DOI: 10.15649/2346030X.3784
Aibi research, management and engineering journal. Volume 12, Number 3, Pag 46-62, September - December 2024, ISSN 2346-030X

SGDWOA: A novel approach in whale optimisation for accurate cell classification in oral squamous cell carcinoma using machine learning.

Anuradha Suresh Pandit[1], Vaibhav V. Dixit[2]
[1]*G. H. Raisoni College of Engineering and Management Ahmednagar, Maharashtra - India*
[2]*G H Raisoni College Of Engineering and Management, Pune - India*
ORCID: [1]*0009-0009-7655-2112*, [2]*0000-0003-4015-2941*

Received: February 08, 2024.
Accepted: June 14, 2024.
Publicado: September 01, 2024.

***Abstract*—** Oral squamous cell carcinoma (OSCC), a common head and neck cancer, is often unnoticed but can be identified early. Diagnosing this heterogeneous tumour requires extensive human experience, and artificial intelligence can help to improve diagnosis. This study used novel methodologies based on feature selection and classification in an attempt to obtain good findings for the early detection of OSCC. By using cutting-edge hybrid strategies to extract features and improve classification, this work seeks to bridge the gap among deep learning and machine learning procedures. Initially, preprocessing is done to address artifacts in the OSCC dataset. The first method uses SMOTE oversampling and feature scaling in conjunction with Resnet 50 and Efficientnet B5 models for feature extraction. In the second method, the best feature set is chosen using the Statistic gain Dynamic Remodelled Whale Optimization Algorithm (SDRWOA), and the Random Forest Classifier is then employed to classify cancer types into poor, moderate, and well categories. The finding shows that the proposed model beats the other classifiers by attaining the maximum overall accuracy, recall and F1-score of 98% and precision of 97.6%. In conclusion, the suggested approach advances the development of extremely precise and effective OSCC diagnosis techniques.

Keywords: oral cancer, oral squamous cell carcinoma, histopathological images, AI-based system, machine learning, data preprocessing, resnet 50, random forest, SMOTE, feature scaling, efficient net B5.

*Corresponding author.
Email: panditanuradha29@gmail.com (Anuradha Suresh Pandit).

How to cite this article: A. Suresh-Pandit y Vaibhav V. Dixit, “SGDWOA: A novel approach in whale optimisation for accurate cell classification in oral squamous cell carcinoma using machine learning”, *Aibi research, management and engineering journal*, vol. 12, no. 3, pp. 46-62 2024, doi: 10.15649/2346030X.3784

I. INTRODUCTION

OSCC is an ulcer proliferative lesion of the oral mucosa that can impact anywhere from the lips to the oropharynx. The tongue, mouth floor, and mandible's gingivobuccal sulcus are common locations [1]. Chewing tobacco and pan masala, a concoction of tobacco, betel nut, and other substances, are the main causes of cancer in India. Because of the tobacco-lime combo, gingivobuccal sulcus becomes the site of classical Indian cancer. The easy availability of pan masala in pre-mixed sachets has contributed to its rising popularity. Over time, fibrosis and sub mucous fibrosis may result from the constant action of chewing or storing pan masala in the oral cavity [2]. The most common pathological study utilized for screening and diagnosis of oral cancer is biopsy. Under a microscope, biopsy slides are visually assessed using a variety of cytological criteria. Thus, maintaining consistency and guaranteeing reproducibility in outcomes present numerous difficulties [3]. This approach, however, is time-consuming, error-prone, labour-intensive, and demands a high level of knowledge. As a result, prompt diagnosis of OSCC is crucial for appropriate treatment, raising survival rates and lowering death rates. The early-stage (initial) survival rate is 80%, whereas the five-year survival rate in OSCC stage IV is 20:30% [4]. Further automated cancer diagnostic research focuses on quantitative techniques. Pathologists can improve inter-observer reliability and diagnostic precision while reducing effort with the help of advances in digital imaging and computational diagnostics [5]. The entire anatomical structure of the oral cavity is seen in Fig. 1.

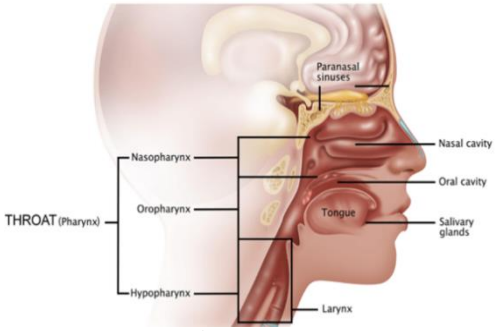

Figure 1: Anatomy of oral cancer sites.
Source: Own elaboration.

There has been study on computer-aided histopathology for a variety of cancer diagnostic and grading applications [6, 7, 8]. Subsequently, a plethora of AI-driven machine learning and deep learning models have been created, capable of accurately forecasting the kind of cancer by extracting and recognizing patterns and relationships from datasets [9, 10, 11, 12]. The effective application of AI algorithms for the diagnosis, staging, and prognosis of all cancer types has been documented in numerous research. Sharma and Mehra suggested a deep learning based algorithm for automatic multiclass categorization of breast cancer based on histopathology pictures. The convolutional neural network (CNN) architecture VGG16, which has 16 weighted layers, combined with the support-vector machine (SVM) classifier produced the highest micro- and macro-average among the various classifier combinations [13].

To train more complicated architectures, recent improvements in deep learning methodology can be summarized as learning methodologies, initializations, and activation functions. Optimizing a neural network with a deep architecture is still challenging, despite the significant progress made in decreasing the effect of exploding/vanishing gradients via activation functions and batch normalization [14]. Data mining techniques have been widely employed by researchers to detect and categorize oral cancer [15]. Lalithamani et al. presented a deep neural-based adaptive fuzzy system that utilized machine learning to identify and diagnose oral cancer with a high classification accuracy [16]. Tabibu et al. demonstrated how deep learning processes may be utilized for both survival outcome prediction and pan-renal cell carcinoma classification by analyzing histopathology images. The highest performing approach, which produced a 0.93 slide-wise AUC, was to combine deep CNN with directed acyclic graph (DAG) and SVM after the multiclass classification job was divided into individual binary tasks [17].

Deep learning approaches have proven to be capable of effectively diagnosing biomedical images, despite their time-consuming nature. Deep learning techniques that work well for this purpose include models based on convolutional neural networks (CNNs) [18]. By combining the features of each new image (testing data) against the features in the recorded image (training data), the CNN models are first trained to identify the traits of each illness. After training, the CNN approach will be able to predict unknown situations [19]. However, the accuracy of CNN models is influenced by a variety of issues, including noise in data set images, a shortage of data sets, unbalanced data sets, the number of layers utilized, activation function, and others [20]. In order to improve histological diagnosis outcomes, which are needed for the initial diagnosis of OSCC, this study set out to explore the difficulties that CNN models face and attempted to resolve them. The goal of this study was accomplished by employing hybrid deep learning and machine learning methods to improve the data set images to reduce noise, solve time-consuming issues, and eliminate the need for expensive systems. Additionally, deep features across deep learning models were used to diagnose OSCC, and a novel optimization method was employed to select the best features from images. The contribution of the proposed study is as follows:

- The study aims to create hybrid systems that combine machine learning and deep learning models, extracting features through deep learning models and selecting optimal features using optimization algorithms, and classifying cancer types using machine learning models.
- This study focuses on categorizing histological images for early diagnosis of OSCC using two approaches.
- The first involves feature extraction using Resnet 50 and Efficientnet B5 models, followed by SMOTE oversampling and normalizer feature scaling.
- The second approach selects the optimal feature set of OSCC using the Statistic gain Dynamic Remodelled Whale Optimization Algorithm (SDRWOA) metaheuristic based optimization algorithm.
- The cancer type is classified into poor, moderate, and well using the Random Forest Classifier.
- This hybridization will contribute to the development of accurate and efficient methods for OSCC diagnosis.
- The further section of the article is arranged as follows: In Section 2, a collection of earlier research is provided. Section 3 presents an examination of the tools and methods utilized for the study and interpretation of the histological pictures of OSCC diagnostic

approaches. Section 4 presents the findings from the evaluation of the suggested techniques, as well as a comparison and analysis of the procedure employed in this study. In Section 5, the study's conclusion is provided.

