New approach Generative AI Melanoma Data Fusion for classification in dermoscopic images with Large Language Model

Adriell Gomes Marques*§, Marcus Vinicius Candido de Figueiredo*§, José Jerovane Da Costa Nascimento‡§, Cidcley Teixeira de Souza*, Carlos Mauricio Jaborandy de Mattos Dourado Junior*§, Victor Hugo C. de Albuquerque‡, Luís Fabrício de Freitas Souza†§

†Universidade Federal do Cariri, Juazeiro do Norte, CE - Brazil
*Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Fortaleza, CE - Brazil
‡Universidade Federal do Ceará, Fortaleza, CE - Brazil
§Laboratório de Inovação em Sistemas de Inteligência Artificial (LISIA), Juazeiro do Norte / Fortaleza, CE
Email: fabricio.freitas@ufca.edu.br

2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) | 979-8-3503-7603-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/SIBGRAPI62404.2024.10716298

***Abstract*—Skin cancer is a disease that causes thousands of deaths each year. Early diagnosis and monitoring the progression of the disease are crucial factors for the treatment and health indicators of a society. This study presents an innovative approach for the detection, segmentation, and classification of melanomas in dermoscopic images using advanced Computer Vision and Artificial Intelligence (AI) methods. Specifically, it applies Large Language Model (LLM) solutions for pre-diagnosis results through generative AI. This work explores combinations of methods for melanoma detection and segmentation based on the YOLO and SAM architectures, achieving 99% accuracy, surpassing various studies in the literature. The classification phase is based on a pipeline integrating feature extraction and selecting the best combination for melanoma region classification, achieving an accuracy of 86.0%, also outperforming different studies in the literature.**

I. Introduction

Melanoma is a form of skin cancer that develops in the pigment-producing cells known as melanocytes. This type of cancer can manifest anywhere on the body, including within pre-existing moles [1]. Epidemiological assessments of global cancer data estimate that in 2020, there were approximately 325,000 new cases and 57,000 deaths due to melanoma [2]. It is projected that by 2040, there will be 510,000 new cases and 96,000 deaths from melanoma [2].

Some warning signs of melanoma, which distinguish it from simple moles or warts, include its asymmetry, with one half of the spot not matching the other; irregular borders, which may be notched or blurred; color variation within the same spot; diameter, generally larger than 6 mm, but it can be smaller; and changes in size, shape, or color over time. These characteristics are identified in new skin spots using the ABCDE guide (Asymmetry, Borders, Color, Diameter, Evolution) [3]. The dermoscopic examination aims to generate high-definition images with good visualization of the regions where lesions are present on the patient's skin. This allows the specialist to conduct a detailed visual inspection, identifying pigmentation patterns, shape, measuring diameter, among other methods used in the diagnosis of this pathology [4].

Detection, segmentation, and classification systems are fundamental methods for pre-diagnosis, optimizing the waiting time for diagnoses and assisting medical diagnosis [5].

Considering the challenges surrounding melanoma diagnosis, its high incidence and mortality rates, characteristics, and existing technologies capable of assisting medical diagnosis in medical imaging, this study proposes the detection, segmentation, and classification of melanoma in dermoscopic images. This work proposes a new model capable of precisely detecting and segmenting the lesion area, as well as classifying its diameter using transfer learning. This approach provides a pre-diagnosis in the form of text, aided by generative AI. The study introduces a method based on a novel multi-layer approach that detects, segments, and classifies, providing a pre-diagnosis through generative AI.

This study addresses different aspects:

- An approach based on transfer learning for detection, segmentation, extraction and classification of melanomas using data fusion.
- Generative AI model based on LLM for pre-diagnostic assistance in melanoma.
- Detection, segmentation, and classification in dermoscopic images for medical diagnostic assistance using generative AI models.

II. Related Work

The application of machine learning and deep learning techniques in dermatology has been widely explored in the literature [6]. Comprehensive analyses of melanoma detection techniques using CNN models revealed the effectiveness of

these approaches, achieving high accuracy on specific datasets [7].

Bi et al. [8] developed an automatic skin lesion segmentation method based on class-specific deep learning with gradual integration grounded in probability (DCL-PSI). The model aims to enhance lesion segmentation by integrating probabilistic steps, achieving an accuracy rate of 95.30%.

