oral lesions, the research makes use of cutting-edge image encoders. The limitations of the traditional models are presented in Table 1.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Limitations of Traditional Models

Author & Year	Algorithm Used	Model Working	Dataset Used	Evaluation Metrics	Limitations
Luo et al., (2023)	DCA-DAFFNet (Attention + Feature Fusion CNN)	Uses deformable attention and adaptive feature fusion for tumor grading from histopathology images	Histopathological laryngeal tumor dataset	Accuracy, Precision, Recall, F1-score	High computational cost, limited generalization
Bhat & Shanthi, (2026)	Deep Learning (CNN-based)	CNN model extracts features from dental radiographs to detect osteoporosis	Dental periapical radiographs	Accuracy, Sensitivity, Specificity	Limited to dental images, lacks multi-modal validation
Zhang et al., (2025)	Mul2UNet (Multi-scale UNet)	Multi-scale feature extraction with deformable convolutions for segmentation	Oral cancer histopathological dataset	Dice coefficient, IoU, Accuracy	Complex architecture, high training time
Xu et al., (2024)	Vision Transformer (MedTrans)	Uses cross-attention transformer for multi-scale feature learning in diagnosis	Medical imaging datasets	Accuracy, Precision, Recall	Requires large dataset, high computational cost
Zhang et al., (2024)	FD-Net (Feature Distillation Network)	Distills deep features for improved segmentation of lymph nodes in hyperspectral images	Hyperspectral OSCC dataset	Accuracy, Dice score, IoU	Limited dataset size, hyperspectral dependency
Das et al., (2024)	Ensemble Deep Learning	Combines multiple CNN models for improved classification performance	Histopathological oral cancer dataset	Accuracy, F1-score, Recall	Increased complexity, ensemble overhead
Ramya et al., (2025)	Xception + Spiking Neural Network	Combines deep CNN with spiking neural network for efficient classification	Histopathological dataset	Accuracy, Precision, Recall	Novelty adds complexity, limited interpretability
Haq et al., (2023)	Hybrid AI Model	Combines machine learning and deep learning for feature extraction and classification	Histopathological images	Accuracy, Sensitivity, Specificity	Hybrid tuning complexity, scalability issues
Goswami et al., (2024)	LightGBM	Gradient boosting-based classification of cancer stages from white light images	White light oral images	Accuracy, Precision, Recall	Limited deep feature extraction capability
Devindi et al., (2024)	Multimodal CNN	Integrates multiple data sources for early detection using CNN pipeline	Multimodal oral cancer dataset	Accuracy, F1-score, AUC	Data integration complexity, high resource requirement

The works discussed are classified into five categories depending on the methodology used. The first methodology is based on the use of CNN architecture or modifications thereof for direct classification of the corresponding medical images. The second methodology is spatial representation through deformable convolution, attention, and adaptive feature fusion, as shown in the case of DCA-DAFFNet. The third methodology uses transformer architecture and multi-scale architecture for combining local and global context information, for example, MedTrans. The fourth methodology is features enhancement through distillation, handcrafted features, ensemble learning, or hybrid machine learning/deep learning approaches. The fifth methodology, by using multimodal learning, provides an additional source of information. It may be concluded from the research carried out in the current manuscript that every methodology has certain advantages and disadvantages.

DCA-DAFFNet, from a critical standpoint, enhances the adaptive representation via deformable attention and feature fusion; however, its complexity of architecture and high computational demands may limit generalization and deployment. Transformer-based MedTrans offers multi-scale representation of the context, but typically needs large amounts of training data and computational resources. Ensemble-based approaches enhance classification by introducing complementary CNN representations, however they make the model more complex and entail higher inference costs. A hybrid method combining gabor features, resnet features, PCA and cat boost can be used to decrease the feature dimension, but will need several processing stages and feature combination. Multimodal approaches are able to offer richer information by utilizing image and patient related data, but the integration of data leads to higher resource needs, and more problems with data availability and consistency. The observations suggest the investigation of a relatively structured feature-processing strategy where the relevance of features and the importance of features are taken into account prior to final classification.

2.1 Problem Statement

OSCC is a very common and life-threatening illness, especially in developing nations, where it is diagnosed at its late stage and therefore greatly lowers the likelihood of survival. Conventional diagnostic tests e.g. histopathological test are very time consuming, prone to errors and very reliant on the interpretation of experts thus resulting in variations and delayed diagnosis. The current machine learning and deep learning oral cancer classification methods are usually constraints in terms of poor feature selection, failure to adequately model the complex relationship between features, and poor optimal classification. Besides, most of the models cannot effectively process high-dimensional image data, which leads to complexity and low-performance. As such, automated, precise, and computationally efficient framework is urgently needed that can be used to conduct early-stage detection and classification of oral cancer. To handle these issues, this study offers a hybrid deep learning model that combines weighted feature selection, multi-level classification, and VGG16 architecture to improve the accuracy of diagnostic performance, speed of processing, and the effectiveness of medical images in predicting oral cancer.

2.2 Difference from Prior Work

The main difference between the proposed WFS-MLC-VGG framework and the reviewed approaches is the place and the application of the proposed optimization. Previous work mainly improves the performance of the diagnostic system by introducing attention modules, deformable convolutions, multi-scale processing, transformers, ensemble or multimodal approaches to the feature-extraction architecture. The proposed framework, on the other hand, adopts the VGG16 as the main deep feature extractor and provides an explicit feature-processing method after feature extraction. Features extracted are then checked for their relevance, irrelevant features are filtered by feature selection and then weights are assigned based on the relationship between features prior to the classification. In the proposed approach, feature quality and prioritization is emphasized instead of simply increasing the complexity of network architecture.

Yet other difference is related to the joint use of feature weighting and multi-level classification. In the proposed WFS-MLC-VGG framework, selected features are given different weights and passed through the hidden layers and classification process. This design is designed to make sure that highly discriminative representations are more positively included in the final decision, whereas less informative/redundant representations are less strongly included. In this study, the authors thus propose a novel approach of adding an explicit feature management stage between the deep feature extraction and classification processes for VGG16-based classification.

The experimental results also highlight the differences between the proposed framework and the comparison models. On the considered images, WFS-MLC-VGG has a feature weight allocation accuracy of 98.4% and oral cancer prediction accuracy of 98.8% while the reported comparison values for corresponding experiments are lower. Furthermore, the times for generation of features subsets and hidden layer processing in the study are still lower than comparison models for the image numbers evaluated. The findings show that feature prioritization offers a viable method of augmenting architectural simplicity over architectural complexity. However, the results used here need to be considered in the context of the experimental data and evaluation procedure used in the present study, and not as proof of clinical superiority.

