Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.5903

A Blockchain-Integrated Pipeline Featuring Generative AI for Melanoma Detection, Segmentation, and Classification in Dermoscopic Images

Marcus Vinicius Candido de Figueiredo [**Federal Institute of Ceará** | ***mvcf10@gmail.com***]
Adriell Gomes Marques [**Federal Institute of Ceará** | ***adriell.gomes@alu.ufc.br***]
Francisco Italo Guilherme Da Silva [**Federal Institute of Ceará** | ***italoguilherme18.1@gmail.com***]
José Jerovane Da Costa Nascimento [**Federal Institute of Ceará**| ***jerovane@alu.ufc.br***]
Cidcley Teixeira de Souza [**Federal Institute of Ceará** | ***cidcley@ifce.edu.br***]
Luís Fabrício de Freitas Souza [**Federal University of Cariri**| ***fabricio.freitas@ufca.edu.br***]

Laboratory of Innovation in Artificial Intelligence Systems (LISIA), Federal University of Cariri (UFCA) - Av. Ten. Raimundo Rocha, 1639 - Cidade Universitária, Juazeiro do Norte / Fortaleza - CE, 63048-080, Brazil.

Received: 13 April 2025 • **Accepted:** 13 January 2026 • **Published:** 03 July 2026

Abstract. Melanoma is one of the most aggressive forms of skin cancer, responsible for thousands of deaths annually worldwide. The effectiveness of early diagnosis and continuous monitoring of disease progression are critical factors for therapeutic success and the improvement of public health outcomes. This study presents an advanced blockchain-integrated pipeline for the automated analysis of dermoscopic images, combining modern Computer Vision and Artificial Intelligence (AI) techniques to enhance the processes of lesion detection, segmentation, and classification. Using state-of-the-art architectures such as YOLO for identifying suspicious regions and the Segment Anything Model (SAM) for precise extraction of affected areas, the proposed method achieved an accuracy of 99% in the initial analysis stages, outperforming previously reported results in the literature. For the classification stage, a pipeline of feature extraction and selection is employed, enabling the identification of regions with a high probability of malignancy, achieving 86% accuracy. The study also incorporates Large Language Models (LLMs) to automatically generate interpretive reports, providing preliminary clinical decision support through generative AI. In the final stage, all generated data and results—including processed images, model outputs, and generated reports—are recorded on a blockchain network. This distributed storage ensures the integrity, traceability, and immutability of the information, enabling the creation of a secure and auditable clinical history. The blockchain-based solution adds transparency to the proposed method and enables controlled data sharing among medical institutions, reinforcing trust and accountability in the use of automated tools for dermatological diagnosis support.

Keywords: Dermoscopic Image of Melanoma, Artificial Intelligence, Large Language Models, Generative AI, Blockchain.

1 Introduction

Melanoma is a form of skin cancer that develops in the cells responsible for producing pigments, known as melanocytes, that is, caused by the development of a malignant tumor in these cells Patil [2022]. Data collected in 2020 regarding cancer estimate that there were about 325,000 new cases and 57,000 deaths due to melanoma Arnold *et al*. [2022]. Alarming figures from future projections indicate that by the year 2040, these numbers may increase, with approximately 510,000 new cases and up to 96,000 deaths caused by the cancer Arnold *et al*. [2022].

Melanoma generally progresses through two distinct phases: the first is characterized by the radial growth of the area, presenting as a flat, small-sized lesion with less aggressive behavior. The next phase, known as the vertical growth phase, involves malignant cells infiltrating deeply into the reticular dermis and potentially reaching the subcutaneous tissue, making the prognosis more difficult [Wainstein and Belfort, 2004].

There are also other signs and symptoms that should raise concern about this type of cancer, known to be the most dangerous form of skin cancer [Alquran *et al*., 2017]. Among them are characteristics that distinguish it from warts or moles, such as asymmetry in the lesion—where one half has a different shape or structure than the other—irregular borders that may be poorly defined or jagged, the presence of multiple shades in the same area, a size generally greater than 6 mm (although it can be smaller), and progressive changes in appearance such as growth, changes in shape, or color. These features align with the ABCDE protocol guidelines (Asymmetry, Border, Color, Diameter, and Evolution) [Duarte *et al*., 2021].

In order to obtain high-definition images that allow for the proper visualization of areas suspected of cancer-related lesions, a non-invasive examination method known as dermatoscopic examination was developed [Frange *et al*., 2009]. With this method, the specialist is able to analyze the area in detail, identifying patterns, shapes, sizes, and other pigmentation characteristics. It also enables the identification of other cutaneous (skin) pathologies [Ring *et al*., 2021].

The early identification of melanoma is extremely impor-

tant for the patient, since this type of skin cancer can progress rapidly and lead to death if not treated in time [Mhaske, 2013]. In this context, several technologies have been developed for the detection, segmentation, and classification of lesions. These technologies play a crucial role in the clinical pre-assessment process by speeding up the time required for diagnosis and providing additional support for medical analysis [Xiao *et al*., 2023].

Blockchain is a technology that has gained prominence due to the functioning of its architecture and operating modes. It is known as a decentralized and immutable ledger, allowing transactions to be carried out without the need for intermediaries [Zheng *et al*., 2017]. Its use is not limited to financial fields but is expanding into various areas, such as the Internet of Things (IoT), healthcare, and decentralized applications [Blaut *et al*., 2023]. In Brazil, there are systems already built on the blockchain network, such as the b-CPF, a technology developed by Dataprev, aimed at sharing the CPF database in the country [Marchisotti, 2025].

Given the challenges related to melanoma diagnosis, its high incidence and mortality rates according to authorities, as well as its characteristics and the existing technologies capable of supporting medical diagnosis through medical imaging, this study proposes the detection, segmentation, and classification of melanoma in dermoscopic images. This work introduces a new model capable of accurately detecting and segmenting the lesion area, as well as classifying its diameter using transfer learning. This approach provides a pre-diagnosis in text format, supported by generative AI and stores this diagnostic information in a blockchain system to ensure immutability, security, and data transparency. The use of blockchain enables auditable tracking of diagnostic information, ensuring that medical records are protected against unauthorized changes and can be securely shared among healthcare professionals. The study presents a method based on a novel multi-layered approach that detects, segments, and classifies, offering a pre-diagnosis through generative AI.

In this study, new fine-tuning methods based on digital image processing were applied to enhance the accuracy in detecting and classifying melanoma in dermoscopic images. These methods were compared with the SAM (Segment Anything Model), improving the efficiency of lesion segmentation. Additionally, new feature extraction models were added to further refine the results and increase the robustness of the system. As part of the methodology's evolution, an additional step of integration with blockchain technology was incorporated, ensuring greater security and integrity of the processed data. This implementation of blockchain enables the immutable tracking of diagnostic information, contributing to transparency and data protection throughout the entire medical process.

Thus, among the main contributions sought in this study, the following stand out: The study focuses on the integration of heterogeneous techniques in a comparative and modular pipeline, combining lesion detection using computer vision models (especially YOLOv8N), fine-tuning using the Segment Anything Model (SAM), and a proof-of-concept for interpretable and contextualized textual pre-diagnosis based on clinical reports augmented by LLMs, with guaranteed auditability and security through the incorporation of blockchain. This end-to-end architecture differs from previous work by offering not only technical performance gains through comparative study and comparison with the literature, but also traceability, transparency, and applicability based on the proof-of-concept developed for blockchain operability. In short, this study seeks to quantitatively and qualitatively understand the results obtained from the experiments performed, through discussions based on technical analysis of the results and comparison with the literature.

This study addresses different aspects:

- A transfer learning-based approach for the detection, segmentation, feature extraction, and classification of melanomas using data fusion.
- A generative AI model based on LLMs for assistance in melanoma pre-diagnosis.
- Detection, segmentation, and classification in dermoscopic images for medical diagnostic support using generative AI models.
- Blockchain, security, and auditable tracking of diagnostic information.