Vasconcelos et al. [9] proposed an automatic segmentation method using geodesic active contour (MGAC) based on mathematical morphology, highlighting its low computational cost. On the PH2 dataset, the method achieved an accuracy of 94.59%, demonstrating its reliability in skin lesion segmentation. It is important to note that, due to the approximate approach of the differential equation, the model is sensitive to variations in lighter lesion areas. Similarly, De et al. [10] proposed a method for skin lesion segmentation using digital image processing, achieving an accuracy of 94.25%.

Al et al. [11] developed a diagnostic aid framework. The authors used the ISIC dataset from different years for training and obtaining results, specifically the versions from 2016, 2017, and 2018. They achieved good accuracy values, above 80%. However, refining the segmentation could potentially improve the area of the segmented lesion, thereby enhancing system performance. Similarly, Yilmaz et al. [12] developed an approach for melanoma classification in dermoscopic images, achieving an accuracy of 82.40%.

With advancements in models and the potential for developing pre-diagnostic approaches based on LLM and clinical workflow knowledge, Wang et al. [13] developed a consultation method known as multi-specialist agent-derived consultation, aiming to perform adaptive fusion of probabilistic distributions from agents regarding potential diseases. The approach requires fewer parameter updates and training time, providing possible pathologies based on the symptoms raised.

III. Materials and Methods

This section will cover the materials, method, evaluation metrics and dataset used in this study.

A. Datasets

Two datasets were used for this study, PH2 and ISIC 2017. The PH2 dataset consists of a total of 200 dermoscopic images provided by the dermatology service of Hospital Pedro Hispano in Portugal [14].

The ISIC 2017 dataset is part of the International Skin Imaging Collaboration Archive. It was used in 2017 for the melanoma detection challenge and includes clinical information about the patient and the lesion [15].

B. YOLO

YOLO (You Only Look Once) is a high-precision and computationally efficient object detection framework, transforming this task into a single regression problem. YOLO is capable of producing bounding boxes, detecting objects of interest in the scene [16].

The derivation of the loss function is crucial for the operation of YOLO, as it measures the error between YOLO's predictions and the true values for each measurement performed [17]. This function is expressed as:

$$\mathcal{L} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] +$$

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] +$$
$$\sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 +$$
$$\sum_{i=0}^{S^2} 1_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad (1)$$

, where λ_{coord} e λ_{noobj} are hyperparameters that control the relative importance of different error terms. 1_{ij}^{obj} e 1_{ij}^{noobj} are binary indicators that represent whether an anchor box contains an object or not. x_i, y_i, w_i, h_i are the predicted coordinates and dimensions of the bounding boxes. $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i$ are the true coordinates and dimensions of the bounding boxes. C_i is the predicted confidence score, and $\hat{C}_i$ is the true confidence score. $p_i(c)$ is the predicted class probability, and $\hat{p}_i(c)$ is the true class probability [17].

C. SAM

The SAM (Segment Anything Model) is a segmentation framework developed by META AI in 2023, capable of identifying objects in an image with high precision and efficiency, using a pipeline of networks combined with a set of processing techniques [18].

The segmentation loss function of SAM, which measures the error between the predicted segmentation mask and the true binary mask, is expressed as:

$$\mathcal{L}_{\text{seg}} = \lambda_{\text{IoU}} \left(1 - \frac{\sum_{i=1}^{N} (y_i \cdot \hat{y}_i)}{\sum_{i=1}^{N} (y_i + \hat{y}_i - y_i \cdot \hat{y}_i)} \right) +$$

$$\lambda_{\text{BCE}} \left(-\frac{1}{N} \sum_{i=1}^{N} \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right) \right) \quad (2)$$

, where λ_{IoU} e λ_{BCE} are hyperparameters that control the relative importance of different error terms. y_i is the value of the true mask at pixel i, provided by the specialist doctor. $\hat{y}_i$ is the value of the predicted mask at pixel i. N is the total number of pixels in the image [18].

D. Extractors and classifiers

The study addressed different feature extractors and classifiers with the aim of identifying the best combination of extractor/classifier for the proposed model. The extractors discussed were LBP (Local Binary Patterns), VGG (Visual Geometry Group), and DenseNet [19]. The classifiers discussed in the models included KNN (K-Nearest Neighbors), MLP (Multilayer Perceptron), Naive Bayes, Random Forest, and SVM (Support Vector Machine) [20].