3. PROPOSED MODEL

There is a strict no-touch zone around the nucleus, cytoplasm, and cytoplasmic background. There are few spots where the preservation has failed, but the important parts, such clusters of nuclei and cells, are mostly there [20]. The information that follows will only fill in the gaps, making it much simpler to selectively process a cell's nucleus and get the necessary phenotypic information [21]. These gaps must be filled in order to extract and determine parameters with any degree of accuracy. Artificial intelligence cannot be realized without the use of machine learning, of which deep learning is a subfield [22]. Global feature descriptors, which are generally used to extract picture features in standard learning algorithms, include histograms of directed gradients, local binary patterns, and color histograms. Only those with substantial expertise in the area should contribute these details. Instead of using pre-defined features, deep neural networks can automatically and hierarchically extract them from images [23]. Lower layers learn about edges and corners, intermediate layers about colors and shapes [24], and higher layers about high-level features describing the item in the oral image [25].

3.1 Dataset Description

The proposed model is applied on publicly available dataset from the link https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images and clinical images collected. The data considered in this research article consists of 700 oral images, which consist of 350 cancer (Oral Squamous Cell Carcinoma) and 350 normal oral images. These images were gathered using a mixture of both clinical and publicly available data sets that were found in an oral and maxillofacial center, and available online repositories like Kaggle. The data contains both clinical and histopathological pictures, which gives a wide range of the oral cancer conditions. Each of the images was downsampled to 224 x 224 x 3 (RGB format) to be compatible with the VGG16 based deep learning structure. Before model training, preprocessing was done by applying image denoising methods to improve the quality of the image. The processed data are then subjected to a structured pipeline which involves extraction of features, selection of features, weighting of features and classification which allows effective and precise model learning.

3.2 Methodology

The CNN is the most well-known network topology in deep learning and image processing. Because of its capacity to learn and classify input based on its hierarchical structure [26], a CNN is also known as a translation-invariant artificial neural network [27]. This research employed CNN for the detection of oral cancer cells. The suggested VGG16 model is a CNN consisting of 13 layers, with only the top three layers being completely linked. The main characteristics include a convolution kernel that is minimal in size, a pooling kernel that is similarly small, and a transconvolution that is fully coupled. Fewer parameters reduce the model's complexity, and an input image dataset of size 224 x 224 x 3 is the bare minimum required. The ability of a multilayer convolution structure to do more nonlinear transformations than a single convolution layer aids in the extraction of higher-level visual information. The proposed model framework is shown in Figure 4.

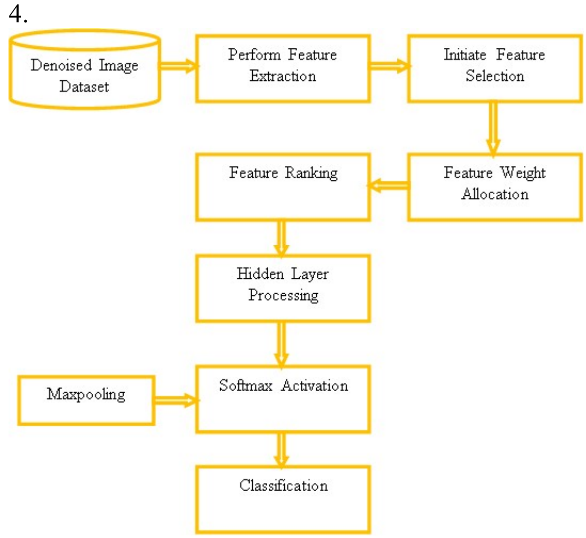

Fig 4: Proposed Model Framework

CNNs are similar in structure to classic neural networks. Training weights, bias values from research weights, or both can be applied to neurons in a neural network. A neuron's typical operation is to take in information, calculate the point product of that information, and then display behavior that is not linear in nature. Each CNN consists of a pooling layer, a subsampling layer, and one or more convolution layers. Classifying cancer cells according to type and stage typically requires a CNN. To accomplish down sampling, CNN's pooling layer is utilized. Cutting down on the number of convolution layers used to extract features speeds up the training process. There are a total of three FC levels, with ReLU activating the first two and Softmax activating the third. Images at a resolution of 224x224 pixels can be read in at the input layer of this design, which is 16 layers deep and uses 138 million parameters. The VGG16 framework is shown in Figure 5.

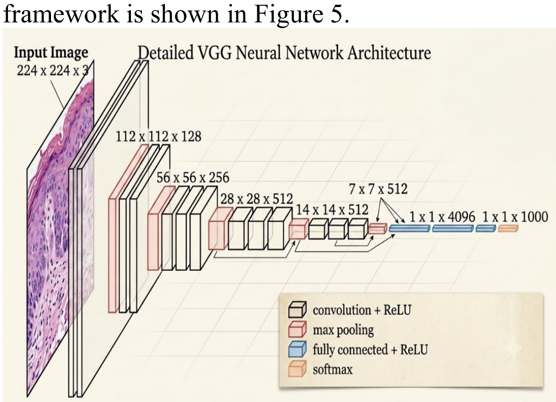

Fig 5: VGG16 Framework

The proposed model network is fed a picture with the coordinates (224, 224, 3) as its input. The 64 channels in the first two levels have the same padding and a filter size of 3 x 3. Two 128-by-3-pixel convolution layers are employed after a 2-stride max pool layer. The following layer is a max-pooling one, and its stride is 2x2. After that, there are two 3x3 sized 256-filter convolution layers. Two sets of three-layer convolution layers and a maximum pool layer follow. Each of the 512 filters has the same padding and is of the size 3x3. Two convolution layers are subsequently superimposed on this picture. This network uses 3*3 filters for feature processing. 1x1 pixels are also utilized in some layers to control the amount of input channels. After each convolution layer, the image is padded with 1 pixel to hide its spatial characteristics. This research develops a Weighted Feature Set model with Multi Level Classification model using VGG 16 model for accurate classification and detection of oral cancer.