Next, the structure of this study is given by the sections of Related Works, Materials and Methods, Methodology, Results and discussion and Conclusion.

2 Related Work

Considering the clinical appeal previously contextualized, as well as the emergence of increasingly robust detection and segmentation models, but also the growing advancement of technologies that seek to simplify their protocols for human interaction. This section deals with the main studies found in the literature that served as motivation for the present proposed method.

Considering their in-depth knowledge in digital image processing, pennisi et al. [Pennisi *et al*., 2016] proposed in their study a fully automatic method for segmentation of skin lesions in dermoscopic images. The model was implemented aiming at the early detection of melanoma in this imaging modality, and consisted of an algorithm, called ASLM, which integrates steps to remove artifacts, such as hair and oil bubbles commonly present in this exam. In addition, the method also sought image equalization, followed by two parallel detection processes: one for detecting the skin in the image and another for segmenting the lesion, through Delaunay triangulation. The method was able to provide a binary segmentation mask by merging the images provided by these two detection steps, applying Hough transformation to circular contours to select the object and draw the contour. In their experiments, which used the PH2 segmentation base, the model proposed by the authors managed to obtain an accuracy of around 89.66%, proving the efficiency of the model. The use of parameter optimization in the preprocessing methods, making them adaptable, could increase the performance of the model, even for more hererogeneous cases.

Taking into account the recent innovations in deep neural networks for semantic segmentation, with increasingly efficient models, Badrinarayanan et al. [Badrinarayanan *et al*.,

2017] proposed the CNN model called SegNet, a fully convolutional encoder-decoder architecture that seeks to transform low-resolution representations into pixel-by-pixel predictions with high accuracy. The proposed CNN model was inspired by the VGG16 structure. However, unlike this, the proposed architecture does not have fully connected layers: SegNet uses the max-pooling indices performed in the encoder stage to perform upsampling in the decoder. In this way, the model eliminates the need for learned interpolation methods, thus optimizing memory usage during the inference process. Finally, the approach is trained end-to-end through batch normalization and stochastic gradient descent techniques. The goal is to allow a more efficient reconstruction of the edge details of the objects segmented by SegNet. The model was also tested with the PH2 base, obtaining 93.36% accuracy, which shows the efficiency of the model. It is worth mentioning that the use of some post-processing in addition to the adjustment in the architecture could increase the segmentation capacity of the network, without increasing the weight of the architecture for this.

Taking into account the challenges in the segmentation of skin lesions, Patino et al. [Patiño *et al*., 2018] proposed a new fully automatic method, based on superpixel merging, for the segmentation of lesions in dermoscopic images. The proposed approach consisted of the application of a supersegmentation process, performed through the SLIC algorithm, which groups pixels according to the color intensity and their spatial coordinates. The output result of the model consists of superpixels that follow the contours of the lesions in the image, approximately. Then, the superpixels are fused through the difference in the average of the RGB values. This fusion is optimized by a binary search, which determines the ideal threshold to separate the region affected by the lesion from the healthy skin. This process does not require post-processing steps. After the fusion, the segmentation mask is generated. The results obtained by the test on the PH2 basis consist of an accuracy of 90.39%, a reasonable result considering the simplicity in terms of processing the application, when compared to CNN models, for example. It is worth mentioning that the use of post-processing adjustment methods could be added to the method, bringing the possibility of a significant improvement in the final segmentation mask.

Al et al. [Al-Masni *et al*., 2018] proposed a method called Full Resolution Convolutional Network (FrCN), a segmentation method that seeks to eliminate the resolution losses caused by standard undersampling in convolutional network architectures. The architecture proposed by the authors seeks to preserve the full spatial resolution of each pixel during the segmentation process. During the learning process, FrCN deals directly with the specific and evident features for segmentation, dispensing with additional steps of artifact removal and contrast adjustment. Among the segmentation results with the PH2 base, the accuracy obtained was 95.08%, a high performance result. The method demonstrated to outperform established techniques, such as FCN, U-Net and SegNet, positioning itself as a promising strategy for systems to support the early diagnosis of melanoma.

Baghersalimi et al. [Baghersalimi *et al*., 2019] proposed a custom architecture called DermoNet. The model was densely connected across layers, aiming to combine high accuracy with computational efficiency. Dense blocks and skip connections between the encoder and decoder layers were implemented, enabling the extraction and reuse of multi-scale features with the aim of recovering fine details of melanoma contours in the image, in addition to attempting to minimize resolution losses resulting from the model's downsampling procedures. The method was also tested on the PH2 database, generating a Jaccard of 85.79%, a good result considering the metric. The results could have been better evaluated if metrics such as accuracy and sensitivity were included in the authors' testing methodology.

Taking into account the growing demand for fast and low-computational cost segmentation methods for applications in IoT systems and edge computing in the medical field, Rebouças et al. [de Souza Rebouças *et al*., 2021] proposed the method called FLog Parzen Level Set (FPLS), an approach that seeks to combine the robustness of the Parzen Window probability density method with the Floor of Log clustering operation to initialize and refine the segmentation result. The level set curve with parzen windowing is initialized with a parameterized force that delimits the region of interest in digital images in an iterative way, until finding what would be the optimal edge limits, and stationing the limits of the region of interest. Among the results with the PH2 segmentation base, the FPLS method obtained an accuracy of 94.25%. It is worth mentioning that a deeper study of the parameterization of Parzen windowing in relation to the PH2 database could increase the effectiveness of this method with the database.

Considering the increasing prevalence of skin cancer and the need for accessible diagnostics, Yilmaz et al. [Yilmaz *et al*., 2021] found themselves in the need to benchmark lightweight deep learning architectures for skin cancer classification using the ISIC 2017 challenge dataset. In the study, the following mobile models were evaluated: MobileNet, MobileNetV2, and NASNetMobile. The architectures were applied in transfer learning and fine-tuning approaches, with experiments involving different batch sizes for the attributes. The results indicated that models trained with smaller batch sizes presented better generalization capacity. NASNetMobile, with batch size 16, achieved the best results among the other models, with an accuracy of 82.00% and precision of 81.77%. This research highlights the viability of mobile models in medical applications, which, because they have fewer parameters and reduced training and inference times, can be dynamically applicable and be ideal for practical applications in portable devices and remote diagnostic systems.

Budhiman et al. [Budhiman *et al*., 2019] developed a skin cancer classification system using the ResNet architecture, trained with data augmentation techniques. The authors studied several network depth configurations and evaluated them: ResNet with 50, 40, 25, 10 and 7 layers. To identify the best structure that allowed the distinction between benign and malignant images, the ISIC dataset was used. The authors applied preprocessing strategies to the base, such as oversampling and undersampling, in addition to the selective use of dropout, to deal with class imbalance. The experiments revealed that the ResNet 50 architecture, trained with data augmentation but without dropout, presented the best results, achieving an accuracy of 83%, with an F1 Score of 0.43 for the malignant class and 0.87 for the benign class. This study

highlights the potential of using deep networks and augmentation techniques to improve the performance of melanoma classification systems, contributing to more accurate and early diagnoses.

Considering the critical importance of early diagnosis and monitoring of skin cancer, Marques et al. [Marques *et al*., 2024] presented an innovative approach, published at the 37th SIBGRAPI 2024 Conference. In this proposal, the authors combine computer vision and artificial intelligence methods, integrating architectures from the YOLO and SAM frameworks for lesion detection and segmentation, achieving an accuracy on the scale of up to 99% for the PH2 segmentation database. The proposed method also incorporates a classification phase, trained with the ISIC 2017 database, and based on a data extraction and fusion pipeline, obtaining melanoma presence classification accuracies of around 86%. Using generative LLM solutions, the method seeks to provide results with good levels of accuracy, in addition to offering textual pre-diagnoses based on the attributes obtained in the classification of melanoma regions. This pioneering approach demonstrates the potential of data fusion and generative artificial intelligence to improve the effectiveness of diagnostic support systems, seeking to contribute to the improvement of melanoma treatment and management protocols.