II. RELATED WORKS

Rahman, T. Y., et al., [21] examined and proposed a framework for the semi-automated diagnosis and categorization of oral cancer using microscopic biopsy pictures of OSCC utilizing morphological and textural parameters that are both clinically and physiologically meaningful. For the investigation, forty biopsy slides were used, and 452 hand-cropped cell nuclei in total were taken into consideration for additional analysis and the extraction of morphological and textural features. A combination strategy was suggested following a comparative analysis of widely utilized segmentation techniques. The best nuclei segmentation was achieved by this methods. The collected characteristics were then fed into five different classifiers: decision tree, logistic regression, SVM, linear discriminant, and k-nearest neighbours (KNN). Training time analysis was also performed on classifiers. An extensive indigenous cell level dataset of OSCC biopsy pictures is another addition made by the research.

Rahman, T. Y., et al., [22] motivated on the use of artificial intelligence in the diagnosis of oral cancer. Traditional biopsy techniques continue to be the gold standard for precise detection despite a great deal of research. The study accuracy histopathology pictures from local hospitals to extract shape, texture, and color properties. A cell-level dataset including 720 nuclei was generated by segmenting and generating a dataset of 42 complete slide slices. For classification, a variety of classifiers were employed, including the Decision Tree Classifier, SVM Logistic Regression, and Linear Discriminant. According to the findings, SVM and linear discriminant classifiers performed the best when it came to texture and color features, which qualified them for use in software as a diagnostic tool.

Fati, S. M., et al., [23] used hybrid approaches based on fused features in an attempt to obtain adequate findings for early OSCC diagnosis. The first approach achieved better outcomes in diagnosing the OSCC data set by utilizing CNN models (AlexNet and ResNet-18) along with the SVM methodology. In the second approach, CNN models were coupled with extracted color, texture, and shape characteristics through the use of methods such as grey-level co-occurrence matrix (GLCM), fuzzy color histogram (FCH), discrete wavelet transform (DWT), and local binary pattern (LBP). After the dimensionality of the data set features was reduced using the principal component analysis (PCA) algorithm, the ANN network based on hybrid features produced the better outcomes in terms of accuracy, specificity, sensitivity, precision, and AUC. However, there is a need to address the drawback of limited data availability, narrow focus of single-model feature extraction and the complexity of cell classification.

Chan, C. H., et al., [24] presented the automatic detection of malignant patches and ROI marking in a single model by means of a deep convolutional neural network (DCNN) and texture map. The model consists of two cooperating branches: a bottom branch for semantic segmentation and ROI marking, and an upper branch for the detection of oral cancer. Whereas the lower branch improves accuracy, the top branch removes malignant areas. A sliding window is used to calculate the standard deviation values once the texture map is retrieved from the input image. The technique is known as the branch-collaborative network based on texture maps. The wavelet transform and Gabor filter-based detection's average sensitivity and specificity are validated by the experimental outcomes.

Subhija, E. N., et al., [25] created a novel algorithm for patch selection to enhance the diagnosis of cancer by obtaining information particular to individual nuclei from picture patches. The technique extracts morphological features from five pre-trained networks and extracts texture from Haar wavelet decomposed components. An additional trees classifier is then used to choose the final feature vector. Six classifiers voting classifier, logistic regression, random forest, Naive Bayes, KNN, and SVM have been used to test the technique. To identify OSCC, the patch-based method performs superior than the image-based technique, indicating its accuracy and efficiency. Oral pathologists will find this procedure to be an accurate and dependable support tool.

III. PROPOSED METHODOLOGY

OSCC represents one of the most prevalent histological neoplasms among head and neck cancer kinds, accounting for the seventh-highest incidence of cancer, which is frequently not seen even though it is confined in an area that is visible and can be identified very early. However, because of tumour heterogeneity, such a diagnosis is laborious and requires a high level of human experience. Artificial intelligence methods thus aid professionals and medical professionals in diagnosing patients accurately. The materials and techniques utilized in this work to categorize histological images for an early diagnosis of OSCC are presented in this sector, as presented in Fig 2. Initially, all histological images were pre-processed because the OSCC data set images included artifacts. The goal of this study was achieved by the use of two methodologies, each using two systems. The initial process was depend on a feature extraction using both Resnet 50 model and Efficient net B5 model followed by SMOTE oversampling and normalizer feature scaling method, while the second approach was to select the optimal feature set of OSCC by the metaheuristic based optimization algorithm of Statistic gain Dynamic Remodelled Whale Optimization Algorithm (SDRWOA) and finally the cancer type is classified into three classes as poor, moderate and well using the Random Forest Classifier. Therefore, the goal of the present research was to create novel systems that combine machine learning and deep learning models, as well as hybrid techniques for using deep learning models to extract features and select the optimal features using optimization algorithm followed by classification using machine learning model. This hybridization will contribute to the development of extremely accurate and efficient methods for OSCC diagnosis.

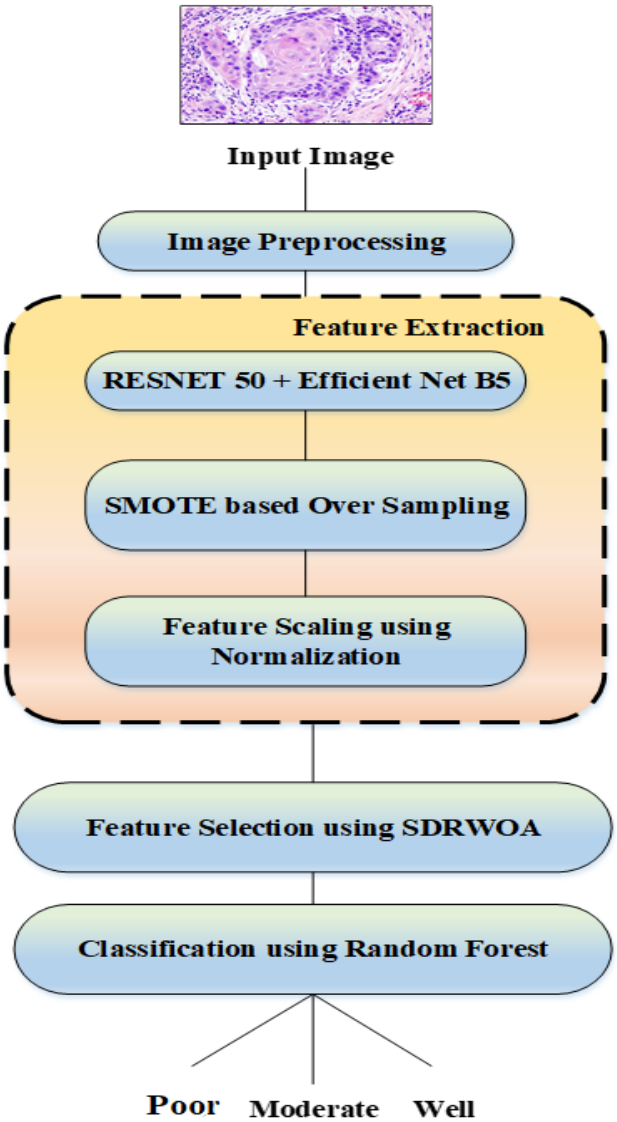

Figure 2: The Proposed Methodology for Histological image diagnosis of oral cancer.
Source: Own elaboration.

a. Dataset Description

The suggested systems in this paper were assessed using a Histological image from the publicly available OSCC data set. The dataset contains 5192 histopathology images collected from biopsy slides at a magnification of 100. Every image in the data set was attained by biopsy under local anaesthesia. A pathologist diagnosed the biopsies, and images were obtained using a technique that allowed for up to 100× magnification under a microscope. 2494 normal histological images, comprising 48% of the total images, and 2698 malignant histological images of OSCC, or 52% of the total images, made up the data set. A pathologist's review of the data set's normal images revealed that the tissue was not malignant. This study focused on histological images, which showed the adipose tissue, connective tissue, and squamous epithelium layer. Fig. 3 describes a set of data set samples of both classes.