Algorithm WFS-MLC-VGG

{

Input: Denosied Image Set {DIset}

Output: Classification Set {Cset}

Step-1: The goal of feature extraction is to extract useful information from massive datasets. In order to analyze more manageable chunks of data, dimensionality reduction is employed to break down massive input datasets into more manageable categories. The proposed image processing application benefit from feature extraction approach by considering all the image features. Features reveal an image's location in terms of storage required, classification efficiency, and, of course, time consumption, as they characterize the image's activity. The feature extraction process is performed as

$$\xi[N] = \sum_{i=1}^{N} \frac{\sum getVal(i,i+1)}{N} + \sum getrange(i,i+1)$$

$$FA[N] = \sum_{i=1}^{N} \max(getrange(DIset(i)))$$

$$FextrSet[N] = \sum_{f=1}^{M} \frac{R(f) + \mathrm{maxrange}(\beta(f,f+1))}{len(M)} + \sum_{f=1}^{N}\sum_{r=1}^{f} \lim_{f\to N}\left(\xi(DIset(f) + \frac{FA(corr(i,i+1))}{i+1}\right)^2$$

R represents model for loading denoised images, i indicates current image, β model extracts numerical features, ξ represents mean range of features and FA considers the feature attribute maximum range. Correlation among the features are represented using corr() model.

Step-2: Feature selection and variable selection are interchangeable terms. It is mostly employed for data decline through the elimination of certain attribute values and makes use of data mining data preparation procedures. Improved data visualization, faster algorithm training, and enhanced prediction performance are all outcomes of this procedure. To reduce the dimensionality of high-dimensional data, feature selection is a crucial technique. In feature selection, the characteristics that have the greatest impact on the target prediction variable or output are chosen, either automatically or by hand. The accuracy of the models is reduced and the model learns based on irrelevant features when there are redundant or irrelevant features in the obtained data. The feature selection process is performed as

$$Fselect[N] = \sum_{i=1}^{N} getVal(FextrSet(i)) + \max(FextrSet(corr(i,i+1))) + simm(getrange(i,i+1))\begin{cases} Fselect \leftarrow attr(i)\ if\ corr(i,i+1)) < Th \\ continue \quad Otherwise \end{cases}$$

Simm() model is used to check the similarity levels of the feature attributes. Th is the threshold value considered to select the features.

Step-3: Feature weighting changes the problem space by assigning varying weights to attributes on the basis that they are not all equally essential. Data mining tasks that explicitly consider the distance between instances use this. The weights are the actual values assigned to each input/feature; they indicate how significant that feature is in predicting the end output. When an algorithm is given, it determines how much weight to give each feature item in the set based on the category information. This is called feature weight. The feature weight allocation is performed as

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$W[N] = \sum_{i=1}^{N} \frac{\mathrm{maxrange}(\mathrm{Fselect}(i, i+1))}{len(Fselect)} + maxattr(i)$$

$$Fweight[N] = \sum_{i=1}^{N} getmax(W(i)) + \lim_{i \to N} \left(maxrange(corr(i, i+1)) + \frac{\min(Fselect(i))}{i} \right)^2 \begin{cases} Fweight \leftarrow W + 1 & if\ W(i) > Th \\ continue & Otherwise \end{cases}$$

Step-4: Five max-pooling layers, three fully linked layers, and thirteen convolutional layers make up VGG16. This means that there are a total of 16 layers with adjustable parameters, consisting of 13 convolutional layers and 3 fully linked layers. The VGG network uses Rectified Linear Unit (ReLU) for all of its hidden layers. The network is fed with a image with dimensions 224 x 224 x 3. With the same padding, the first two layers include 64 channels with a 3x3 filter size. Following a stride 2x2 max pool layer, two subsequent layers have 128-and 3-filter-size convolution layers, respectively. The next layer is an identical max-pooling layer with stride 2x 2. Two convolution layers with 256 filters each are then added. The filter sizes are 3x3. The hidden layer processing is done as

$$Hneurons[N] = \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{i} \max(Fweight(i,j)) + getVal(Fselect(j)) + \begin{bmatrix} Fselect_i & Fselect_{i+1} \\ Fselect_{i-1} & Fselect_i \end{bmatrix} * maxrange(Fweight(Fselect(i)))$$

$$Hprocess[N] = \prod_{i=1}^{N} \frac{Hneurons(i)}{N-i} + \sum_{i=1}^{N} \sum_{j=1}^{i} \frac{\sum maxrange\left(Hneurons(Fselect(i))\right) + maxrange(Fweight(j))}{N - (i * j)}$$

Step-5: A max pool layer follows by two sets of three convolution layers. Every one of them has 512 filters with the same padding and a size of 3x3. After that, the two convolution layers are given this image to work with. The filters utilized in these max-pooling and convolution layers are 3*3 in size. Additionally, it makes use of 1*1 pixels in a few layers to control the amount of input channels. In order to decrease the dimensionality and parameter count of the feature maps produced by each convolutional phase, a pooling layer is added after many convolutional layers. The max pooling operation is performed by setting the convolution as

$$Conv[N] = \sum_{i=1}^{N} hl^{i-1} * w^{i}$$

$$Conv_i[N] = \sum_{i=1}^{N} W_i * h_{l-i}$$

$$Maxpool[N] = \sum_{i=1}^{N} \sum_{j=1}^{i} \max(i,j) * hl^{i-1}(Hprocess(i,j))$$

Step-6: The basis model for each classifier was the VGG16 architecture with the Softmax layer removed. The category classification softmax activation layer follows the output layer. By utilizing softmax, CNNs are able to generate a probability distribution including all potential classes. The ability for the CNN to generate more precise forecasts is why this is crucial. The softmax function is applied as

$$Softmax[N] = \sum_{i=1}^{N} \frac{e^{conv(i)}}{\sum_{i=1} e^{conv(i+1)}} \mid maxpool(Fweight(i))$$

Step-7: Classification include training a model to predict the right label from a set of input data. Classification involves first training the model on the training data, then testing it on test data, and then utilizing it to make predictions on unknown data. The VGG model's capability to fit more complicated functions grows in direct proportion to the number of layers. In order to produce a classification, an image classifier runs an image through these layers. The oral cancer classification is performed as

$$Cset[N] = \sum_{i=1}^{N} \max\left(Softmax(Fweight(i, i+1))\right) + \frac{\sum_{i=1}^{N} maxpool(Hprocess(Fselect(i, i \mid 1)))}{N-i} + \max(diff(Fweight(i, i+1))) - \min(Fweight(i)) \begin{cases} Cset \leftarrow 1 & if\ diff > Th \\ Cset \leftarrow 0 & Otherwise \end{cases}$$