Bi et al. [Bi *et al*., 2019] developed an automatic skin lesion segmentation method based on class-specific deep learning with probability-based stepwise integration (DCL-PSI). The model aims to improve lesion segmentation by incorporating probabilistic steps, achieving an accuracy rate of 95.30%.

Vasconcelos et al. [Vasconcelos *et al*., 2019] proposed an automatic segmentation method using geodesic active contouring (MGAC) based on mathematical morphology, highlighting its low computational cost. In the PH2 dataset, the method achieved an accuracy of 94.59%, demonstrating its reliability in the segmentation of skin lesions. It is worth mentioning that, due to the approximate approach adopted in the differential equation, the model is sensitive to variations in areas with lighter lesions. Similarly, De et al. [de Souza Rebouças *et al*., 2021] proposed a method for segmenting skin lesions through digital image processing, which obtained an accuracy of 94.25%.

Al et al. [Al-Masni *et al*., 2020] developed a diagnostic aid framework. The authors used the ISIC dataset from different years (2016, 2017 and 2018) for training and obtaining results, achieving accuracy values above 80%. However, segmentation refinement could improve the area of the segmented lesion, potentially increasing the system's performance. Similarly, Yilmaz et al. [Yilmaz *et al*., 2021] developed an approach for melanoma classification in dermoscopic images, achieving an accuracy of 82.40%.

With advances in models and the potential for developing pre-diagnostic approaches based on LLM and knowledge of clinical flows, Wang et al. [Wang *et al*., 2024a] developed a query method called "multi-expert agent-derived query". This method aims to adaptively fuse probabilistic distributions from agents to potential pathologies, which requires fewer parameter updates and training time, and provides diagnostic hypotheses based on the presented symptoms.

Therefore, taking into account the conceptual and technological context addressed in this section, the present study seeks to implement a system to aid medical diagnosis through a pipeline for detecting, segmenting, and classifying melanoma in dermoscopic images. The application refines the pre-diagnosis with LLM processing contextualized with the pathology and the classification of clinical attributes. The application had its data transferred through a blockchain architecture, for applicability, integrity, and traceability tests. Among the methodological contributions of this study, the use of a comparative pipeline, optimized with the search for optimal hyperparameters, refinement of diagnosis with LLM, and study on the flow of data traffic along the lines of blockchain communication protocols stand out, opening up the possibility of broader discussions on the use for medical data.

In Aishwarya et al.[Aishwarya *et al*., 2025], YOLOSkin was presented, a unified framework designed to aid in the identification of different types of skin cancer, using simplified versions of the YOLO detectors. The research analyzed the performance of the YOLOv3tiny, YOLOv4tiny, YOLOv5s, YOLOv7tiny, and YOLOv8s architectures, using the ISIC dataset, which contains nine skin cancer categories. The models were implemented on the Nvidia Jetson Nano, aiming for fast processing on local devices. By combining the YOLOv5s and YOLOv8s models, taking into account their certainty levels, the average precision (mAP@0.5) jumped from 91.5% to 94.3%, while the accuracy increased from 89.6% to an impressive 97.87%. Although the system shows promise on embedded platforms, more in-depth testing in real medical environments and the inclusion of more dermatological databases could consolidate its practical use.

Also using YOLO architectures, Gul et al. Gül *et al*. [2025] presented YOLOSAMIC, a hybrid approach for automatic skin cancer segmentation, which combined YOLOv8 for lesion detection and SAM for precise edge segmentation. The study used a hybrid dataset composed of 3,463 public and 765 private images, incorporating various clinical scenarios, such as skin tone variations, hair presence, complex backgrounds, and varying lighting conditions at the time of capture. The model achieved Dice and Jaccard indices of 93.9% and 91.1% in the public dataset and 89.90% and 84.45% in the hybrid dataset, respectively. It is worth noting that despite the high performance, the authors pointed out that the set of more complex clinical cases still reduces mask overlap, which opens up space for research into expanding the model's results.

Aishwarya et al. Aishwarya *et al*. [2023] proposed a YOLO-based deep neural network for skin cancer diagnosis, capable of classifying nine different tumor classes. The study used a set of 4,389 images, augmented by data augmentation techniques, and divided into nine distinct categories. During the experiments, the YOLOv3 and YOLOv4 models achieved a mean precision (mAP) of 88.03% and 86.52%, respectively, while the mean accuracy, precision, recall, and F1-score metrics were 98.06%, 92.75%, 91%, and 92%. Although the method showed promising results, the authors highlighted that adapting it to real clinical scenarios requires further validation and optimization to address inter-patient variations and different image capture conditions.

Thapar et al. [Thapar *et al*., 2024] proposed a hybrid method that integrates Grasshopper algorithm optimization with deep learning techniques for skin lesion segmentation and melanoma classification. The authors applied the method

to dermoscopic images without data augmentation and reported satisfactory accuracy rates of 98%–89% across different test scenarios. The method outperformed competing approaches in the presented experiments.

Nagaoka et al. Nagaoka [2025] presented a three-step approach for skin cancer classification, combining lesion detection, semantic segmentation, and classification. The method used YOLOv8 for initial detection in HAM10000, followed by the Segment Anything Model (SAM) and MedSAM for edge segmentation in ISIC 2018, and finally CNNs such as InceptionV3, DenseNet, Xception, VGG16, EfficientNet, and CoCa for classification. The best results were obtained with the CoCa model, achieving an accuracy of 85.33%, outperforming the other architectures evaluated in conjunction with the research. Despite the competitive performance, the authors highlighted the need for broader clinical validations and integration with diverse medical databases to ensure robustness and generalizability.

Abedini et al. [Abedini, 2024] developed a three-step process for skin cancer classification, starting with lesion detection using YOLOv8, followed by segmentation using the Segment Anything Model (SAM) and MedSAM on the ISIC 2018 dataset, and finishing with CNNs for classification, including InceptionV3, DenseNet, Xception, VGG16, EfficientNetV2-B3, and CoCa. The method was trained on the HAM10000 (10,015 images) and ISIC 2018 (2,594 images) datasets, applying preprocessing techniques such as digital hair removal and data augmentation. In the experiments, MedSAM achieved better segmentation performance (78.55% IoU versus 70.03% for SAM). For classification, the CoCa model outperformed the other architectures, achieving an accuracy of 85.33%, precision of 84.66%, recall of 91.56%, and an F1-score of 87.97%. Despite the competitive results, the authors emphasized the need to validate the approach in real clinical scenarios to ensure greater generalizability and practical applicability.

The studies by Rezk et al. Rezk *et al*. [2023], Lyakhova et al. Lyakhova and Lyakhov [2024] and Naseri et al. present comprehensive and systematic reviews on the application of artificial intelligence in skin cancer pathology, with an emphasis on the diagnosis and prognosis of melanoma from dermatoscopic images. Taken together, these studies show that modern architectures such as deep learning, including CNNs, DenseNet, ViT, and multimodal ensembles, can achieve accuracy levels comparable to or exceeding that of dermatologists in databases such as ISIC and HAM10000 through testing. However, they point to important limitations for clinical translation, such as low data diversity (skin tones, pediatric populations), class imbalance, the use of small or poorly standardized datasets, the absence of external validation, and several other challenges regarding interpretability and robustness to changes in data distribution. These reviews seek to reinforce the need for more diverse, unified, and clinically validated expert databases, as well as explainable and computationally efficient models, for the safe incorporation of AI into dermatological practice.