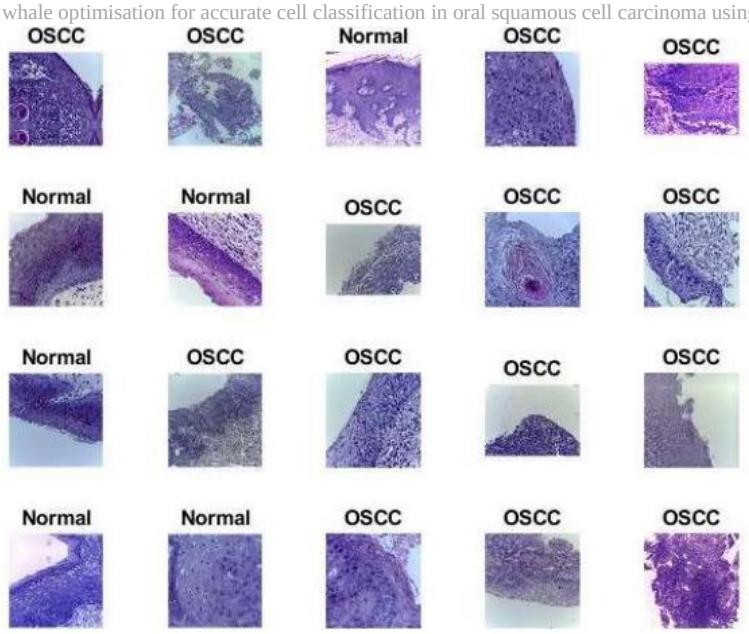

Figure 3: Samples of histological images in the dataset.
Source: Own elaboration.

b. Preprocessing of Histological Images

Preprocessing is an essential phase in biomedical image processing because it allows pictures to be properly coordinated to achieve high accuracy. The suggested paradigm necessitates costly calculations and uniform input image formats. There are dark spots on the biopsy slides, and some of them are stained with medical solutions and blood; as a result, the images on the slides have different colors. This led to the calculation of each image's average RGB color. Next, each image's scale was adjusted to determine the color consistency. To minimize noise in the image patches, such as bright and dark pixels related to taking the pictures under the microscope, a 3 × 3 median filter has been applied. The boundaries of the regions of interest were finally made visible by Gaussian and Laplacian filters, which also eliminated artifacts and improved image contrast. The low-frequency data was then retained after the images were run through a Gaussian noise filter, which eliminated high-frequency data. It is important to note that the linear low-frequency spatial filter used for noise reduction and blurry images is the Gaussian filter smoothing factor. Equation (1) explains the operation of a Gaussian filter.

$$H(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (1)$$

where σ is the standard deviation of x and μ denotes the mean of x. Equation (2) was then used to create the images and pass them through a Laplacian filter in order to reveal the borders of the lesion in the images of diseased tissue.

$$\nabla^2 f = \frac{df}{d^2x} + \frac{d^2f}{d^2y} \quad (2)$$

where x and y stand for the image's pixel coordinates. The next step involved overlapping the filter outputs to create an improved image, which was then produced by subtracting the improved Laplacian input from the Gaussian filter output, as shown in Equation (3) to enhance the histopathology images.

$$O(x) = h(x) - \nabla^2 f \quad (3)$$

After the enhancing process, which included eliminating noise, boosting contrast, and highlighting the margins of the area of focus, a set of optimal pathological images are displayed in Fig 4.

Figure 4: Pre-processed image for Normal and OSCC.
Source: Own elaboration.

c. Feature Extraction using Resnet 50 and Efficientnet B5

The first approach was based on a feature extraction using both Resnet 50 model and Efficientnet B5 model which was described below:

1. Resnet50 Architecture

ResNet50 is a deep convolution neural network framework created by Microsoft Research in 2015. It has 50 layers for pooling and convolutional operations. The remaining links in the design enable data to move directly from the input to the output layers, skipping the intermediate layers. This lessens the possibility of the vanishing gradient problem in DNN. A thorough explanation of the ResNet50 architecture as shown in Fig .6 may be found below.

- Input Layer: accepts a 224 × 224 x 3 input image.
- Convolution Layer: The convolution layer, which makes up the first layer of ResNet50, has 64 filters with a stride of two and a size of 7 by 7. This layer produces a feature map with dimensions of 112 × 112 x 64.
- Layer of Batch Normalization: In this layer, the output from the preceding layer is normalized by dividing it by the standard deviation and subtracting the mean. This expedites the training process and lessens the impact of covariate shifts.
- Max Pooling Layer: This layer uses a 3x3 pool size and a 2 stride to carry out the max pooling operation. This layer produces a 56x56x64 feature map as its output.
- Residual Blocks: ResNet50 consists of 16 residual blocks, each of which combines multiple convolution layers with an input-to-output skip connection. These building components aid in solving the vanishing gradient issue and improve the network's accuracy.
- Global Average Pooling Layer: This layer creates a single value for each feature map by executing an average pooling operation overall feature maps. This layer produces a vector with a size of 2048.
- Fully Connected Layer: This layer generates a probability distribution across the output classes using the output of the preceding layer as input. The number of neurons in this layer is the similar as the number of output classes.

ResNet50 has excelled at a variety of computer vision tasks, containing object detection, semantic segmentation, and image classification. About 4 billion (FLOPs) are needed to process a single image with ResNet50. The deep features from the pre-processed histological images are extracted using the Resnet 50 model as shown in the Fig 5.

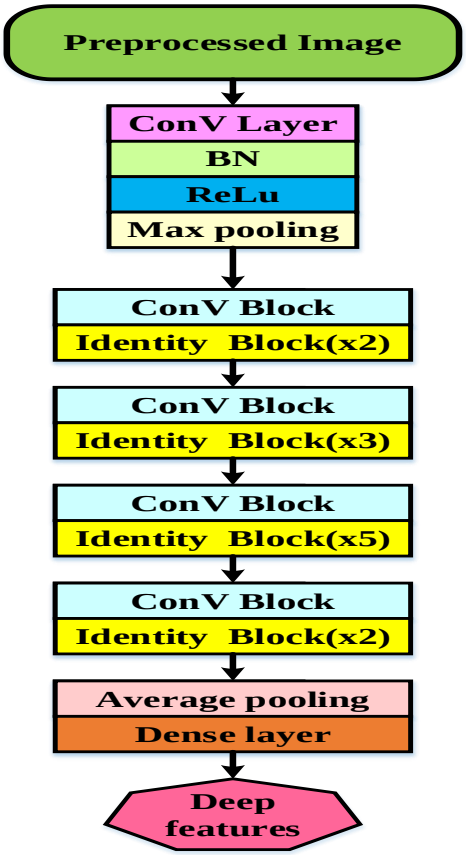

Figure 5: The Resnet 50 model for the feature extraction from the pre-processed images.
Source: Own elaboration.

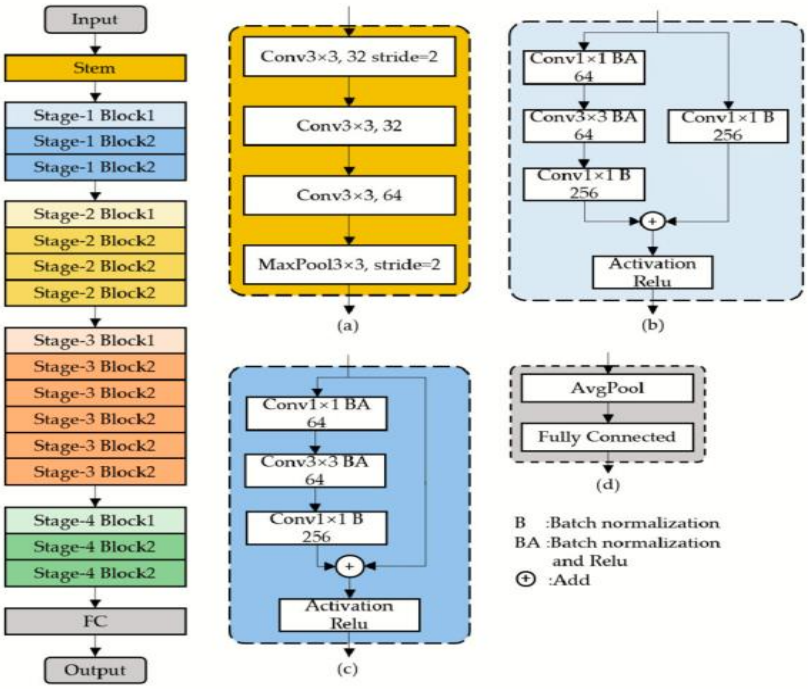

Figure 6: The Architecture of RESNET 50.
Source: Own elaboration.