}

4. RESULTS

The most frequent form of oral cancer is called OSCC. Despite their significant impact on mortality, OSCCs are typically detected at an advanced stage since effective screening procedures are not yet available [28]. A better curative outcome and lower recurrence rates following surgical therapy could be achieved with early diagnosis and correct recognition of OSCCs. In this research, VGG16 model is applied on histopathology pictures of oral lesions to assess their malignancy. Cancer of the mouth and jaw, often known as oral cancer, is a common malignancy. Research on mouth malignant tumours can help clinicians treat oral cancer at an earlier stage. Using a computer-aided diagnostic system based on deep learning has been shown to be effective in a number of contexts, and it has the potential to provide a rapid and accurate diagnosis of oral malignant lesions. The proposed model is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

implemented in python and executed in Google Colab. This research develops a Weighted Feature Set model with Multi Level Classification model using VGG 16 (WFS-MLC-VGG) model for accurate classification and detection of oral cancer. The proposed model is compared with the traditional End-to-End Network With Deformable Fusion Attention and Deep Adaptive Feature Fusion for Laryngeal Tumor Grading From Histopathology Images (DCA-DAFFNet), Intelligent Computing for Medical Diagnosis Using Multiscale Cross-Attention Vision Transformer (MedTrans) and Xception Spiking Fractional Neural Network for Oral Squamous Cell Carcinoma Classification Based on Histopathological Images (XSFNN).

Finding and removing as many superfluous features as possible is an important part of feature subset selection. Because the input is now lower dimensional, learning algorithms can process it more rapidly and efficiently. In order to simplify the model's inputs, Feature Selection finds and removes features that aren't relevant. One part of developing a machine learning model is selecting features automatically according to the problem's characteristics; this process is called automatic feature selection. The Feature Subset Generation Time Levels of the proposed and existing models are shown in Table 2 and Figure 6.

Table 2: Feature Subset Generation Time Levels

Images Considered	Models Considered			
	WFS-MLC-VGG Model	IMTM-OCC Model	BDL-OCIC Model	ADCOL-OCPMD Model
100	15.4	17.9	22.9	21.5
200	15.6	18.1	23.1	21.7
300	15.8	18.3	23.3	21.9
400	16.1	18.5	23.6	22.1
500	16.3	18.8	23.8	22.4
600	16.5	19	24	22.6

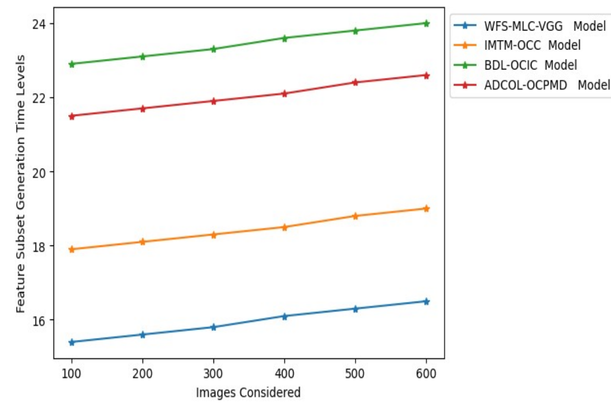

Fig 6: Feature Subset Generation Time Levels

The proposed model selects the features that are more relevant from the extracted set. Weights are allocated to the features based on correlation range and these weights helps in training the model. The Table 3 and Figure 7 show the Feature Weight Allocation Accuracy Levels of the proposed and existing models.

Table 3: Feature Weight Allocation Accuracy Levels

Images Considered	Models Considered			
	WFS-MLC-VGG Model	IMTM-OCC Model	BDL-OCIC Model	ADCOL-OCPMD Model
100	97.3	91.7	94	93.8
200	97.6	91.9	94.2	94
300	97.8	92.1	94.4	94.2
400	98	92.3	94.6	94.4
500	98.2	92.5	94.8	94.6
600	98.4	92.7	95	94.8

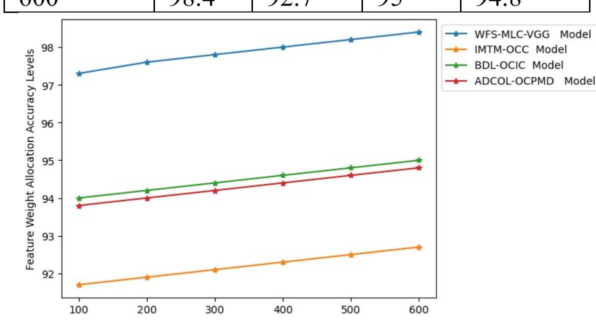

Fig 7: Feature Weight Allocation Accuracy Levels

A hidden layer is a part of an artificial neural network that processes the weighted inputs and then generates the activation function-driven output. It resides between the input and output layers. A hidden layer is interposed between the input and output of a neural network approach. This layer is responsible for routing the inputs through an activation function and assigning weights to them.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The result is the intended output. The hidden layers implement a nonlinear approach to altering the inputs to the network. The hidden layer types used by a neural network are defined by its goal, and the layers themselves might serve various functions depending on their weights. The Hidden Layer Processing Time Levels of the existing and proposed models are shown in Table 4 and Figure 8.

Table 4: Hidden Layer Processing Time Levels

Images Considered	Models Considered			
	WFS-MLC-VGG Model	IMTM-OCC Model	BDL-OCIC Model	ADCOL-OCPMD Model
100	14.7	20.7	21.4	19.6
200	14.9	20.9	21.6	19.8
300	15.1	21.2	21.8	20
400	15.3	21.4	22	20.2
500	15.5	21.6	22.2	20.4
600	15.7	21.8	22.4	20.6

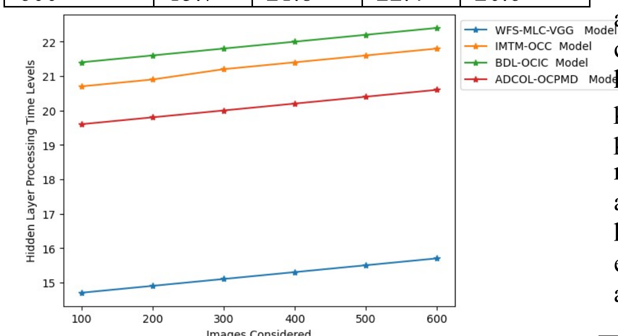

Fig 8: Hidden Layer Processing Time Levels

Each of the mathematical functions that make up a hidden layer is designed to provide a distinct output, and these layers are nested within each other. One kind of hidden layer is squash functions. If users expect a probability as a consequence of the method, these functions will come in helpful because they take an input and return a value that fits the criteria for calculating probability. The hidden layers of a neural network allow one to break down the data manipulations that make up the network's overall function. There is an optimized function in each of the hidden layers that produces a specific outcome. The Table 5 and Figure 9 shows the Hidden Layer Processing Accuracy Levels of the proposed and existing models.