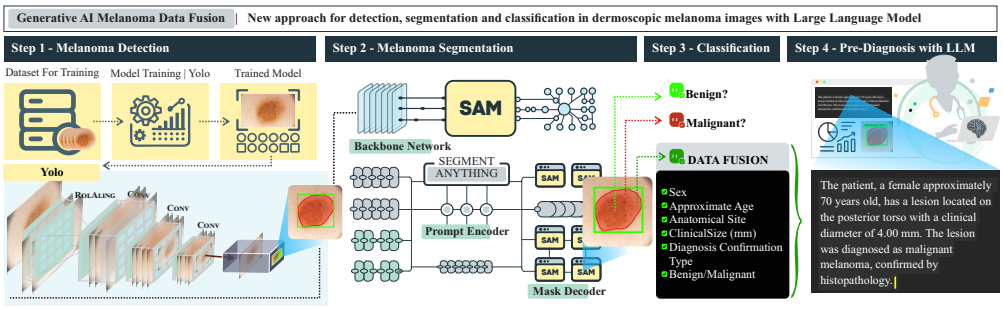

Fig. 1. Proposed approach, Step 1 - Melanoma detection, involving training for bounding box detection and detection mask. Step 2 - Melanoma segmentation using the SAM framework, generating a new detection mask from the YOLO mask. Step 3 - Melanoma classification through the developed pipeline, including grid search, transfer learning, and cross-validation. Step 4 - Pre-diagnosis with LLM through interpretation of attributes extracted by the pipeline.

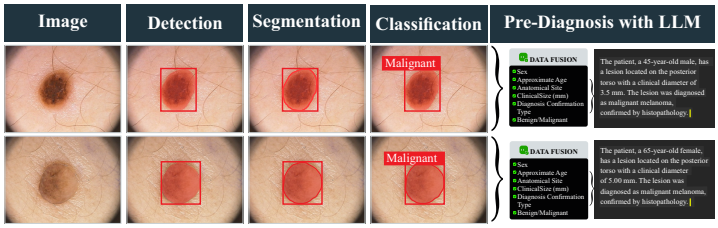

Fig. 2. Results from the different stages of the proposed methodology: two input images were used. In the detection stage, the melanoma region is marked with a bounding box and a binary mask. During segmentation, the edges of the binary mask are refined through concatenation with SAM. Classification pertains to malignancy and other medical attributes. Pre-diagnosis with LLM gathers information obtained from classification and presents it in the form of textual pre-diagnosis.

E. LLM Model

Large Language Models (LLMs) are deep neural networks trained on large volumes of textual data to perform a wide range of natural language processing (NLP) tasks. LLMs employ advanced architectures and are designed for understanding, generating, assimilating, and abstracting text. Words are processed as tokens, which are interconnected to form various contexts [21].

Training LLM models involves obtaining the cross-entropy loss function, which measures the difference between the predicted probability distribution by the model ($\hat{y}$) and the actual probability distribution (y):

$$\mathcal{L}_{\text{cross-entropy}} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{3}$$

where N is the total number of classes. y_i is the true probability of class i. $\hat{y}_i$ the probability predicted by the model for class i [21].

IV. Methodology

The Figure 1 presents the proposed methodology for this study. The method has been divided into 4 main Steps, with Steps 2 and 3 containing their respective Stages. In Step 1, initial melanoma detection occurs using the YOLO framework. Step 2 involves melanoma segmentation, combining the YOLO framework with SAM. In Step 3, classification of the region of interest takes place, with extraction of medical attributes from the image, which ultimately leads to a descriptive pre-diagnosis in Step 4, assisted by the LLM model used.

The PH2 databases have highly defined GTs, which are used in training detection and segmentation models. The ISIC 2017 database has clinical information related to the patient, which is used in the attribute estimation processor in the classification stage.

At step 1 in Figure 1, the dermatoscopic image is inputted into the model and processed using the YOLO framework, which has been previously trained to detect melanoma in two stages: In the first stage, YOLO detects melanoma in the region and outlines it with a bounding box; in the second stage, the model classifies the pixels within the bounding box region to form the corresponding binary mask of the melanoma in the figure. The YOLO versions used in this approach were: YOLOv8l, YOLOv8m, YOLOv8n, YOLOv8s, YOLOv8x, YOLOv9c, and YOLOv9e.