2. EfficientNet-B5

In comparison to state-of-the-art works, Google AI's Efficient Net can enhance the performance of DR severity categorization. Through Efficient Net scaling, CNN's breadth, depth, and resolution may all be collectively increased. There have been seven iterations of Efficient Net released thus far, numbered B0 through B7. These variants can be distinguished by the number of layers that are employed. For instance, 237 and 813 layers, respectively, are employed in EfficientNet-B0 as well as EfficientNet-B7. In this case, the image size is well-defined as (456, 456, and 3) and we utilize EfficientNet-B5 as shown in Fig 7. This shows that the channels, width, and height of the image are 3, 456, 456, and 456, respectively. Here, a batch size of 4 is utilized. After extracting the deep features from the Resnet 50 model, the remaining features are extracted using the Efficientnet B5.

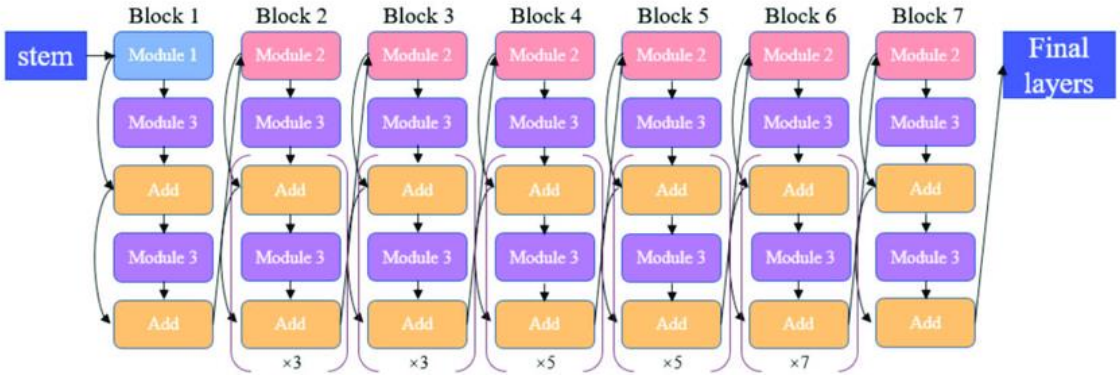

Figure 7: The Efficient net B5 Architecture.
Source: Own elaboration.

d. Oversampling Method for Imbalance Data

An imbalanced data is one in which a single class predominates over the others (i.e., one class has a significantly higher record count than the others). This presents challenges for machine learning algorithms. It distorts the classification, which has an adverse effect on the results. SMOTE was employed in this work to address the issue of unbalanced data during the feature selection procedure. An imbalanced data is one in which a single class predominates over the others (i.e., one class has a significantly higher record count than the others). This presents challenges for machine learning algorithms. It distorts the classification, which has an adverse effect on the results. SMOTE was employed in this work to address the issue of unbalanced data during the feature selection procedure.

1. SMOTE

SMOTE is an oversampling approach suggested by Chawla et al. [26]. This generates novel minority class occurrences at random from the closest minority class neighbors in the sample. In order to produce original minority class instances, these instances are constructed with

consideration for the attributes present in the original dataset. It is utilized to fix imbalances in a several domains, including cervical, liver, and breast cancer detection. SMOTE employs Equation (4) to raise the minority class.

$$x_{syn} = x_i + (x_{knn} - x_i) \times t \quad (4)$$

SMOTE first recognizes the feature vector x_i and finds the KNN x_{knn}. The variation among the feature vector and the KNN is then calculated. The variation is then multiplied by a random value among 0 and 1. To find a novel point on the line segment, it then appends the output number to the feature vector. Finally, it finds feature vectors by repeating the previous processes.

e. Feature scaling

Feature scaling, also referred to as standardization, is a step in the data preprocessing procedure. The purpose of it is to standardize the data within a given range. Feature scaling aids in speeding the algorithm's computations. Feature scaling of the data is a common condition for all research that uses Scikit-learn, Keras, and deep learning. There are variables in the dataset utilized in this experiment with varying scales. Feature scaling of the dataset is done to transform the feature vector into a format that is more appropriate for deep learning algorithms. For feature scaling of the dataset, numerous scalers are available. The most often utilized ones are RobustScaler(), Normalizer(), MinMaxScaler(), and StandardScaler().The proposed model utilizing the Normalizer()scaler for feature scaling.

Normalizer() scales each sample's data to a unit norm. This scaling method scales the SciPy sparse matrix as well as NumPy arrays. For instance, if the feature values were x, y, and z, the transformed normalized value for x would be as follows:

$$z = \frac{x_i}{\sqrt{x_i^2 + y_i^2 + z_i^2}} \quad (5)$$

f. Feature Selection

In this feature selection approach the optimal feature set of OSCC was selected by the proposed metaheuristic based optimization algorithm of Statistic gain Dynamic Remodelled Whale Optimization Algorithm (SDRWOA) which is depend on the traditional whale optimization algorithm.

1. Whale optimization Algorithm (WOA)

The WOA is a bio-inspired optimization method that was recently suggested. Finding and pursuing prey mimics the communal hunting behaviour of a humpback whale. As seen in Fig. 8 WOA simulates the upward-spirals and double loops bubble-net hunting method, in which whales dive below, start to bubble-net their prey in a spiral, and then swim up towards the surface.

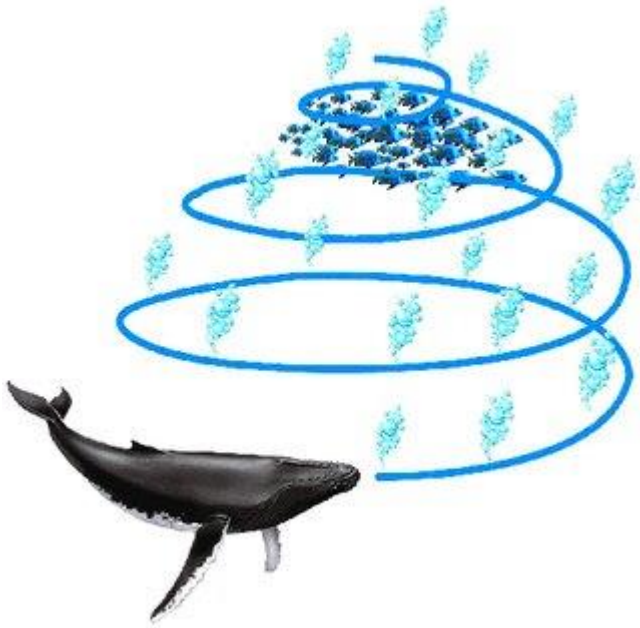

Figure 8: Humpback whales' bubble-net hunting method.
Source: Own elaboration.