Table 5: Hidden Layer Processing Accuracy Levels

Images Considered	Models Considered			
	WFS-MLC-VGG Model	IMTM-OCC Model	BDL-OCIC Model	ADCOL-OCPMD Model
100	97.8	94	93.6	92
200	98	94.2	93.8	92.4
300	98.2	94.5	94	92.6
400	98.4	94.7	94.3	92.8
500	98.6	95	94.5	93
600	98.8	95.2	94.7	93.3

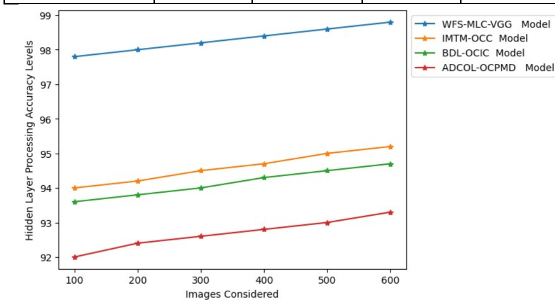

Fig 9: Hidden Layer Processing Accuracy Levels

To build a down sampled feature map, the Max Pooling operation takes the feature map's patches and determines their maximum value. After a convolution layer, this one is often applied. By having successive convolution layers examine progressively larger windows, in terms of the portion of the original input it covers, max-pooling reduces the amount of feature-map coefficients that are processed and induces the spatial-filter hierarchies. The max pooling accuracy levels of existing and proposed models are shown in Table 6 and Figure 10.

Table 6: Max Pooling Accuracy Levels

Images Considered	Models Considered			
	WFS-MLC-VGG Model	IMTM-OCC Model	BDL-OCIC Model	ADCOL-OCPMD Model
100	97.6	92.3	94.8	92.8
200	97.8	92.5	95	93
300	98	92.7	95.2	93.3
400	98.2	92.9	95.4	93.6
500	98.4	93.1	95.6	93.8
600	98.6	93.4	95.8	94

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

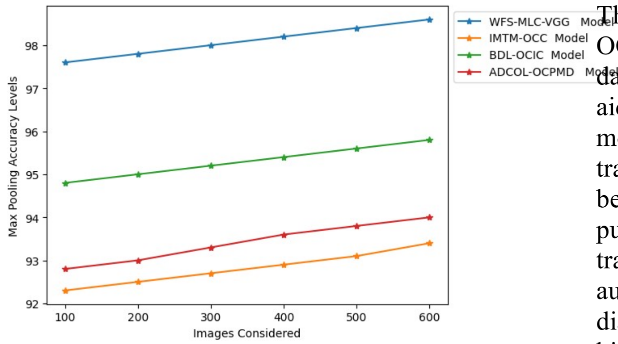

Fig 10: Max Pooling Accuracy Levels

Multiple factors have been linked to oral cancer's increased mortality rate after treatment. The use of appearance forms and stereotypes allows for the non-specific recognition of biological designs, as well as lesion-free tumour models used in medicine. Leukoplakia, erythroplasia, and oral submucosal fibrosis are all precancerous lesions seen often in the high-risk population. Differentiating among benign and malignant tumours is also essential. The proposed model accurately detects the oral cancer in very less time when compared to the traditional model. The Oral Cancer Prediction Time Levels are depicted in Table 7 and Figure 11.

Table 7: Oral Cancer Prediction Time Levels

Images Considered	Models Considered			
	WFS-MLC-VGG Model	IMTM-OCC Model	BDL-OCIC Model	ADCOL-OCPMD Model
100	17.5	20.4	19.2	24.1
200	17.7	20.6	19.4	24.4
300	17.9	20.8	19.6	24.6
400	18.1	21	19.8	24.8
500	18.4	21.2	20.1	25
600	18.6	21.4	20.3	25.2

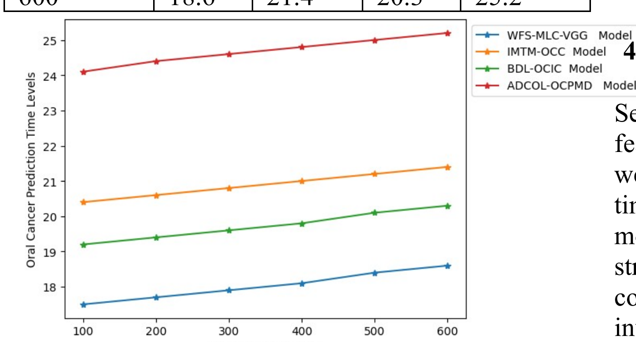

Fig 11: Oral Cancer Prediction Time Levels

The development of AI has the potential to enhance OC screening. Accurately analyzing a massive dataset from multiple imaging modalities, AI can aid oncologists in their work. This study employed mouth Squamous Cell Carcinoma biopsy images to train a VGG model, a type of CNN, to distinguish between malignant and healthy oral tissue. The purpose of this research was to use in-depth training and CNN techniques to explore novel automation approaches for making accurate diagnoses of Oral Squamous Cell Carcinoma from high-quality images. The Table 8 and Figure 12 shows the Oral Cancer Prediction Accuracy Levels of the proposed and existing models.

Table 8: Oral Cancer Prediction Accuracy Levels

Images Considered	Models Considered			
	WFS-MLC-VGG Model	IMTM-OCC Model	BDL-OCIC Model	ADCOL-OCPMD Model
100	97.7	94.3	95.2	93.6
200	97.9	94.5	95.4	93.8
300	98.1	94.7	95.6	94
400	98.4	94.9	95.8	94.2
500	98.6	95.1	96	94.5
600	98.8	95.4	96.2	94.7

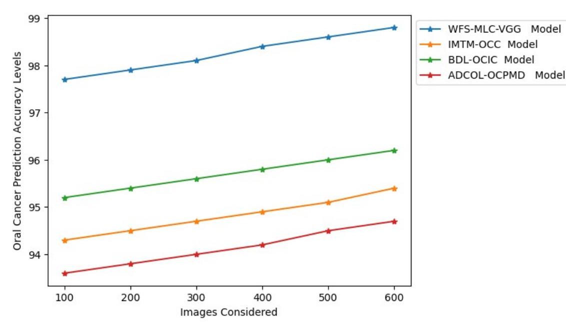

Fig 12: Oral Cancer Prediction Accuracy Levels

4.1 Time Complexity

Several sequential steps, such as convolutional features extraction, features selection, features weighting and classification, mainly determine the time complexity of the proposed WFS-MLC-VGG model. The convolutional steps in the VGG16 structure are also a contributing factor to the computational cost in that they are required to involve multiple layers of processing high-dimensional image data. Moreover, the feature selection process involves pair-wise comparison/correlation to the features, which adds to the computation requirement further. Based on these aspects, the time complexity of the whole model may be represented as

$$O(N*L*F + N*F^2)$$

with N being the number of input images, L being the network layers, and F being the number of features extracted. The feature selection aspect prevails in the case of large feature space, and the convolutional operations predominate in the course of deep feature extraction.