The integration of distributed ledger technologies, such as blockchain, has gained prominence in several areas of healthcare, offering innovative solutions to persistent challenges. [Kotter *et al*., 2021].

A relevant study highlights the benefits and threats of blockchain as a disruptive innovation in the healthcare sector, exploring its implications for data security and interoperability [AbdelSalam, 2023].

Furthermore, recent research proposes an innovative solution that combines Ethereum-based smart contracts with language models to detect fraud in health insurance claims, demonstrating the effectiveness of this approach with a detection rate of up to 99% [Islayem *et al*., 2025].

These studies illustrate how blockchain, combined with emerging technologies, can transform data and process management in the healthcare sector, promoting greater security, transparency, and efficiency. [Kotter *et al*., 2021; AbdelSalam, 2023; Islayem *et al*., 2025].

The table 1 provides a comparative summary of the works discussed in the related works section, highlighting the main points of the studies, as well as the database used, methodology applied, main results, and suggestions for improving the results.

3 Materials and Methods

This section presents the materials, methods, evaluation metrics, and datasets used in the development of this study.

3.1 *Datasets*

To support the proposed approach, several dermatology-related datasets were investigated. Two of them stood out in the analyses and were selected for use in 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. It was developed for research and benchmarking purposes in the context of segmentation and classification algorithms in dermatology [Mendonça *et al*., 2015]. The dataset is divided into 80 images of common nevi, 80 of atypical nevi, and 40 of melanoma, all from patients with Fitzpatrick skin types II and III. Each image has a resolution of 768×560 pixels in 8-bit RGB format, providing sufficient detail for the evaluation of segmentation and classification algorithms. [Sachdeva, 2009].

The second dataset used in this study is ISIC 2017, which is part of the International Skin Imaging Collaboration (ISIC) Archive. It contains a total of 2,000 lesion images, 2,000 corresponding segmentation masks, as well as clinical metadata related to the patient and the lesion. This dataset was used in the 2017 ISIC challenge on melanoma detection, which aimed to develop tools capable of performing analyses for melanoma diagnosis. The images are available in varying resolutions, typically ranging from 1024×768 to 6748×4499 pixels, offering high granularity for both segmentation and classification tasks. [Berseth, 2017].

To ensure reproducibility and fair comparison, both PH2 and ISIC 2017 datasets were partitioned into independent subsets. For PH2, 70% of the images (140 samples) were used for training, 15% (30 samples) for validation, and 15% (30 samples) for testing. For ISIC 2017, 60% (1,200 images) were used for training, 20% (400 images) for validation, and

Table 1. Comparative summary of the works discussed in the related works section. The table aims to synthesize the main points of each article to facilitate comparison with the method proposed in this study.

Author(s)	Proposed architecture	Dataset used	Methodology	Main results	Suggested advancements
Rezk et al. [45]	–	–	Review on AI for skin cancer	Summarizes performance and gaps of DL methods	End-to-end secure pipeline (this work).
Lyakhova et al. [31]	–	–	Systematic review on AI methods	Broad overview, limited clinical generalization	Practical application in a real pipeline.
Naseri et al. [38]	–	–	Review of ML/DL for melanoma	High accuracy, few works on prognosis	Add explainability and blockchain.
Pennisi et al.	ASLM	PH2	Automatic lesion segmentation	Reasonable binary masks in dermoscopy	Integrate into the proposed full pipeline.
Patiño et al.	Superpixels	PH2	Superpixel-based segmentation	Better contours than classical methods	Couple with classification and reports.
Al-Masni et al.	FrCN	ISIC 2017, PH2	End-to-end CNN for segmentation	Good segmentation on multiple datasets	Combine with YOLO and blockchain.
Badrinarayanan et al.	SegNet	PH2	Generic encoder–decoder segmentation	Widely used segmentation *baseline*	Specialize and embed into a pipeline.
Baghersalimi et al.	DermoNet	ISIC 2016, PH2	Dense segmentation of lesions	Good metrics on dermoscopic data	Extend to a full end-to-end solution.
de Souza Rebouças et al.	FPLS	PH2 and others	Parzen-based Level Set segmentation	Stable and robust lesion contours	Use as refinement within the pipeline.
Marques et al.	YOLO + SAM	PH2, ISIC 2017	Detection–segmentation–classification pipeline	High accuracy, no LLM or blockchain	Extend with LLM and blockchain (this work).
Aishwarya et al. (YOLOSkin)	YOLOSkin	Public dataset (ISIC)	Detection on embedded hardware	Good accuracy on Jetson Nano	Evolve to a complete clinical pipeline.
Aishwarya et al. (YOLO)	YOLO-based diagnosis	Public dataset (ISIC)	YOLO detection/classification	Good malignant vs. benign performance	Add textual report and logging.
Abedini et al. (CoCa)	YOLOv8 + SAM + CNNs	HAM10000, ISIC 2018	Three-stage pipeline (D–S–C)	High metrics in all three stages	Include LLM and auditable blockchain.
Nagaoka et al.	Detection + segmentation	HAM10000, ISIC 2018	Integrated lesion analysis flow	Integration improves overall performance	Add SNOMED, LLM and blockchain.
Yilmaz et al.	Lightweight DL models	ISIC 2017	Benchmark of compact architectures	Good accuracy with low cost	Embed them in a traceable pipeline.
Budhiman et al.	ResNet + distillation	HAM10000, PH2	Binary melanoma classification	High melanoma vs. non-melanoma accuracy	Use as a module inside the pipeline.
Thapar et al.	Metaheuristic + DL	Public dermoscopic datasets	Hybrid segmentation–classification pipeline	Good results with heuristic optimization	Add LLM and immutable recording.
Gül et al. (YOLOSAMIC)	YOLOv8 + SAM-Box	3,463 public + 765 private	Joint detection + segmentation framework	Gains on large hybrid dataset	Extend with attributes, reports and blockchain.
Al et al.	CNN-based CAD support	ISIC 2016/17/18	CNN support system for diagnosis	Good accuracy on several ISIC sets	Complete with segmentation and blockchain.
Wang et al. (LLM agent)	Multi-specialist LLM agent	–	Agent-based clinical architecture	Shows potential of clinical LLM agents	Apply to melanoma + imaging pipeline.
Wang et al. (LLM+retrieval)	LLM with *retrieval*	–	Medical QA with knowledge retrieval	Improved clinical question answering	Generate reports from image-derived attributes.
Islayem et al.	LLM + blockchain (fraud)	–	Fraud detection in healthcare	Increases trust using blockchain	Migrate concept to image-based diagnosis.
Kötter et al.	Blockchain for images	–	Framework for medical images on-chain	Ensures integrity and traceability of images	Connect framework to melanoma pipeline.
Marchisotti et al.	B-CPF (citizen data)	–	Sensitive data in blockchain	Shows governance and interoperability	Apply principles to dermoscopic reports.
AbdelSalam et al.	–	–	Blockchain benefits/risks review	Analyzes pros and cons of blockchain	Concretize in a diagnostic prototype.
Zheng et al.	–	–	Review of blockchain architecture	Summarizes components and applications	Apply to dermoscopic telemedicine.

20% (400 images) for testing. All quantitative results reported in this paper refer exclusively to the held-out test split, which was not used in any phase of model training or hyperparameter tuning.

3.2 YOLO

YOLO (*You Only Look Once*) is a widely used approach for tasks where object detection in images or scenes is essential [Redmon *et al*., 2016]. It performs object detection in a precise and computationally efficient manner by framing the task as a single regression problem. Additionally, YOLO is capable of producing bounding boxes, which facilitates the understanding and localization of detected objects in the scene [Jiang *et al*., 2022].

This architecture operates based on a loss function, which evaluates how far the network's predictions are from the ground truth annotations in the dataset—i.e., it quantifies the error between YOLO's predictions and the actual values for each measured object [Hussain, 2023].