As step 2 in Figure 1, the binary detection mask generated by YOLO is used to initialize the model based on the SAM framework, which generates its own melanoma-specific mask. Then, the two binary masks from YOLO and SAM are combined through an intersection process, resulting in the final segmentation binary mask with improved edges. This fine-tuning process using SAM on YOLO's initialization enhances the performance of the developed segmentation models, leading to higher segmentation accuracy.

As Step 3 in Figure 1 the process involves classifying the segmented melanoma from Step 2, where data related to malignancy classification, size of the melanoma, patient age, and various attributes can be obtained. To achieve this, the

proposed model combines classical and deep feature extractors with machine learning classifiers, aiming to find the best combination for the proposed approach. Additionally, the grid search method is used to optimize hyperparameters of some classifiers. The classifiers used include: KNN, MLP, Naive-Bayes, RandomForest, linear SVM, polynomial SVM, and RBF SVM. Thus, Step 3 of the proposed model is a pipeline for image feature extraction and classification using transfer learning, where deep attributes are classified by machine learning classifiers. Cross-validation was employed to validate results and generate means and deviations.

After classifying the melanoma as benign or malignant, the best extraction and classification set is adopted to obtain the remaining labels, which are: diagnosis_confirm_type, clin_size_long_diam_mm, anatom_site_general e age_approx

The class "diagnosis_confirm_type" refers to the method of confirming the diagnosis of the lesion, which can be histopathology, follow-up, consensus, among others. The class "clin_size_long_diam_mm" refers to the diameter in millimeters of the melanoma in the image. The class "anatom_site_general" refers to the general anatomical location of the lesion on the body surface, which can be head/neck, upper extremity, lower extremity, and torso. Finally, the class "age_approx" refers to the estimated age of the patient.

As step 4 in Figure 1, The process of pre-diagnosing melanoma occurs through the combination of extracted attributes with the pipeline from the previous step. The LLM model is trained to generate formal text that incorporates classified medical information. It integrates the obtained labels into the sentence to succinctly explain essential information to the application user for a preliminary diagnosis. The text provides the patient's age, lesion size, malignancy, location of the lesion on the body, and basic recommendations regarding actions to take concerning the lesion, such as seeking medical assistance.

V. Results and Discussion

This section will address the results obtained in the different stages of the proposed method. The discussion has been divided into three main subsections: detection and segmentation, classification, and comparison with state-of-the-art methods. The evaluation metrics adopted for this study were: Accuracy, Jaccard Index, Dice Coefficient, Sensitivity, Specificity, Precision, Recall, and F1-score [22]. The statistical test used was the Bartlett test.

Figure 2 presents the results of detection, segmentation, and classification for two input images. The network output consists of the segmented melanoma region along with the textual diagnosis, generated from information obtained during the classification stage.

A. Melanoma detection and segmentation

Table I presents the results of the detection stage using different versions of the YOLO framework. The metrics obtained relate to the comparison of YOLO's binary detection mask with the gold standard. The database used in detection and segmentation was the PH2 database.

TABLE I
SEGMENTATION RESULTS FOR THE DIFFERENT MODELS GENERATED FROM THE COMBINATION OF VARIOUS YOLO MODELS WITH THE SAM FRAMEWORK.

Detection Model	ACC(%)	JAC(%)	DICE(%)	SEN(%)	SPE(%)
SAM + YOLOv8n	**99.13 ± 0.40**	**93.81 ± 1.48**	**96.80 ± 0.79**	95.89 ± 2.21	**99.57 ± 0.63**
SAM + YOLOv8m	98.98 ± 0.50	92.91 ± 2.16	96.31 ± 1.19	96.14 ± 2.30	99.35 ± 0.83
SAM + YOLOv8l	98.91 ± 0.52	92.36 ± 2.19	96.02 ± 1.21	95.95 ± 2.36	99.29 ± 0.84
SAM + YOLOv8s	99.03 ± 0.43	93.17 ± 2.02	96.45 ± 1.11	**96.15 ± 2.37**	99.42 ± 0.76
SAM + YOLOv8x	98.91 ± 0.54	92.35 ± 2.52	96.00 ± 1.40	95.25 ± 2.82	99.39 ± 0.89
SAM + YOLOv9c	98.78 ± 0.64	91.59 ± 3.31	95.58 ± 1.88	95.33 ± 3.01	99.27 ± 0.95
SAM + YOLOv9e	98.65 ± 0.81	90.55 ± 4.18	94.99 ± 2.43	94.55 ± 3.40	99.17 ± 1.23

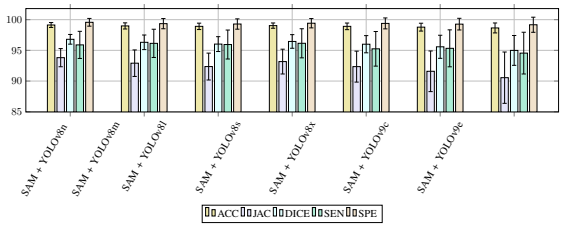

Fig. 3. Comparison graph of the segmentation comparison table of the models developed in this study.