4.2 Computational Complexity

The computational complexities of the proposed model are highly controlled by the application of VGG16 architecture which comprise of around 138 million parameters. This translates into huge memory and processing specifications especially during the training stage. The space complexity of the model can be stated as $O(P + N*F)$ with P being the number of model parameters and N*F being the amount of storage used to store the feature representations. Though the dimensionality reduction and weighting mechanism in its inclusion can aid in efficiency and minimize dimensionality, the fact that the convolutional structure is deep makes the total computational cost very expensive. Therefore, this model needs to be run on GPUs in order to be effectively processed and might not be available to be deployed into resource-constrained or real-time systems without additional optimization.

4.3 Limitations of Proposed Model

Although the proposed model has high classification accuracy, there are a number of limitations that it has to overcome. The size of the dataset is not that large since it is only 700 images, which can reduce the generalization of the model and raise the likelihood of overfitting. The model is yet to be tested in clinical settings, which limits the use of the model in clinical practice. The use of the VGG16 architecture implies that the model has high computational and memory demands, which is why it is not as appropriate when it comes to the deployment in low-resource systems. The feature selection and weighting method could be biased when not optimized. The model lacks resiliency analysis to noisy or adversarial input, and the scalability of the model has not been verified on large-scale data, which is the subject of future research and validation.

4.4 Problems and Open Research Issues

Although WFS-MLC-VGG was employed to achieve promising results, there are still some problems left for further investigation. The first is related to the current experimental data, which consists of 700 images, 350 of which are cancer and 350 healthy images. The current study is thus a proof of the effectiveness of the model on the analyzed data set in absence of population-level generalization.

Second, the issue of external and clinical validation is yet to be resolved. The proposed framework has not been tested, in a prospective clinical workflow, or validated with multiple institution data. External validation with independent databases from various institutions, scanners, laboratories, patient cohorts and imaging acquisition settings should be explored in future research.

Third, computational efficiency is still an open question. The primary architecture used, VGG16, has about 138 million parameters and thus has a high memory and processing requirement, but it can be reduced by feature selection and weighting of the processed representations. This analysis shows that the current implementation is suitable for GPU execution and that the implementation can be made more efficient for deployment to resource-constrained or real-time applications. Lightweight CNN backbones, model-compression strategies such as knowledge distillation and quantization, parameter-efficient transfer learning, and pruning should be studied in future research.

Fourth, the feature-weighting mechanism is also robust, and its robustness should be investigated further. The present study does not systematically assess the noise sensitivity of feature selection and weighting, variations in image quality, adversarial perturbations, or variations in staining and acquisition conditions. High performance of the classification on a limited dataset does not necessarily imply good performance under real-world variation, which is important for medical image analysis.

Lastly, there are opportunities for research on interpretability and explanations for clinical situations. The proposed model identifies and weights discriminative features; the current study does not offer a detailed visualization of the anatomical/histopathological regions involved in individual predictions. In future, explainable-AI methods, visualization of attention, saliency analysis and clinician-centered assessment should be explored to establish if the model's decisions align with medically relevant visual features.

The findings of the experiments show that WFS-MLC-VGG has a good performance in terms of classification in the experimental set-up provided. In this model, the prediction accuracy for oral cancer is achieved as 98.8% at 600 considered images and 98.4% for feature-weight allocation is achieved at 600 considered images. The results indicate that the assumption that explicit feature selection and explicit feature weighting can improve the processing of deep representations for classification is supported.

However, the results must also be critically analysed. The reported accuracy is based on a relatively small number of images (700) and a current study does not have an independent external validation or prospective clinical evaluation. The model, in other words, is not ready for clinical use, but for experimental use. Moreover, VGG16 has around 138 million parameters, making the practical benefit of the proposed framework non-viable in high resource constrained environments. The feature weighting strategy could also depend on the nature of the training data and its stability has not been proven when facing shifts in the data sets. As such, one of the main results of this research must be interpreted as that weighted feature allocation and multi-level classification is a feature processing approach that can be used for oral cancer image classification and has the potential to outperform all the existing deep-learning architectures, not that it is the best of them all.

The significance of this study is that it shows that better oral cancer image classification is not always achieved by building up the architectural complexity of the feature extractor. In contrast, a feature-centric strategy is implemented in the proposed WFS-MLC-VGG framework, where deep features extracted from VGG16 are carefully chosen, weighted and then classified at multiple levels. It is therefore a design principle for deep-learning-based oral cancer classification that the study adds: feature allocation and hierarchical classification can be used to enhance the effectiveness of classification based on the CNN features it extracted.

5. CONCLUSION

This research aimed at exploring the possibility of improving the performance of oral cancer image classification systems using explicit feature prioritisation and multi-level classification, based on the VGG16 model. The inspiration was based on the fact that existing deep learning methods have shown many times that they have enhanced the representation of diagnoses by adopting increasingly complex architectures, attention mechanisms, transformers, ensembles, segmentation networks and multimodal processing. Instead, the proposed WFS-MLC-VGG framework adopts feature-centric approach where deep features extracted by VGG16 are passed through the feature selection, feature weighting and hierarchical classification layers prior to making the final prediction.

These experimental results illustrate the effectiveness of this design for the experimental conditions considered. The proposed framework's feature-weight allocation accuracy and oral cancer prediction accuracy was 98.4% and 98.8% respectively at 600 considered images. The reported featureset generation and hidden layer processing results also show reduced processing time as compared to the models compared in the study. The results seem to validate the research hypothesis of explicitly determining and prioritizing deep features that are discriminating in the classification of oral cancer using CNN.