In this study, YOLO was employed as the detection method, due to its speed, effectiveness and analytical reliability, with a single-stage actuation architecture capable of delimiting the melanoma region with a bounding box.

The training images were resized to 640×640 pixels, following Ultralytics standards. The training protocol was configured with 300 iterations, mini-batches according to default settings, and the optimizer and learning rate also kept as in the original YOLO implementation, ensuring both efficiency and reproducibility.

3.3 Fine-tuning Methods

Fine-tuning methods applied to segmentation generally aim to improve the output result by enhancing the boundaries or positioning of the segmented region of interest. These methods can consist of image processing techniques based on pixel manipulation principles or more sophisticated methods such as specialized CNN segmentation models.

In this study, fine-tuning models are used to increase the effectiveness of the segmentation stage, applied through the creation and combinatorial comparison of digital image processing methods. Thus, greater possibilities and variability of results are expected for the different images, in order to select the best combination according to the applied database.

3.3.1 SAM

The *Segment Anything Model* (SAM) is a segmentation model developed by META AI in 2023. It stands out for its high capacity to identify objects of interest in images with excellent accuracy and efficiency, utilizing an architecture formed by multiple neural networks and specific image processing techniques [Deng *et al*., 2023].

3.3.2 Level Set Based on Parzen Windowing

In order to segment images, the *level set* method based on Parzen windowing combines non-parametric density estimation with boundary evolution to construct a probability density map. This methodology employs a level function $\phi(x)$ to represent the contour of a region of interest. The density evaluation through Parzen windowing allows the calculation of the probability of certain pixels belonging or not to this region, based on extracted properties such as intensity or texture, as well as the proximity relationships between pixels. The evolution or shrinkage of the region's boundaries based on the level function is guided by parameters such as the energy function, which includes data fidelity terms – obtained through density evaluation – and regularization terms, such as the contour's smoothness.

3.3.3 Region Growing

The region growing method is a segmentation technique based on initialization with one or more seed points, and from these, the region is expanded iteratively, incorporating neighboring pixels that satisfy a similarity criterion, such as intensity or color. In short, for each pixel p in the proximity of the region R, the algorithm evaluates if the difference between the characteristic of p and the mean (or median) of the pixels already belonging to R is smaller than a predefined threshold T. If the criterion is met, p is included in R. This iterative process continues until no further changes are possible by a specific set of iterations, resulting in the end of the iterative cycle and the segmentation of the object. The simplicity of the technique makes it quite intuitive, although it is sensitive to the choice of seeds, the selected threshold value, and the size of the pixel analysis kernel during iteration.

3.3.4 K-means Clustering

The K-means algorithm is an iterative clustering method that partitions a dataset into k clusters. The method minimizes the sum of squared distances between the points and the centroids of the clusters. Initially, k centroids are randomly selected. Then, each point x is assigned to the cluster whose centroid μ_i is closest in the attribute space, according to the Euclidean distance $\|x - \mu_i\|$. After the assignment, the centroids are recalculated as the mean of the points belonging to each cluster. This assignment and update are repeated until convergence, meaning when the centroids no longer change their position over a set number of iterations, achieving stability.

3.4 Extractors and Classifiers

This study explored different feature extraction methods and classification algorithms in order to identify the most effective combination of extractors and classifiers for the proposed model. The extractors analyzed were LBP (*Local Binary Patterns*), VGG (*Visual Geometry Group*), and *DenseNet* [Xu *et al*., 2022]. Among the classifiers evaluated, KNN (*K-Nearest Neighbors*), MLP (*Multilayer Perceptron*), *Naive Bayes*, *Random Forest*, and SVM (*Support Vector Machine*) [Nayeem *et al*., 2021] were considered.

Transfer learning, applied in conjunction with machine learning classifiers, is a methodology that usually yields satisfactory results across a variety of databases. Classical extractors often encounter limitations in extracting features from certain databases due to their high degree of complexity. Performance improvements are also limited by the simple application of well-known CNN methods. Thus, transfer learning, combined with machine learning methods, seeks to expand the possibility of adapting performance by customizing parameters based on the deep features generated by the network extraction steps.

3.4.1 Extractors

LBP (*Local Binary Patterns*) is a feature extractor that performs texture description using image processing. It compares each pixel in the image with all of its neighbors, encodes them in a binary pattern format, which is then converted into a decimal value representing the texture of the area being analyzed in the image [Xu *et al*., 2022].

VGG is a CNN (*Convolutional Neural Network*) architecture developed by the Visual Geometry Group at the University of Oxford. This architecture has several variations in its internal convolution layers. The VGG-16 model, for instance, has 16 layers, and is used for deep feature extraction from images, followed by their classification. The features

extracted from VGG-16 can be classified by learning methods through a *transfer learning* process [Xu *et al*., 2022].

The last extractor studied is another CNN architecture called DenseNet. It is characterized by the fact that its layers are intensely connected, meaning each layer receives input from all previous layers, passing information to each future layer. This processing method helps mitigate the vanishing gradient problem, improving the propagation of features for the given problem [Xu *et al*., 2022].

3.4.2 Classifiers

KNN (K-Nearest Neighbors) is a machine learning algorithm used for both classification and regression. In this work, it was used as a classification approach. Its processing works by identifying the K nearest neighbors of the sample whose class is to be determined. For classification, the sample is assigned the class based on the majority vote of its neighbors, while for regression, it would be the average of the neighbors [Nayeem *et al*., 2021].

MLP (*Multilayer Perceptron*) is a type of artificial neural network that is part of artificial intelligence (AI) designed to teach computers how to process data similarly to the human brain. Its network architecture consists of at least three layers of nodes: an input layer, one or more hidden layers, and an output layer. Each node in one layer is interconnected with all nodes in the next layer. MLP uses backpropagation for training, adjusting the weights of the connections based on the error between the network's prediction and the real value. MLPs are powerful and can model complex relationships, being used for classification, regression, and signal processing tasks [Singh and Sachan, 2014].

The next classifier discussed in this work is *Naive Bayes*. This machine learning method belongs to a family of classification algorithms based on Bayes' Theorem, which assumes that all features are statistically independent of each other, simplifying the probabilistic calculations and making *Naive Bayes* a fast and effective classifier for complex problems [Nayeem *et al*., 2021].

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve classification accuracy, enhancing predictive performance. The trees in the forest are trained with random subsets of data and features, and their predictions are combined by majority voting or averaging, for regression [Nayeem *et al*., 2021].

Lastly, the final classifier discussed in this study is SVM (*Support Vector Machine*). It is a supervised learning algorithm used for both classification and regression. Its data processing architecture involves finding support vectors by obtaining the best hyperplane that separates the data into different classes in the feature space. Some of the kernels used for mapping data include linear, polynomial, *radial basis function* (RBF), and sigmoid. SVM is effective in high-dimensionality problems and is known for its robustness and precision [Nayeem *et al*., 2021].

3.5 Grid Search

Grid Search is an optimization technique through exhaustive search of hyperparameters in machine learning models. In its application, a finite set of possible values is defined for each hyperparameter, and then the model goes through training and evaluation for each combination of parameter values, allowing the identification of the set that results in the best performance for the dataset conditions.

3.6 LLM

Large Language Models (LLMs) are deep neural networks trained on large volumes of textual data to perform Natural Language Processing (NLP) tasks including translation, text generation, question answering, and summarization [Topsakal and Akinci, 2023]. These models are based on sophisticated architectures and are designed to perform functions such as language understanding, interpretation, and synthesis. Words are represented as *tokens*, which combine to construct different linguistic contexts [Wang *et al*., 2024b].