As observed in Table I, the use of SAM for fine-tuning improved the segmentation model's performance, with an increase of 0.02% in accuracy, 0.18% in Jaccard index, 0.1% in Dice coefficient, 0.66% in sensitivity, and a decrease of 0.09% in specificity. Thus, there is a noticeable improvement in the mask for 4 out of the 5 adopted metrics, indicating enhancements in the ability to detect true positive pixels.

Figure 3 illustrates the results of different metrics for the proposed models using a bar graph

TABLE II
BARTLETT'S STATISTICAL TEST FOR SEGMENTATION RESULTS (SAM + YOLO).

	Metrics				
Comparison(S+V8n)	ACC	JAC	DICE	SEN	SPE
S+V8m	=	=	=	=	=
S+V8l	x	x	x	x	x
S+V8s	=	=	=	=	=
S+V8x	x	x	x	x	=
S+V9c	=	=	=	=	=
S+V9e	x	x	x	x	x

Table II presents the results of Bartlett's statistical test for the different Segmentation models developed. The best model had its result vectors, used to generate the mean and standard deviation in the table, compared with the result vectors of the other models for dataset test split. The table indicates statistical equality between the SAM+YOLOv8n model and the SAM+YOLOv8m, SAM+YOLOv8s, and SAM+YOLOv9c models. There is statistical difference for the SAM+YOLOv8l, SAM+YOLOv8x, and SAM+YOLOv9e models.

The lower performance of some of the heavier models, with more robust architecture compared to the lighter models, may be due to the smaller amount of images provided by the dataset used. Generally, more robust models like YOLOv8n require less data for good performance.

TABLE III
CLASSIFICATION RESULTS FOR THE COMBINATION OF FEATURE EXTRACTORS AND CLASSIFIERS ADDRESSED IN THIS STUDY

Feature Extractor	Classifier	Accuracy	Precision	Recall	F1-Score	Matthews	Train Time	Predict Time
LBP	KNN	85.0±1.0	62.0±4.0	52.0±1.0	50.0±1.0	9.0±3.0	0.05±0.01	0.01±0.01
	MLP	**86.0 ± 1.0**	**74.0 ± 23.0**	50.0±1.0	47.0±1.0	7.0±6.0	3.64±0.83	0.01±0.01
	NaiveBayes	77.0±2.0	59.0±2.0	**63.0 ± 2.0**	**60.0 ± 2.0**	**22.0 ± 4.0**	**0.01 ± 0.01**	**0.01 ± 0.01**
	RandomForest	**86.0 ± 1.0**	68.0±6.0	51.0±1.0	49.0±2.0	9.0±4.0	45.18±0.71	0.01±0.01
	SVM_Linear	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	0.0±0.0	0.64±0.12	0.01±0.01
	SVM_Polynomial	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	0.0±0.0	3.82±0.18	0.01±0.01
	SVM_RBF	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	0.0±0.0	1.94±0.15	0.02±0.01
DenseNet121	KNN	86.0±1.0	72.0±9.0	53.0±1.0	52.0±2.0	15.0±5.0	0.35±0.02	0.02±0.01
	MLP	83.0±1.0	62.0±3.0	57.0±1.0	58.0±1.0	19.0±4.0	11.86±2.16	0.01±0.01
	NaiveBayes	71.0±2.0	58.0±1.0	63.0±2.0	57.0±2.0	20.0±3.0	0.04±0.01	0.01 ± 0.01
	RandomForest	85.0±1.0	62.0±19.0	50.0±0.0	47.0±1.0	4.0±4.0	483.26±13.20	0.01±0.01
	SVM_Linear	84.0±1.0	64.0±3.0	57.0±2.0	58.0±2.0	19.0±5.0	399.82±85.72	0.05±0.01
	SVM_Polynomial	85.0±1.0	48.0±11.0	50.0±0.0	47.0±1.0	2.0±4.0	15.15±2.02	0.08±0.01
	SVM_RBF	85.0±1.0	51.0±11.0	51.0±1.0	47.0±2.0	3.0±6.0	29.26±2.99	0.30±0.13
VGG16	KNN	85.0±1.0	56.0±7.0	51.0±1.0	49.0±2.0	4.0±5.0	0.18±0.01	0.01±0.01
	MLP	81.0±2.0	60.0±3.0	57.0±3.0	58.0±3.0	17.0±6.0	26.86±2.59	0.01±0.01
	NaiveBayes	58.0±5.0	56.0±2.0	62.0±3.0	51.0±4.0	17.0±4.0	0.02±0.01	0.01±0.01
	RandomForest	86.0±1.0	65.0±22.0	50.0±0.0	47.0±1.0	5.0±6.0	200.23±5.17	0.01±0.01
	SVM_Linear	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	-0.0±1.0	17.18±1.62	0.03±0.01
	SVM_Polynomial	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	0.0±0.0	7.34±0.75	0.04±0.01
	SVM_RBF	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	0.0±0.0	8.90±0.30	0.07±0.01