The results suggest that feature selection and weighted feature allocation can be combined with hierarchical CNN representation and multi-level classification, thus providing an alternative to the methods relying more on the expansion of the architecture. This adds to the existing knowledge base to illustrate how the organization and prioritization of extracted representations can be an important design factor when developing an automated oral cancer image analysis.

The results should not be interpreted as clinical deployment ready, though, at the same time. It is based on a relatively limited number of images (700) and the model has not yet been independently validated across multiple institutions or validated in a prospective clinical setting. Moreover, the approximately 138-million-parameter VGG16 is relatively heavy for computations.

Future studies should therefore be aimed at the large-scale and multi-institutional validation, robustness to image quality and acquisition fluctuations, explainable prediction mechanisms, lightweight model architectures, model compression, and assessment on clinically representative datasets. These investigations will help to clarify if the proposed feature-weighting principle is able to persist beyond the current experimental setup and if eventually, these principles can help to make computer-assisted oral cancer screening reliable.

All these together suggest that the most current challenge is not just whether deep learning can be used to classify oral cancer images, but how deep learning systems can perform well and reliably in a clinically generalizable manner, while being interpretable and computationally efficient. In the current work, we focus on one aspect of this overarching problem and show the benefits of feature selection and weighted feature allocation in a classification system utilizing a VGG16 model. The other challenges in dataset diversity, external validation, computational scalability, robustness, and explainability set the next agenda for research before these systems can be adopted for reliable clinical decision support.

Declarations

Ethical approval

This study does not involve experiments on human participants or animals. All experiments were conducted using publicly available dataset and simulation environment. Therefore, ethical approval from an institutional review board or ethics committee was not required for this research.

Consent to participate

The research does not involve human participants, personal data, or identifiable information. Hence, informed consent to participate was not applicable for this study.

Consent to publish

The research does not contain any individual person's data in any form. All authors have reviewed the manuscript and consent to its publication.

Ethics approval

Ethics approval was not required for this study as it did not involve human or animal subjects.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Clinical trial number:

Not applicable

Data availability

The datasets used and/or analyzed during the current study are publicly available at the link link https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images which is discussed in the manuscript.

Authors contribution

The introduction, proposed model and results are generated and written by P.Sandhya Krishna and supervised and verified by Peram Subba Rao and P.Sandhya Krishna concluded the paper.

Acknowledgement

The authors would like to thank their institution and peers for their support and valuable support in writing this paper.

Clinical trial number: Not applicable

Human Participants and/or Animals

This study did not involve human participants or animals.

REFERENCES

[1]. J. Luo et al., "DCA-DAFFNet: An End-to-End Network With Deformable Fusion Attention and Deep Adaptive Feature Fusion for Laryngeal Tumor Grading From Histopathology Images," in IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-15, 2023, Art no. 5031115, doi: 10.1109/TIM.2023.3328088.

[2]. S. A. Bhat and P. B. Shanthi, "Osteoporosis Detection From Dental Periapical Radiographs Using Deep Learning Approaches," in IEEE Access, vol. 14, pp. 16774-16791, 2026, doi: 10.1109/ACCESS.2026.3658368.

[3]. L. Zhang, W. Xiong, D. Wu, G. Liu, K. Guo and X. Zhou, "Mul2UNet: Deformable Multi-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Scale UNet With Decoupled Channel Fusion for Oral Squamous Cell Carcinoma Segmentation," in IEEE Transactions on Consumer Electronics, vol. 71, no. 3, pp. 8477-8491, Aug. 2025, doi: 10.1109/TCE.2025.3582774.

[4]. Y. Xu, Y. Hong, X. Li and M. Hu, "MedTrans: Intelligent Computing for Medical Diagnosis Using Multiscale Cross-Attention Vision Transformer," in IEEE Access, vol. 12, pp. 146575-146586, 2024, doi: 10.1109/ACCESS.2024.3450121.

[5]. X. Zhang et al., "FD-Net: Feature Distillation Network for Oral Squamous Cell Carcinoma Lymph Node Segmentation in Hyperspectral Imagery," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1552-1563, March 2024, doi: 10.1109/JBHI.2024.3350245.

[6]. M. Das, R. Dash, S. Kumar Mishra and A. Kumar Dalai, "An Ensemble Deep Learning Model for Oral Squamous Cell Carcinoma Detection Using Histopathological Image Analysis," in IEEE Access, vol. 12, pp. 127185-127197, 2024, doi: 10.1109/ACCESS.2024.3450444.

[7]. S. Ramya, R. I. Minu and K. T. Magesh, "Xception Spiking Fractional Neural Network for Oral Squamous Cell Carcinoma Classification Based on Histopathological Images," in IEEE Access, vol. 13, pp. 98574-98591, 2025, doi: 10.1109/ACCESS.2025.3574327.

[8]. I. U. Haq, M. Ahmed, M. Assam, Y. Y. Ghadi and A. Algarni, "Unveiling the Future of Oral Squamous Cell Carcinoma Diagnosis: An Innovative Hybrid AI Approach for Accurate Histopathological Image Analysis," in IEEE Access, vol. 11, pp. 118281-118290, 2023, doi: 10.1109/ACCESS.2023.3326152.

[9]. B. Goswami, M. K. Bhuyan, S. Alfarhood and M. Safran, "Classification of Oral Cancer Into Pre-Cancerous Stages From White Light Images Using LightGBM Algorithm," in IEEE Access, vol. 12, pp. 31626-31639, 2024, doi: 10.1109/ACCESS.2024.3370157.

[10]. G. A. I. Devindi et al., "Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted Early Detection of Oral Cancer," in IEEE Access, vol. 12, pp. 124375-124390, 2024, doi: 10.1109/ACCESS.2024.3454338.

[11]. Yadav, Dhirendra & Sharma, Bhisham & Noonia, Ajit & Mehbodniya, Abolfazl. (2025). Explainable label guided lightweight network with axial transformer encoder for early detection of oral cancer. Scientific Reports. 15. 10.1038/s41598-025-87627-y.

[12]. B. Song, S. Sunny, S. Li, K. Gurushanth, P. Mendonca et al., “Bayesian deep learning for reliable oral cancer image classification,” Biomedical Optics Express, vol. 12, no. 10, p. 6422, 2021.

[13]. G. Tanriver, M. S. Tekkesin and O. Ergen, “Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders,” Cancers, vol. 13, no. 11, pp. 2766, 2021.