In this work, we used the LLaMA 2-13B (Ollama) model, from Meta's LLaMA 2 family, an autoregressive Transformer with about 13 billion parameters, pre-trained on approximately 2 trillion public tokens until September 2022 and fine-tuned for dialogue through SFT and RLHF [AI, 2023]. The model was run locally, without *fine-tuning*, ensuring greater control over costs and privacy, and was contextualized to the clinical domain through *prompt engineering*. For this, the features extracted in Step 3 were inserted into a clinical *template* enriched with standardized descriptors (SNOMED CT, ICD-10) and prevalence statistics derived directly from the PH2 and ISIC 2017 sets, considering the class distribution. Choosing the 13B model variant balances capacity and computational requirements, making it viable for both lab and production scenarios, in compliance with the Llama 2 Community License.

Thus, the LLM acts as a generator of textual technical reports, which include standardized clinical descriptions and observations derived from the classified attributes. Furthermore, the model is capable of producing illustrative descriptions of the segmented images and highlighting potential clinical implications, acting as a communication layer between the automated system and the specialist accessing the results. This strategy, by employing only *prompt engineering* instead of *fine-tuning*, reduces the need for large volumes of annotated medical data and facilitates the adaptation of the system to different pathologies.

3.7 Implementation Details

The YOLO model was trained with resized inputs of 640×640 pixels, adopting the Ultralytics default protocol with 300 iterations, batch size, optimizer, and learning rate preserved to ensure comparability with the baseline.

The SAM module was triggered through the detection of homogeneous regions indicated by YOLO bounding boxes. In this configuration, SAM generated refined masks based on region prompts, which were concatenated with YOLO outputs via interpolation to better align pixel-level boundaries of melanoma lesions.

For feature space reduction, K-means clustering was applied with $k = 127$, thereby reducing grayscale levels. Region Growing was configured with a 5×5 kernel and an

adaptive threshold, defined as ± 40 grayscale levels around the initial seed, in order to maintain robustness while preserving lesion details.

3.8 Blockchain Implementation

For the implementation, a private Ethereum blockchain was simulated using Ganache [Bhosale *et al*., 2019]. The smart contracts were developed in Solidity and managed by the Truffle framework [Dhanvardini *et al*., 2023], and were deployed on the local network. The processed images were stored on IPFS [Kumar and Tripathi, 2019], and their cryptographic hashes were recorded on the blockchain, along with the classification results and generated reports. Interaction with the network was carried out through a web interface developed in React [Alickovic, 2023], integrated with the MetaMask wallet for transaction management and user authentication [Lee, 2023]. The computational performance required to handle a 10 MB image on the chosen blockchain essentially involves two main stages: an initial local processing phase and the subsequent distribution of data across the network. At the beginning of the process, the node itself divides the image into smaller chunks, generates hashes for each of them, and organizes this information. For a 10 MB file, this phase typically completes in about five seconds. From that point on, the primary determining factor is network bandwidth, which governs the sharing of these data with other participants in the system that request them. Consequently, both transfer and access times depend largely on the quality of the connection and the number of nodes available in the network to assist in replicating the image.

3.9 Evaluation Metrics

This section will address the metrics used to evaluate the results obtained by the models. Both segmentation metrics and classification metrics will be covered. It is worth noting that the segmentation metrics were selected based on the analysis of several studies found in the literature. Thus, the metrics seek to analyze the pipeline's output performance using both overlap and similarity criteria in segmentation, and imbalance in the proportion of false positives and false negatives in classification.

Segmentation Metrics

Below is a brief explanation of the metrics adopted in the evaluation of detection and segmentation methods:

Accuracy: Accuracy in segmentation measures the proportion of correctly classified pixels relative to the total number of pixels in the image [Müller *et al*., 2022].

Sensitivity: Also known as Recall, sensitivity indicates the model's ability to correctly identify the pixels belonging to the region of interest, among the total pixels identified [Müller *et al*., 2022].

This metric is crucial when it is important to minimize the omission of relevant pixels as belonging to the region of interest.

Specificity: Specificity is a metric that seeks to quantify the ability of a segmentation model to correctly identify pixels that do not belong to the region of interest, i.e., the background region [Müller *et al*., 2022].

It is especially useful when the cost of false positives is high for the application.

Dice Coefficient: The Dice coefficient is a measure that evaluates the overlap between the generated segmentation and the ground truth segmentation [Müller *et al*., 2022].

Values close to 1 indicate a high match between the binary masks [Müller *et al*., 2022].

Jaccard Index: Also known as the Jaccard Overlap Coefficient [Müller *et al*., 2022].

This index penalizes false positives and false negatives more severely compared to Dice, making it a stricter metric for evaluating overlap [Müller *et al*., 2022].

Classification Metrics

Below is a brief explanation of the metrics adopted for evaluating extraction/classification methods:

Accuracy: In classification, accuracy is the fraction of correctly classified examples out of the total classifications made by the model [Ferrer, 2022].

Despite its simplicity, it can be misleading in imbalanced datasets, often not providing a complete measure of a classification method's performance [Ferrer, 2022].

Precision: Precision measures the proportion of true positive classifications among all positive predictions made by the model for a given class [Ferrer, 2022].

This metric is important when the cost of false positives is high for the classification method [Ferrer, 2022].

Recall: Also known as the true positive rate.

It emphasizes the model's ability to capture positive cases in the dataset [Ferrer, 2022].

F1-Score: The F1-Score is the harmonic mean of precision and recall, balancing both metrics in its evaluation [Ferrer, 2022]. It is given by:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{1}$$

This metric is particularly useful for datasets with class imbalance, providing a robust measure of performance in such cases [Ferrer, 2022].

Matthews Correlation Coefficient (MCC):

The MCC is a robust metric that uses all values from the confusion matrix in its calculation [Ferrer, 2022]. As shown by:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{2}$$

it ranges from -1 (total error) to +1 (perfect classification), with 0 representing random performance for the model [Ferrer, 2022].

Bartlett's Statistical Test

The Bartlett test is a parametric test commonly applied to check whether the variances of a set of k groups are equal to each other. For the null hypothesis H_0, all populations are assumed to have equal variances [Phoon *et al*., 2003]. A

common way to calculate the statistic of this test is through the following formula:

$$\chi^2 = \frac{(N-k)\ln(S_p^2) - \sum_{i=1}^{k}(n_i-1)\ln(s_i^2)}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{N-k}\right)}, \quad (3)$$

where:

- k is the number of groups to be compared;
- n_i is the sample size of group i, and $N = \sum_{i=1}^{k} n_i$ is the total number of observations across all groups;
- s_i^2 is the sample variance of group i;
- S_p^2 is the pooled variance, defined by

$$S_p^2 = \frac{\sum_{i=1}^{k}(n_i-1)s_i^2}{N-k} \quad (4)$$

In this equation, the numerator compares the logarithm of the pooled variance with the sum of the logarithms of the individual variances of the datasets being tested, weighted by the degrees of freedom of each group. The denominator adjusts the statistic to account for the variability in the sample sizes of each group. If H_0 is true, the statistic χ^2 follows approximately a chi-square distribution with $k-1$ degrees of freedom. Thus, a p-value smaller than the significance level (e.g., 0.05, or 0.01 for more stringent tests) indicates that the variances of the groups are significantly different [Phoon *et al*., 2003].

4 Methodology

The methodology of the present study was divided into five stages:

- **Step 1 - Melanoma Detection**: Conducted through a comparative analysis of different YOLO models.
- **Step 2 - Segmentation with Fine-Tuning**: Using different fine-tuning models based on Image Processing (IP) and Convolutional Neural Networks (CNNs).
- **Step 3 - Clinical Attribute Classification**: Using a comparative Extraction/Classification pipeline.
- **Step 4 - Pre-diagnosis with LLM**: Using a model contextualized to the pathology and incorporating classification results.
- **Step 5 - Blockchain Structuring with Health of Things (HoT)**: Communication within a restricted system.