B. Melanoma classification

Table III presents the classification results for the pipeline developed to classify melanoma as benign or malignant. The primary metric used to select the best combination was accuracy. The database used in the classification was the ISIC 2017 database.

As observed in Table III,the best combination obtained for classifying the ISIC 2017 dataset was using LBP as the feature extractor and MLP as the classifier. A maximum of 1000 iterations was used for the automatic adjustment of MLP's internal hyperparameters, which was excluded from the grid search applied to the other classifiers.

The accuracy obtained by the best combination, 86.0% $\pm$ 1.0%, indicates a satisfactory ability to correctly classify labels. The precision of 74.0% indicates a satisfactory result in the model's ability to classify true positives within the melanoma region, compared to other models.

The LBP as a feature extractor may have outperformed CNN extractors due to the texture variability among melanomas, as homogeneous or heterogeneous coloration in melanoma is indicative of its malignancy. MLP, overall, excels in its ability to handle non-linear problems. Coupled with the lower complexity of features extracted by LBP, which are significant for malignancy labeling, this combination resulted in good performance.

TABLE IV
CLASSIFICATION RESULTS OF THE OTHER LABELED DATA FROM THE ISIC 2017 DATASET FOR THE BEST PROPOSED EXTRACTION/CLASSIFICATION COMBINATION.

Feature Extractor	Classifier	Label	Accuracy
LBP	MLP	diagnosis_confirm_type	0.90 ± 0.01
		benign_malignant	86.0 ± 1.0
		clin_size_long_diam_mm	0.79 ± 0.03
		anatom_site_general	0.41 ± 0.02
		age_approx	0.12 ± 0.01

Table IV presents the average classification accuracies of the other classes in the ISIC 2017 dataset for melanoma images using the best extraction/classification combination, LBP + MLP. Overall, the pipeline was effective in classifying diagnosis confirmation type, malignancy, and melanoma diameter in millimeters.

For the lesion location on the patient's body and age classes, the system showed low accuracy for a simple reason: to improve the data extraction about melanoma, extraction and classification were performed on the segmentation of the melanoma in the dermoscopic image. These two classes, which had lower accuracies, require context from the image beyond the melanoma to be more accurate, such as the texture of the patient's skin.

C. Comparison with the state of the art

Table V presents the comparison with the state of the art for the proposed model and methods found in the literature that also used the PH2 dataset to segment melanoma. The segmentation metrics strategically compared were Accuracy, Dice coefficient, and Specificity, aiming to evaluate the models' performance in terms of pixel accuracy in segmentation, similarity to ground truth (GT), and rate of segmented true positive pixels.

TABLE V
COMPARISON WITH THE STATE OF THE ART FOR MELANOMA SEGMENTATION CONCERNING THE PH2 DATASET.