WOA is used to generate a set of random solutions, or candidate solutions, to a given optimization problem at the start of the search process. Once the termination requirements are satisfied, a population of search agents will modify their positions in favour of the best search agent. The WOA mathematical model is described by equation 6, where a probability of 0.5 is expected to select either updating the spiral mechanism during optimization or the shrinking encircling mechanism:

$$\overrightarrow{Y}(t+1) = \begin{cases} \overrightarrow{Y^*}(t) - \vec{A} \cdot \overrightarrow{D}, & if\ P < 0.5 \\ e^{bl} \cdot Sin(2\pi l) + \overrightarrow{Y'}(t), & if\ P \geq 0.5 \end{cases} \quad (6)$$

Where b is a constant determining the spiral shape, p random number is $\in [0,1]$, t is the current iteration, Y' is the best solution thus far, Y is the position vector, and l random number is $\in [1,1]$;

$\vec{D}$ is provided by:

$$\vec{D} = \left|\vec{C} \cdot \overrightarrow{Y'}(t) - \vec{Y}(t)\right| \quad (7)$$

While, $\vec{A}$ and $\vec{C}$ are coefficient vectors, determined by:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (8)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (9)$$

where $\vec{a}$ dropped linearly during the length of iterations from 2 to 0, and $\vec{r}$ is a random vector with a range of $\in [0,1]$.
The following factors show how far away the i^{th} whale is from its prey:

$$\vec{D} = \left|\overrightarrow{Y'}(t) - \vec{Y}(t)\right| \quad (10)$$

To possess a global optimizer, exploration is done using the vector A; $1 < \vec{A} < -1$. W hereby, the search agent position is upgraded in accordance with a search agent selected at random using $\overrightarrow{Y_{rand}}(t)$**:**

$$\vec{D} = \left|C.\overrightarrow{Y_{rand}}(t) - \vec{Y}(t)\right| \quad (11)$$

$$\vec{Y}(t+1) = \overrightarrow{Y_{rand}}(t) - \vec{A} \cdot \vec{D} \quad (12)$$

2. Remodelled Whale Optimization Algorithm

The WOA algorithm's desired path toward convergence can be broken down into two opposing stages: exploration vs. exploitation. Whales do not congregate near the local minima during the exploration phase; instead, they disperse across the whole search space. Whales attempt to converge on the global minimum during the exploitation phase by seeking locally near-found solutions.

Equation 8 is used by the WOA algorithm to linearly reduce the distance control parameter a from 2 to 0 when it transitions between the exploration and exploitation phases. When $|A| \geq 1$, the pivot is the best option to update the search agents' positions because half of the iterations are used for exploration. The ideal solution discovered so far acts as the pivot point, with the other half being devoted to exploitation; $|A| < 1$.

$$A = 2\left(1 - \frac{t}{t_{max}}\right) \quad (13)$$

Where the symbols t and t_{max} denote, respectively, the current iteration and the maximum number of iterations. In general, more exploration is correlated with more randomness, while more exploitation is correlated with less randomness and is likely to produce subpar optimization results. To ensure a precise approximation of the global optimum, the proposed Cosine Remodelled WOA algorithm (RWOA) tries to strike the correct balance among the exploration and exploitation phase.

For the control parameter decay throughout the duration of iterations, RWOA uses a cosine function rather than a linear function, as seen in equation 13.

$$A = 1 - \text{Cosin}(\pi\frac{t}{t_{max}}) \quad (14)$$

Additionally, Dynamic Remodelled WOA (DRWOA) is suggested to improve the exploitation capacity (local search) of the RWOA algorithm. The mutation operator in the DRWOA algorithm tries to alter the result around the best result so far, Y′, or around a randomly chosen result, Y_{rand}. Additionally, the Crossover operator is used to produce a solution that is halfway between the solution produced by the mutation operation Y_{mut} and the solution Y_t.

3. Dynamic Remodelled WOA for Feature Selection Problem

An innovative Statistic Gain DRWOA method (SDRWOA) is offered to tackle the feature selection problem. The SDRWOA algorithm is designed to address binary optimization issues. As a result, the whale position is denoted by a binary vector, with 1 denoting that a certain feature is picked and 0 denoting that it is not. The vector's length is determined by the quantity of features in the original dataset. For population initialization, SDRWOA adapted Statistic Gain (SG), where features with appropriate entropy are represented by 1; otherwise, the value is adjusted to 0. As the placements of the agents are frequently the ones nearest to the ideal solution, the SG initialization procedures of SDRWOA are employed to ensure a large initialization in order to boost the local search ability.

The two primary aims of feature selection are to increase classification accuracy while decreasing the number of features. SDRWOA is used to adaptively search for the ideal feature combination while keeping these two objectives in mind. The fitness function required to assess every whale's position is given by the following equations:

$$\text{Fitness} = \alpha(1 - E_R) \quad (15)$$

where E_R is the classification error rate. The term α, where $\alpha \epsilon (0.5,1)$, indicates the relative significance of the number of selected attributes and the classification accuracy. The SDRWOA pseudocode is displayed in Algorithm 1:

Algorithm 1 Pseudo code of SDRWO Algorithm
Input:
Number of whales n
Number of optimization iterations Max_Iter
Output:
Optimal whale binary position Y^*
Step 1: Compute the entropy of each feature $f \in dataset$.
Step 2: Initialize the n whale's population positions $\in entropy(f) > 1$.
Step 3: Initialize a, A and C.
Step 4: $t = 1$
Step 5: while $t \leq Max\_Iter$ do
Step 6: Compute the fitness of every search agent.
Step 7: Y^* = the best search agent.
Step 8: for each search agent do
Step 9: Update a b y equation 13
Step 10: Update A, C and l
Step 11: Generate randomly $p \in [0,1]$
Step 12: if $p < 0.5$ then
Step 13: if $|A| < 1$ then
Step 14: perform $Y_{mut} = mutation\ (Y^*)$
Step 15: update $Y_{t+1} = Crossover(Y_{mut}, Y_t)$
Step 16: else if $|A| \geq 1$ then
Step 17: Select a random search agent Y_{rand}
Step 18: perform $Y_{mut} = mutation\ (Y_{rand})$
Step 19: update $Y_{t+1} = Crossover(Y_{mut}, Y_t)$
Step 20: end if
Step 21: else if $p > 0.5$ then
Step 22: Upgrade position Y_{t+1} by equation 6(b)
Step 23: end if
Step 24: Compute the fitness of each search agent
Step 25: Upgrade Y^* if a better solution exists
Step 26: end for
Step 27: $t = t + 1$
Step 28: end while
Step 29: Return Y^*

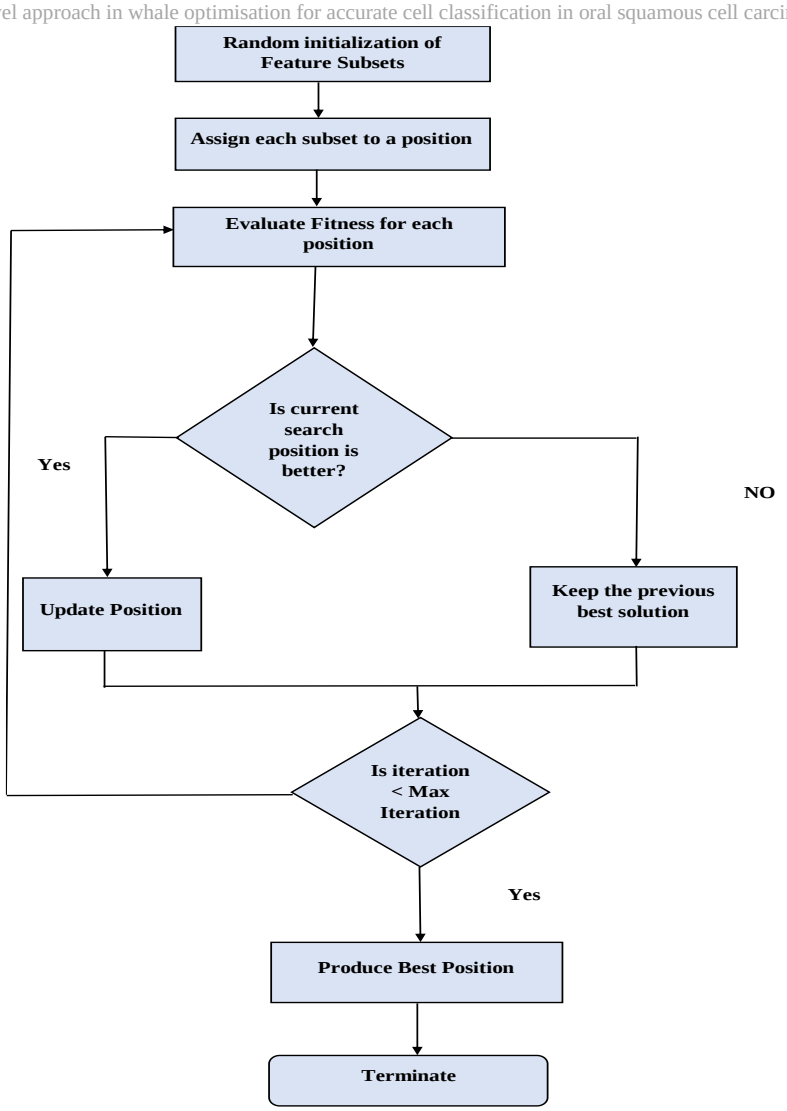

Figure 9: Flowchart for the proposed Feature Selection.
Source: Own elaboration.