Figure 1 presents the complete flowchart of the methodology proposed in this study. The method is divided into five stages, which operate sequentially, with each stage processing the output of the previous one, culminating in a refined textual and visual pre-diagnosis of the dermoscopic image of the melanoma.

4.1 Step 1 - Melanoma Detection

Stage 1 in Figure 1 consists of the initial detection of melanoma using the YOLO framework. In this stage, a set of YOLO models is pre-trained on the PH2 dataset to perform melanoma segmentation on dermoscopic images. After training, input images from the test split are fed into the system, passing through the two-step YOLO detection process, undergoing YOLO's two-step detection process. Initially, the model architecture generates bounding boxes by dividing the image into quadrants and placing detection anchors in each quadrant. These anchors are used to predict the presence of melanoma through regression parameters and region proposal estimations. These proposals are then concatenated to form a single bounding box. Subsequently, a second layer of YOLO—typically a fully convolutional layer—constructs a binary map of the detected region within the bounding box by extracting features from the region and performing pixel-by-pixel classification. The output of YOLO is a binary detection mask, which functions as an "X-ray" of the melanoma region: a binary image where the melanoma area appears in white (255), and the rest of the image appears in black (0).

The approach involved a comparative analysis of the following versions of the YOLO framework: YOLOv8l, YOLOv8m, YOLOv8n, YOLOv8s, YOLOv8x, YOLOv9c, and YOLOv9e. All models were trained using the same dataset split, and their results were compared through similarity and overlap calculations against the ground truth. These qualitative measures were quantitatively assessed using the metrics: Accuracy, Sensitivity, Specificity, Dice Coefficient, and Jaccard Index.

4.2 Step 2 - Segmentation with Fine-Tuning

On Stage 2 of Figure 1, the binary detection masks generated by the different YOLO versions were refined through post-processing using fine-tuning techniques based on digital image processing and Convolutional Neural Networks (CNNs).

In general, several models were developed by combining, in an ordered manner, the outputs of the YOLO models with different segmentation methods: Level Set based on Parzen Windowing (Parzen); Level Set based on Parzen Windowing applied to images clustered using the K-means method (Parzen_Clustering); reconstruction of the binary mask through Region Growing followed by Level Set based on Parzen Windowing; and concatenation of the YOLO masks with the binary detection mask produced by the SAM framework.

Each YOLO model had its binary mask applied to each of these four fine-tuning variations, in order to modify the edges of the detection binary mask in different ways, aiming to make the segmentation as accurate and as closely aligned as possible with the ground truth, which was annotated by medical specialists.

The binary masks post-processed through fine-tuning are referred to as segmentation masks, and their similarities and overlaps are also measured using the following metrics: Accuracy, Sensitivity, Specificity, Dice, and Jaccard. In summary, the objective of applying different YOLO versions and various fine-tuning methods is to obtain the best combination, considering the prior comparison with the dataset used during training. This can potentially offer more adaptability to the pipeline, making it more robust to variations in the dataset's conditions, allowing for interchangeability between more robust YOLO and fine-tuning models and lighter, less

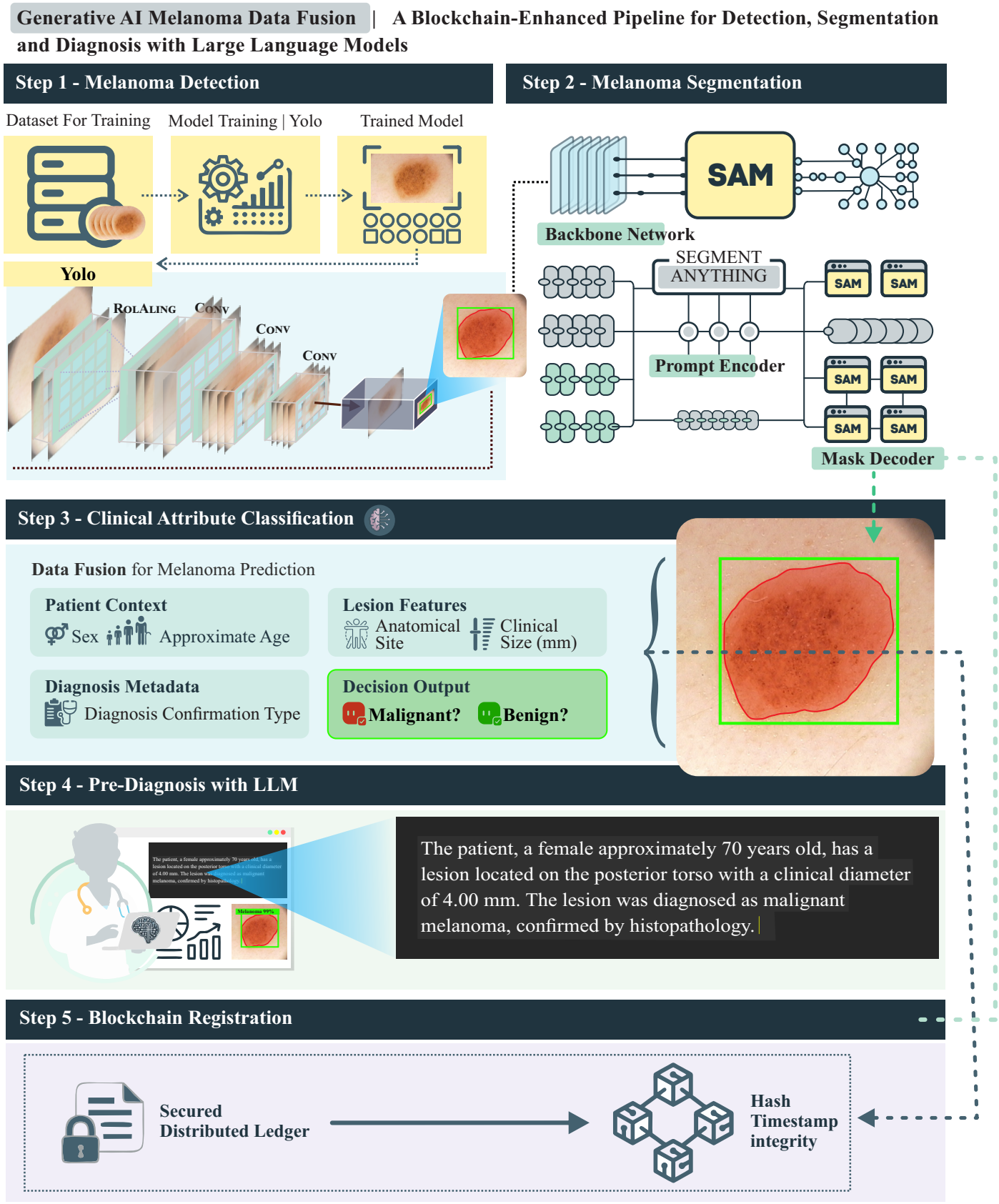

Figure 1. Proposed methodological flow for the automated analysis of dermatoscopic images focusing on melanoma detection and classification. The methodology consists of five sequential steps: (1) lesion detection with YOLO models trained on the PH2 dataset; (2) segmentation of the detected region with fine-tuning methods based on digital image processing (PDI) and neural networks, including the SAM model; (3) clinical classification with fusion of structured attributes (gender, age, location, diameter, confirmation type, and malignancy) extracted from the ISIC 2017 dataset; (4) generation of pre-diagnosis in natural language with trained LLMs; and (5) distributed and secure registration of data and reports on blockchain, ensuring integrity, traceability, and medical interoperability.

computationally intensive versions.

4.3 Step 3 - Clinical Attribute Classification

In step 3 of Figure 1, the new input image, which has already had its melanoma region segmented in step 2, undergoes a process of feature extraction using different extractors, followed by classification with a set of machine learning classifiers. In summary, this step involves building different extraction/classification models to select the best combination through metric-based comparison. The use of this model redundancy can also improve the adaptability of the classification system, especially when considering real-world applications subject to clinical image variability. From a research perspective, it is also valuable to compare the effectiveness of different extraction/classification models across the various cases present in the dataset.