[14]. S. Camalan, H. Mahmood, H. Binol, A. L. D. Araújo, A. R. S. Silva et al., “Convolutional neural network-based clinical predictors of oral dysplasia: Class activation map analysis of deep learning results,” Cancers, vol. 13, no. 6, p. 1291, 2021.

[15]. J. H. Lim, C. S. Tan, C. S. Chan, R. A. Welikala, P. Remagnino et al., “D’OraCa: deep learning-based classification of oral lesions with mouth landmark guidance for early detection of oral cancer,” in Annual Conf. on Medical Image Understanding and Analysis, MIUA 2021: Medical Image Understanding and Analysis, Lecture Notes in Computer Science Book Series, Cham, Springer, vol. 12722, pp. 408–422, 2021.

[16]. H. Lin, H. Chen, L. Weng, J. Shao and J. Lin, “Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis,” Journal of Biomedical Optics, vol. 26, no. 08, pp. 086007–086017, 2021.

[17]. R. A. Welikala et al., "Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer," in IEEE Access, vol. 8, pp. 132677-132693, 2020, doi: 10.1109/ACCESS.2020.3010180.

[18]. C. Huang, G. Zhang, S. Chen and V. H. C. de Albuquerque, "An Intelligent Multisampling Tensor Model for Oral Cancer Classification," in IEEE Transactions on Industrial Informatics, vol. 18, no. 11, pp. 7853-7861, Nov. 2022, doi: 10.1109/TII.2022.3149939.

[19]. N. Sharma et al., "Multifractal Texture Analysis of Salivary Fern Pattern for Oral Pre-Cancers and Cancer Assessment," in IEEE Sensors Journal, vol. 21, no. 7, pp. 9333-9340, 1 April1, 2021, doi: 10.1109/JSEN.2021.3053262.

[20]. M. Z. M. Shamim et al., "Automated Detection of Oral Pre-Cancerous Tongue Lesions Using Deep Learning for Early Diagnosis of Oral Cavity Cancer," in The Computer Journal, vol.

65, no. 1, pp. 91-104, Jan. 2020, doi: 10.1093/comjnl/bxaa136.
[21]. K. Caughlin et al., "Aligning Small Datasets Using Domain Adversarial Learning: Applications in Automated in Vivo Oral Cancer Diagnosis," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 1, pp. 457-468, Jan. 2023, doi: 10.1109/JBHI.2022.3217015.
[22]. H. Yan et al., "Diverse Region-Based CNN for Tongue Squamous Cell Carcinoma Classification With Raman Spectroscopy," in IEEE Access, vol. 8, pp. 127313-127328, 2020, doi: 10.1109/ACCESS.2020.3006567.
[23]. J. Rimal, A. Shrestha, I. K. Maharjan, S. Shrestha and P. Shah, "Risk assessment of smokeless tobacco among oral precancer and cancer patients in eastern developmental region of nepal", *Asian Pacific J. Cancer Prevention*, vol. 20, no. 2, pp. 411-415, Feb. 2019.
[24]. H. Amarasinghe, R. D. Jayasinghe, D. Dharmagunawardene, M. Attygalla, P. A. Scuffham, N. Johnson, et al., "Economic burden of managing oral cancer patients in sri lanka: A cross-sectional hospital -based costing study", *BMJ Open*, vol. 9, no. 7, Jul. 2019.
[25]. R. K. Gupta, M. Kaur and J. Manhas, "Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium", J. Multimedia Inf. Syst., vol. 6, no. 2, pp. 81-86, Jun. 2019.
[26]. P. R. Jeyaraj and E. R. Samuel Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm", J. Cancer Res. Clin. Oncol., vol. 145, no. 4, pp. 829-837, Apr. 2019.
[27]. S. Xu, Y. Liu, W. Hu, C. Zhang, C. Liu, Y. Zong, et al., "An early diagnosis of oral cancer based on three-dimensional convolutional neural networks", IEEE Access, vol. 7, pp. 158603-158611, 2019.
[28]. N. Haron, R. B. Zain and A. Ramanathan, "m-Health for early detection of oral cancer in low-and middle-income countries", *Telemed. e-Health*, vol. 26, no. 3, pp. 278-285, 2019.
[29]. M.-J. Lian et al., "Construction of an oral cancer auto classfy system based on machine learning for artificial intelligence", Proc. Comput. Sci. Inf. Technol. Comput. Sci. Conf. Proc., pp. 33-39, 2019.
[30]. N. Chamidah, K. H. Gusti, E. Tjahjono and B. Lestari, "Improving of classification accuracy of cyst and tumor using local polynomial estimator", Telkomnika, vol. 17, no. 3, pp. 1492-1500, 2019.
[31]. C.-H. Chan, T.-T. Huang, C.-Y. Chen, C.-C. Lee, M.-Y. Chan and P.-C. Chung, "Texture-map-based branch-collaborative network for oral cancer detection", IEEE Trans. Biomed. Circuits Syst., vol. 13, no. 4, pp. 766-780, Aug. 2019.
[32]. S. Panigrahi and T. Swarnkar, "Automated classification of oral cancer histopathology images using convolutional neural network", Proc. IEEE Int. Conf. Bioinf. Biomed., pp. 1232-1234, 2019.
[33]. R. Selvan and E. B. Dam, "Tensor networks for medical image classification", Med. Imag. Deep Learn., pp. 721-732, 2020.
[34]. Z.-Z. Sun, C. Peng, D. Liu, S.-J. Ran and G. Su, "Generative tensor network classification model for supervised machine learning", Phys. Rev. B, vol. 101, no. 7, 2020.
[35]. N. R. Rao, A. Villa, C. B. More, R. D. Jayasinghe, A. R. Kerr and N. W. Johnson, "Oral submucous fibrosis: A contemporary narrative review with a proposed inter-professional approach for an early diagnosis and clinical management", J. Otolaryngol. Head Neck Surg., vol. 49, no. 1, pp. 3, Dec. 2020.
[36]. A.C. Pinto, J. Caramês, H. Francisco, A. Chen, A. M. Azul and D. Marques, "Malignant transformation rate of oral leukoplakia—Systematic review", Oral Surg. Oral Med. Oral Pathol. Oral Radiol., vol. 129, no. 6, pp. 600-611, Jun. 2020.
[37]. Y. Panagakis et al., "Tensor methods in computer vision and deep learning", Proc. IEEE, vol. 109, no. 5, pp. 863-890, 2021.
[38]. R. O. Alabi et al., "Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer", Int. J. Med. Informat., vol. 136, 2020.