4. Proposed Methodology for Feature Selection

The specifics of the Statistic gain Dynamic Remodelled Whale Optimization Algorithm (SDRWOA) for feature selection that has been proposed are shown in Algorithm 1. The Crossover operator is used in this technique to perform feature selection in order to obtain a solution that is halfway between the resultant solution from the mutation operation and the solution and then convert the resulting solution into binary solution. In order to evaluate the quality of the features that are chosen and get the highest fitness for oral cancer detection, the K-Nearest Neighbour method is also used. Algorithm 2 also provides a high-level summary of the suggested feature selection algorithm. This algorithm performs the SDRWOA to obtain the desired characteristics at each iteration of the procedure. KNN is used to evaluate these chosen features with the use of a fitness function. The algorithms then return the set of features that outperform all other feature sets in terms of fitness at the end of these iterations. The basic flowchart for the proposed feature selection process is shown in Fig 9.

Algorithm 2: The Proposed Feature Selection Algorithm

Step 1: Initialize SDRWOA population, parameters, and configurations
Step 2: while $t \leq Max\_Iter$ do
Step 3: Apply the SDRWO algorithm
Step 4: Calculate the objective function and select the best solution
Step 5: Train KNN and calculate the error
Step 6: Compute Fitness
Step 7: Upgrade Fitness
Step 8: end while
Step 9: Return features that achieve the best fitness

g. Random Forest Classification

Random Forest is a well-known machine learning method that is employed for a variety of categorization problems. An ensemble of classifiers with a tree topology is called a Random Forest. Every tree in the forest has a unit vote, indicating which class designation is more likely based on every input. It is an effective ensemble that can discover non-linear patterns in the data and is a quick method that is resistant to noise. It simply handles both category and numerical data. One of Random Forest's primary benefits is that it does not suffer from over-fitting, even with more trees added to the forest. Its parallel architecture handles data imbalance and makes it a quick classifier.

The number of trees that make up the forest, or n_trees, is the parameter utilized in random forests. It is assumed to be 100 here. The number of decision trees is controlled by the following hyper parameters, which are necessary because a random forest generates ensembles of many trees:

1. max_features (the quantity of characteristics to be chosen at random from the data)
2. max_depth (for tree pre-pruning)
3. For classification, max_features = sqrt(n_features).

By increasing the accuracy of predictive models, ensemble learning approaches improve their performance. Random Forest constructs a tree as an ensemble of decision trees that employ bagging and random sampling of training points. It starts with an uncorrelated forest of trees, creates several decision trees, and then combines them to provide the classifier a forecast that is more reliable and accurate. Finally, the cancer type is classified into poor, moderate, and well using the Random Forest Classifier based on the optimized features obtained by the proposed SDRWOA algorithm.

IV. RESULTS AND DISCUSSION

This section discusses the suggested method's efficiency as well as the outcomes of its application. The comparative results between this study and earlier publications are also shown. The PYTHON program is used to implement and assess the suggested techniques on an Intel Core 5 configuration computer with 16 GB of RAM and a 2.6 GHz processor. Through the use of models, data pre-processing, and Python 3 libraries (such as NumPy, pandas, seaborn, and Sk learn packages), the suggested machine learning-based approach to a standard filter is assessed. Additionally, the Keras Python library and Tensor Flow 2.10 are used to generate the model. The histological images from the OSCC oral dataset publicly available in kaggle is utilized for the proposed model.

1. Data Set Splitting

The primary goal of this study was to use novel approaches depend on feature extraction, classification, and machine learning algorithms in conjunction with hybrid techniques between Resnet 50 and Efficientnet B5 models to categorize histological images for the early identification of oral OSCC. The 5192 histological images from the biopsy that make up the OSCC data set are split into three classes: mild, moderate and well. At random, the data set was split into two groups: 20% was used for testing and the remaining 80% for training and validation.

2. Evaluation parameters

The confusion matrix is used to evaluate performance. The positive classes that have been deemed to be positive are indicated by the term "true positive" (TP), which also indicates the true positive rate. True negative (TN) is a true negative rate that emphasizes the negative class that has been defined as negative; false positive (FP) is a false positive rate that shows the negative class that has been defined as positive; and false negative (FN) is a positive class that has been defined as negative. The effectiveness of suggested deep learning algorithms is calculated using recall, accuracy, F1-score, precision and ROC curve.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (16)$$

Precision is defined as the percentage of actual positive cases relative to all projected positive patterns. It can be determined by applying the subsequent formula.

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

The percentage of true positives a classifier receives is used to calculate recall, which gauges how well it can recognize positive class patterns. The method might be sensitive without being specific, or precise without being sensitive. Recall may be calculated using.

$$Recall = \frac{TP}{TP+FN} \quad (18)$$

The F1-score is a more accurate predictor of incorrectly detected patterns than the accuracy. It is computed by using

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (19)$$

SDRWOA Algorithm is used in the suggested system to select the best data features from the histological images based on fitness at different iterations. As a result of employing the SDRWOA Algorithm, the suggested system exhibits high efficiency with little parameterization as shown in Fig 10.

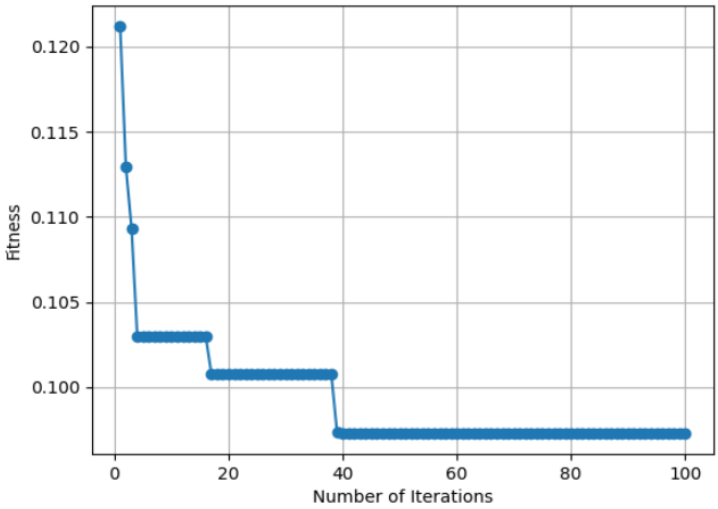

Figure 10: Fitness using SDRWOA Algorithm.
Source: Own elaboration.

Typically, ROC curves have the true positive rate (TPR) and false positive rate (FPR) shown on the Y and X axes of Fig 11, accordingly. The "ideal" placement is therefore the upper left corner of the plot, which has a TPR of 1 and an FPR of 0. This implies that a larger area under the curve (AUC) is generally preferred, even though it is not very practical. The "steepness" of ROC curves is also significant, as it is optimum to make best use of the TPR whereas reducing the FPR. The suggested system uses three classes (cancer types) ranging from class 0 to class 2; in terms of TPR and FPR values in order to accomplish the multiclass ROC.