Methods	ACC(%)	DICE(%)	SPE(%)
Proposed Method(S+v8n)	**99.13 ± 0.40**	**96.80 ± 0.79**	**99.57 ± 0.63**
Deep class (2019) [8]	95.30	92.10	94.52
Geodesic (2019) [9]	94.59	92.17	97.99
FLog Parzen (2020) [10]	94.25	92.49	93.21

The proposed method achieved state-of-the-art performance in all three adopted metrics, with accuracy 3.83% higher (99.12% - 95.30%) compared to the second-best model, Deep Class [8]. The DICE coefficient of the proposed model was 4.75% higher (96.80% - 92.10%). Finally, specificity was 5.05% higher (99.57% - 94.52%).

Table VI presents the comparison with the state of the art for melanoma classification using the ISIC 2017 dataset. The proposed model achieved state-of-the-art performance in melanoma classification, with accuracy 4.71% higher (86.0% - 81.20%) than the method proposed by Al et al. [11], which used Inception-v3 for feature extraction and melanoma classification.

Figure 4 presents the graphical illustration of the results compared in Table VI.

TABLE VI
COMPARISON WITH THE STATE OF THE ART FOR MELANOMA CLASSIFICATION CONCERNING THE ISIC 2017 DATASET.

Author	Feature Extractor	Classifier	Accuracy(%)
Proposed	LBP	MLP	**86.0 ± 1.0**
al [11]	Inception-v3	Inception-v3	81.29 ± -
al [11]	ResNet-50	ResNet-50	81.57 ± -
al [11]	Inception-ResNet-v2	Inception-ResNet-v2	81.34 ± -
al [11]	DenseNet-201	DenseNet-201	73.44 ± -
yilmaz [12]	NASNetMobile - 16	NASNetMobile - 16	82.00 ± -
yilmaz [12]	MobileNetV2 - 16	MobileNetV2 - 16	81.45 ± -
yilmaz [12]	MobileNet - 32	MobileNet - 32	80.73 ± -
budhiman [23]	ResNet 50	ResNet 50	78.50 ± -
budhiman [23]	ResNet 25	ResNet 25	80.70 ± -
budhiman [23]	ResNet 7	ResNet 7	82.40 ± -

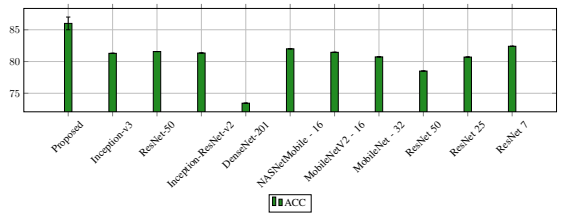

Fig. 4. Graphical comparison of the accuracy of proposed methods and state-of-the-art methods.

VI. CONCLUSION

This study developed an innovative approach that combines various aspects of artificial intelligence for the detection, segmentation, and classification of melanoma in dermoscopic images. Detection and segmentation, inspired by YOLO and SAM architecture models, achieved detection and segmentation accuracy above 99%, surpassing the state of the art. For classification, a pipeline of feature extraction and classification was created using grid search, transfer learning, and cross-validation to find the best extractor-classifier combination for the proposed approach, achieving a detection accuracy of 86% as its best result with LBP-MLP

The proposed model emerged from the experimental selection of a set of methods, varying models and methods for detection, segmentation, and classification. The best combination was compared with various studies found in the literature to validate the results obtained with the datasets approached, surpassing several studies found in the literature.

As future work, we aim to apply the proposed methodology to other datasets, including different datasets such as brain tumors and hemorrhagic stroke. We plan to propose new CNN models via transfer learning, with different classifiers, custom CNN architectures, and other LLM models for the textual pre-diagnostic prompt.

REFERENCES

[1] K. Saginala, A. Barsouk, J. S. Aluru, P. Rawla, and A. Barsouk, "Epidemiology of melanoma," *Medical sciences*, vol. 9, no. 4, p. 63, 2021.

[2] M. Arnold, D. Singh, M. Laversanne, J. Vignat, S. Vaccarella, F. Meheus, A. E. Cust, E. De Vries, D. C. Whiteman, and F. Bray, "Global burden of cutaneous melanoma in 2020 and projections to 2040," *JAMA dermatology*, vol. 158, no. 5, pp. 495–503, 2022.

[3] A. F. Duarte, B. Sousa-Pinto, L. F. Azevedo, A. M. Barros, S. Puig, J. Malvehy, E. Haneke, and O. Correia, "Clinical abcde rule for early melanoma detection," *European Journal of Dermatology*, vol. 31, no. 6, pp. 771–778, 2021.