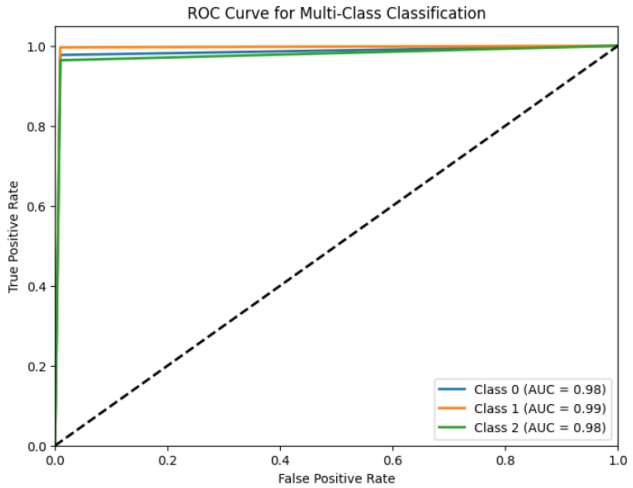

Figure 11: Multiclass ROC Curve of the proposed model.
Source: Own elaboration.

The Random Forest Classifier method's accuracy to classify the cancer types has led to an increase in TPR values for every class. The reduction of false positive predictions by Ada-Max leads to a lower FPR for every class. By carefully selecting the threshold, the model may find cases of actual positivity while lowering false positive rates. By using a threshold optimization process, the multiclass ROC is guaranteed to provide good TPR and FPR values.

The confusion metrics pertaining to TP, FP, TN, and FN have a major impact on the processing and classification of the oral cancer by the proposed system. The number of histological images that are properly classified the oral OSCC as well, moderate and poor is also determined by the confusion measures. The confusion metrics of the proposed model are shown in Figure 12. The prediction values for each class are displayed as percentages.

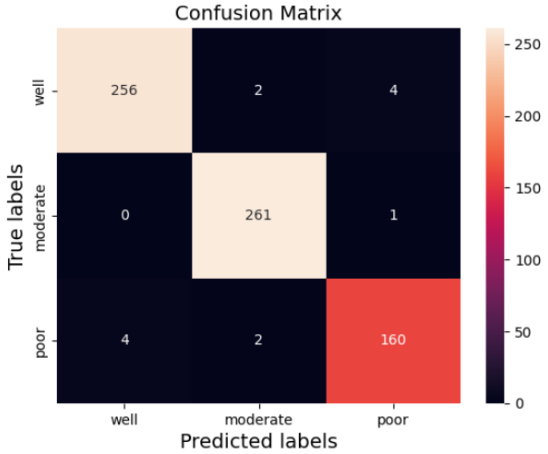

Figure 12: Confusion Matrix for the proposed model.
Source: Own elaboration.

The diagonal values of the confusion matrix correspond to accurately anticipated examples of a certain class. Figure 13 shows how successfully the proposed models classify the OSCC oral cancer types. There are three classes in total, ranging from class 0 to class 2, and each class is represented by one of the three types of cancer that are currently affected in the given histological images, which include well, moderate and poor. As a result, the TP value is predicted for 160 data from total data volume in class 0, 261 data from total data volume in class 1, 256 data from total data volume in class 2 respectively. The greatest predicted value (TP) for the type of oral cancer for the true labels poor, moderate and well is 261 for class 1, respectively.

3. Comparison

This section uses cutting-edge machine learning methods to compare the suggested model's performance measures, such as accuracy, precision, recall, and F1-score.

Table 1: Accuracy Comparison of classifiers.

Algorithms	Accuracy
KNN	83
Naïve Bayes	73
SVM	96
Logistic Regression	97
Proposed	98

Source: Own elaboration.

Table .1 depicts the comparison of the suggested classifier method with other standard machine learning classifiers in terms of accuracy. The suggested classifier model is related with the implemented standard classifiers such as KNN, Naïve Bayes, SVM and Logistic Regression classifier models. These models are implemented along with the proposed model and compared in terms of the performance metrics such as accuracy, precision, recall and F1-score. The accuracy finding shows that the suggested classifier beats the other classifiers by achieving 98% as shown in Fig 13 due to the enhanced classification process with optimal feature selection by employing the SDRWOA Algorithm.

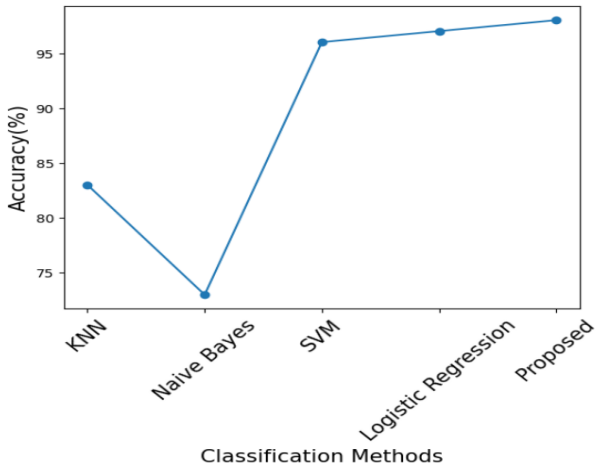

Figure 13: Accuracy Comparison.
Source: Own elaboration.

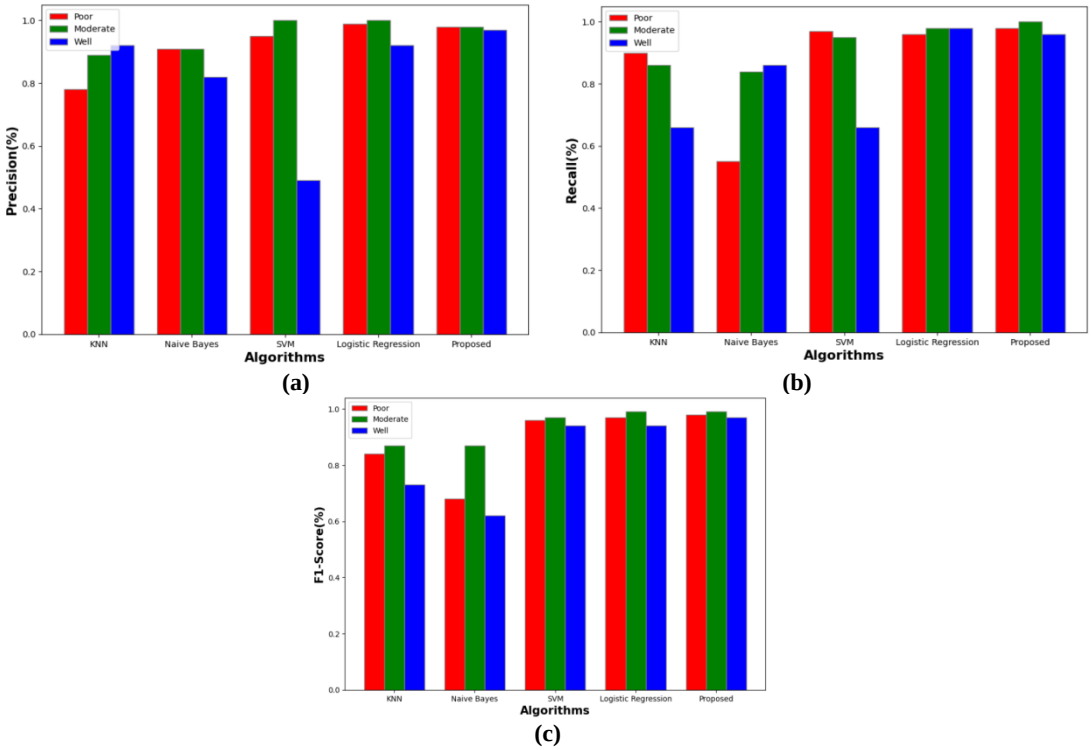

Figure 14: Performance Comparison with classifiers (a) Precision (b) Recall (c) F1-score.
Source: Own elaboration.

Table 2: Performance Comparison with other classifiers.

	Precision			Recall			F1-score		
Algorithms	Poor	Moderate	Well	Poor	Moderate	Well	Poor	Moderate	Well
KNN	78	89	92	90	86	66	84	87	73
Naïve Bayes	91	91	82	55	84	86	68	87	62
SVM	95	100	49	97	95	66	96	97	94
Logistic Regression	99	100	92	96	98	98	97	99	94
Proposed	98	98	97	98	100	96	98	99	97

Source: Own elaboration.

Table .2 depicts the comparison of the suggested model with other state of art machine learning classifiers in terms of precision, recall and F1-score. The suggested classifier model is related with the implemented standard approches such as KNN, Naïve Bayes, SVM and Logistic Regression classifier models. These models are implemented along with the proposed model to classify the cancer image into three classes such as poor, moderate and well and compared in terms of the performance metrics such as accuracy, precision, recall and F1-score for each three classes. The finding shows that the suggested classifier beats the other classifiers by achieving the highest value of overall precision of around 97.6 % and overall recall and F1-score of about 98% as shown in Fig 14 due to the utilization of the SDRWOA Algorithm along with Random forest classifier. However, in some instance the precision for SVM and Logistic Regression classifier exceeds 2% than proposed model only for moderate class and similarly the recall for the Logistic Regression classifier exceeds 2% than proposed model only for well class.

V. CONCLUSION

In conclusion, the research conducted in this paper demonstrates promising developments in the early detection of OSCC, a predominant head and neck cancer. By leveraging novel methodologies rooted in feature selection and classification, this work underscores the possible of artificial intelligence in enhancing OSCC diagnosis. The integration of cutting-edge hybrid techniques that combine deep learning and machine learning strategies has proven instrumental in addressing the challenges associated with the early identification of this heterogeneous tumour. The suggested classifier method is related with the implemented standard classifiers such as KNN, Naïve Bayes, SVM and Logistic Regression classifier models in terms of several performance measures such as accuracy, precision, recall and F1-score. Thus, with the proposed model showcasing superior performance, attaining an impressive overall accuracy, recall, F1-score, and precision of 98% and 97.6% accordingly, this approach covers the way for the progress of highly precise and effective diagnostic tools for OSCC, marking a major stride in the field of cancer diagnosis and treatment.

Conflict Of Interest
They author declare that have no conflict of interest

Ethical Approval
Institutional Review Board approval was not required.

Consent for Participate
All contributors agreed and given consent to participate.

Consent for Publication
All contributors agreed and given consent to Publish.

Data availability
No data, models, or code were generated or used during the study

Competing interests
None

Funding
The authors state that this work has not received any funding.

Author Contribution
The authors confirm contribution to the paper as follows and all authors reviewed the results and approved the final version of the manuscript.

Acknowledgements
The authors would like to thank the Deanship of G. H. Raisoni College of Engineering & Management for supporting this work.

VI. REFERENCIAS

[1] S. Warnakulasuriya, & J.S. Greenspan, "Textbook of oral cancer: Prevention," diagnosis and management, vol. 1, 2020, pp. 1-452.
[2] V. Deshmukh, & K. Shekar, "Oral squamous cell carcinoma: Diagnosis and treatment planning," Oral and maxillofacial surgery for the clinician, 2021, pp. 1853-1867.
[3] A. Romano, D. Di Stasio, M. Petruzzi, F. Fiori, C. Lajolo, A. Santarelli, & M. Contaldo, "Noninvasive imaging methods to improve the diagnosis of oral carcinoma and its precursors: State of the art and proposal of a three-step diagnostic process," Cancers, vol. 13, no. 12, 2021, pp. 2864.
[4] P.M. Speight, S.A. Khurram, & O. Kujan, "Oral potentially malignant disorders: risk of progression to malignancy," Oral surgery, oral medicine, oral pathology and oral radiology, vol. 125, no. 6, 2018, pp. 612-627.
[5] B. Acs, M. Rantalainen, & J. Hartman, "Artificial intelligence as the next step towards precision pathology," Journal of internal medicine, vol. 288, no. 1, 2020, pp. 62-81.
[6] N. Haefner, J. Wincent, V. Parida, & O. Gassmann, "Artificial intelligence and innovation management: A review, framework, and research agenda," Technological Forecasting and Social Change, vol. 162, 2021, pp. 120392.
[7] K. Kaba, M. Sarıgül, M. Avcı, & H. M. Kandırmaz, "Estimation of daily global solar radiation using deep learning model," Energy, vol. 162, 2018, pp. 126-135.
[8] I. Lorencin, N. Anđelić, V. Mrzljak, & Z. Car, "Genetic algorithm approach to design of multi-layer perceptron for combined cycle power plant electrical power output estimation," Energies, vol. 12, no. 22, 2019, pp. 4352.
[9] H. Chen, & J.J. Sung, "Potentials of AI in medical image analysis in Gastroenterology and Hepatology," Journal of Gastroenterology and Hepatology, vol. 36, no. 1, 2021, pp. 31-38.
[10] S. Stolte, & R. Fang, "A survey on medical image analysis in diabetic retinopathy," Medical image analysis, vol. 64, 2020, pp. 101742.
[11] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, & C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical image analysis, vol. 42, 2017, pp. 60-88.
[12] A. Singh, S. Sengupta, & V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," Journal of imaging, vol. 6, no. 6, 2020, pp. 52.
[13] S. Sharma, & R. Mehra, "Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight," Journal of digital imaging, vol. 33, 2020, pp. 632-654.
[14] H.H. Tan, & K.H. Lim, "Vanishing gradient mitigation with deep learning neural network optimization," In 2019 7th international conference on smart computing & communications (ICSCC), pp. 1-4.
[15] A. Tetarbe, T. Choudhury, T.T. Toe, & S. Rawat, "Oral cancer detection using data mining tool," In 2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2017, pp. 35-39.
[16] K. Lalithmani, & A. Punitha, "Detection of oral cancer using deep neural based adaptive fuzzy system in data mining techniques," Int J Rec Tech Eng, vol. 7, 2019, pp. 397-404.
[17] S. Tabibu, P.K. Vinod, & C.V. Jawahar, "Pan-Renal Cell Carcinoma classification and survival prediction from histopathology images using deep learning," Scientific reports, vol. 9, no. 1, 2019, pp. 10509.
[18] R. Yamashita, M. Nishio, R.K.G. Do, & K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights into imaging, vol. 9, 2018, pp. 611-629.
[19] N. Rajawat, B.S. Hada, M. Meghawat, S. Lalwani, & R. Kumar, "C-covidnet: A cnn model for COVID-19 detection using image processing," Arabian Journal for Science and Engineering, vol. 47, no. 8, 2022, pp. 10811-10822.
[20] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, & L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," Journal of big Data, vol. 8, 2021, pp. 1-74.
[21] T.Y. Rahman, L.B. Mahanta, H. Choudhury, A.K. Das, & J.D. Sarma, "Study of morphological and textural features for classification of oral squamous cell carcinoma by traditional machine learning techniques," Cancer Reports, vol. 3, no. 6, 2020, pp. e1293.
[22] T.Y. Rahman, L.B. Mahanta, A.K. Das, & J.D. Sarma, "Automated oral squamous cell carcinoma identification using shape, texture and color features of whole image strips," Tissue and Cell, vol. 63, 2020, pp. 101322.
[23] S.M. Fati, E.M. Senan, & Y. Javed, "Early diagnosis of oral squamous cell carcinoma based on Histopathological images using deep and hybrid learning approaches," Diagnostics, vol. 12, no. 8, 2022, pp. 1899.
[24] C.H. Chan, T.T. Huang, C.Y. Chen, C.C. Lee, M.Y. Chan, & P.C. Chung, "Texture-map-based branch-collaborative network for oral cancer detection," IEEE transactions on biomedical circuits and systems, vol. 13, no. 4, 2019, pp. 766-780.

[25] E.N. Subhija, & V.G. Reju, “An image patch selection algorithm for the detection of Oral Squamous Cell Carcinoma using textural and morphological features”.
[26] N.V. Chawla, K.W. Bowyer, L.O. Hall, & W.P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” Journal of artificial intelligence research, vol. 16, 2022, pp. 321-357.