[4] C. Ring, N. Cox, and J. B. Lee, "Dermatoscopy," *Clinics in Dermatology*, vol. 39, no. 4, pp. 635–642, 2021.

[5] H. Xiao, L. Li, Q. Liu, X. Zhu, and Q. Zhang, "Transformers in medical image segmentation: A review," *Biomedical Signal Processing and Control*, vol. 84, p. 104791, 2023.

[6] T. Mazhar, I. Haq, A. Ditta, S. A. H. Mohsan, F. Rehman, I. Zafar, J. A. Gansau, and L. P. W. Goh, "The role of machine learning and deep learning approaches for the detection of skin cancer," in *Healthcare*, vol. 11, no. 3. MDPI, 2023, p. 415.

[7] A. K. Tiwari, M. K. Mishra, A. R. Panda, and B. Panda, "Survey on computer-aided automated melanoma detection," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, no. 7, p. 2300257, 2024.

[8] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," *Pattern recognition*, vol. 85, pp. 78–89, 2019.

[9] F. F. X. Vasconcelos, A. G. Medeiros, S. A. Peixoto, and P. P. Reboucas Filho, "Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour," *Cognitive Systems Research*, vol. 55, pp. 44–59, 2019.

[10] E. de Souza Rebouças, F. N. S. de Medeiros, R. C. P. Marques, J. V. S. Chagas, M. T. Guimaraes, L. O. Santos, A. G. Medeiros, and S. A. Peixoto, "Level set approach based on parzen window and floor of log for edge computing object segmentation in digital images," *Applied Soft Computing*, vol. 105, p. 107273, 2021.

[11] M. A. Al-Masni, D.-H. Kim, and T.-S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Computer methods and programs in biomedicine*, vol. 190, p. 105351, 2020.

[12] A. Yilmaz, M. Kalebasi, Y. Samoylenko, M. E. Guvenilir, and H. Uvet, "Benchmarking of lightweight deep learning architectures for skin cancer classification using isic 2017 dataset," *arXiv preprint arXiv:2110.12270*, 2021.

[13] H. Wang, S. Zhao, Z. Qiang, N. Xi, B. Qin, and T. Liu, "Beyond direct diagnosis: Llm-based multi-specialist agent consultation for automatic diagnosis," *arXiv preprint arXiv:2401.16107*, 2024.

[14] T. Mendonça, M. Celebi, T. Mendonca, and J. Marques, "Ph2: A public database for the analysis of dermoscopic images," *Dermoscopy image analysis*, vol. 2, 2015.

[15] M. Berseth, "Isic 2017-skin lesion analysis towards melanoma detection," *arXiv preprint arXiv:1703.00523*, 2017.

[16] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia computer science*, vol. 199, pp. 1066–1073, 2022.

[17] M. Hussain, "Yolo-v1 to yolo-v8, the rise of yolo and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol. 11, no. 7, p. 677, 2023.

[18] R. Deng, C. Cui, Q. Liu, T. Yao, L. W. Remedios, S. Bao, B. A. Landman, L. E. Wheless, L. A. Coburn, K. T. Wilson *et al.*, "Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging," *arXiv preprint arXiv:2304.04155*, 2023.

[19] Y. Xu, M. A. Dos Santos, L. F. F. Souza, A. G. Marques, L. Zhang, J. J. da Costa Nascimento, V. H. C. de Albuquerque, and P. P. Rebouças Filho, "New fully automatic approach for tissue identification in histopathological examinations using transfer learning," *IET Image Processing*, vol. 16, no. 11, pp. 2875–2889, 2022.

[20] M. J. Nayeem, S. Rana, F. Alam, and M. A. Rahman, "Prediction of hepatitis disease using k-nearest neighbors, naive bayes, support vector machine, multi-layer perceptron and random forest," in *2021 international conference on information and communication technology for sustainable development (ICICT4SD)*. IEEE, 2021, pp. 280–284.

[21] J. Wang, Z. Yang, Z. Yao, and H. Yu, "Jmlr: Joint medical llm and retrieval training for enhancing reasoning and professional question answering capability," *arXiv preprint arXiv:2402.17887*, 2024.

[22] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, p. 210, 2022.

[23] A. Budhiman, S. Suyanto, and A. Arifianto, "Melanoma cancer classification using resnet with data augmentation," in *2019 international seminar on research of information technology and intelligent systems (ISRITI)*. IEEE, 2019, pp. 17–20.