The set of feature extractors used in this version of the experiment included: DenseNet121, DenseNet169, DenseNet201, LBP, VGG16, and VGG19. The set of classifiers used included: KNN, MLP, NaiveBayes, RandomForest, SVM_Linear, SVM_Polynomial, and SVM_RBF.

The interaction between the CNN extractors and the machine learning classifiers occurs through transfer learning, where the deep features generated by the convolutional layers of DenseNet and VGG are transferred to the machine learning classifiers. Grid Search was applied to determine the best hyperparameters for the classifiers based on the deep and conventional features (extracted by LBP). The optimal search ranges used in the Grid Search can be found in Table 2.

Table 2 presents the range values applied to each classifier through Grid Search, aiming to find the optimal hyperparameters for each classifier according to the characteristics of the features generated by each extraction model.

Table 2. Range table of hyperparameters set for each classifier. The ranges were used to obtain the optimal hyperparameters through Grid Search.

Classifier	Parameters of Search	Intervals of Search
Naive Bayes	-	-
MLP	Hidden layers	[2, 1001]
Nearest Neighbors	Number of Neighbors	1, 3, 5, 7, 9, 11, 13, 15
Random Forest	Maximum depth Bootstrap Criterion	6, Unlimited True, False Gini, Entropy
SVM Linear	Regularization parameter (C)	$2^x \vert x \in [-5, 15]$
SVM Polynomial	Degree Regularization parameter (C)	3, 5, 7, 9 $2^x \vert x \in [-5, 15]$
SVM RBF	Gamma Regularization parameter (C)	$2^x \vert x \in [-15, 3]$ $2^x \vert x \in [-5, 15]$

The classes obtained in the classification stage of the ISIC 2017 dataset are: "diagnosis_confirm_type", which refers to the method used to confirm the lesion diagnosis and may include histopathological analysis, follow-up, consensus, among others; the class "clin_long_diameter_mm", which refers to the melanoma diameter in the image, measured in millimeters; the class "anatom_site_general", which refers to the general anatomical location of the lesion on the body surface, such as head/neck, upper extremity, lower extremity, and trunk; and finally, the class "age_approx", which refers to the patient's estimated age.

4.4 Step 4 - Pre-diagnosis with LLM

In step 4 of Figure 1, the output result from step 3 (feature extraction/classification) is processed as input to an LLM model contextualized with clinical statistics and pathologies. The model generates a textual technical report based on the classified attributes, as well as an illustration of the segmented image, and includes the possibility of providing insights based on classification values. Thus, the LLM acts as a refinement in the communication between the system and the accessing specialist.

4.5 Step 5 - Blockchain Structuring with Health of Things (HoT)

In step 5 of Figure 1, blockchain is incorporated into the system architecture through a locally simulated environment. For this purpose, Ganache was used—a tool that simulates an Ethereum blockchain directly in the local development environment, allowing for testing and execution of smart contracts without financial costs or dependency on public networks.

This additional step was implemented to test a more secure data flow for the transmission of sensitive information. The data flow was transferred through the communication protocol, infrastructure, and blockchain-based data transfer to ensure traceability and integrity after the transition.

The environment setup began with the installation of the Ganache GUI (graphical interface), which provides an intuitive view of blocks, transactions, accounts, and real-time events. Upon starting the application, the local blockchain network is automatically created at `http://127.0.0.1:7545`, with ten pre-configured accounts, each with a fictitious ETH balance simulating a real test environment.

To ensure data persistence and the possibility of decentralized file sharing, IPFS (InterPlanetary File System) was integrated into the project. IPFS was configured using the IPFS Desktop graphical interface, which allowed the local daemon to run on port `5001` and enabled CORS support. This configuration was essential for the React front-end to interact with the IPFS network, allowing for efficient and secure file uploads and retrievals.

Additionally, it was necessary to configure the MetaMask wallet in the browser. MetaMask enables users to interact with the local blockchain, serving as a bridge between the web interface and the Ethereum network. The Ganache local network was manually added to MetaMask, and the first account generated in Ganache was imported using the corresponding private key, allowing transactions and smart contract function calls to be made directly from the browser.

The project structure was based on an integration between the Truffle framework and a React application. Truffle was used to compile, migrate, and test smart contracts written in Solidity. After cloning the project repository, environment dependencies were installed in both the main directory and

the *client* directory, which contains the front-end application. Migrations were executed with the command `truffle migrate --reset`, which deployed the contracts to the simulated Ganache network.

Finally, the React application was launched using the `npm start` command, enabling full interaction with the smart contracts. The interface offers users the ability to register and query data, store files via IPFS, and verify transactions executed on the local blockchain. This integration between React, Truffle, Ganache, IPFS, and MetaMask provided a complete, modular, and realistic environment for developing and validating the proposed model.

4.6 Methodological Flowchart

The Figure 2 illustrates the five main stages of the proposed methodology, organized hierarchically to represent the complete pipeline — from data acquisition to decentralized report storage — integrating hybrid computer vision, machine learning, and generative AI techniques.

In **Stage 1 — Data Collection and Preprocessing**, dermoscopic images are obtained from the PH2 dataset, which includes high-resolution lesion samples with clinical metadata. Images are standardized through a resizing process to ensure compatibility with the detection models. This dataset serves as the training and inference base for the system.

Stage 2 — Suspicious Region Detection using YOLO involves the application of multiple variants of the YOLO (You Only Look Once) family — YOLOv3, YOLOv5, YOLOv8L, YOLOv8M, YOLOv8X, and YOLOv9e. Each model detects the region of interest and generates a binary lesion mask, where lesion pixels are labeled as 255 and the background as 0.

Model performance is quantitatively evaluated using Precision, Recall, F1-score, and Intersection over Union (IoU). The optimal YOLO model is then selected for the subsequent segmentation stage. A detection re-evaluation loop is activated whenever segmentation metrics indicate suboptimal performance, allowing iterative model refinement.

In **Stage 3 — Advanced Lesion Segmentation**, the regions detected by YOLO are refined using both classical and modern segmentation approaches. Algorithms such as Region Growing, Level Set, K-means + Level Set, and Parzen Window are combined with the Segment Anything Model (SAM) to integrate parametric and prompt-based segmentation.

The resulting segmentations are compared in the Mask Comparison module, evaluated through quantitative metrics — Accuracy, Dice Coefficient, Jaccard Index, Specificity, and Sensitivity.

The best-performing segmentation mask is forwarded to the feature extraction module. When the segmentation quality falls below the defined threshold, a feedback loop is triggered to reapply or fine-tune YOLO, enabling continuous optimization of the detection-segmentation process.

Stage 4 — Feature Extraction and Classification analyzes the segmented lesion in detail. Morphological and textural features are extracted using Local Binary Patterns (LBP) and Convolutional Neural Networks (CNNs). These features are fed into a supervised Machine Learning (ML) classification pipeline comprising K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Naive Bayes, Random Forest, and Support Vector Machines (SVM) with linear, polynomial, and RBF kernels.

This stage outputs structured clinical and morphological descriptors such as patient gender, age, anatomical location, clinical diameter, and lesion type (benign or malignant). A secondary feedback connection allows the refinement of segmentation parameters based on inconsistencies detected in the classification results.

Finally, **Stage 5 — Report Generation and Storage with Blockchain** integrates Large Language Models (LLMs) and decentralized technologies to ensure interpretability and data integrity.

Using the extracted features, an LLM (LLaMA 2–13B) automatically generates a structured clinical report through prompts standardized by medical terminologies (SNOMED CT and ICD-10). The generated report is then stored in a decentralized IPFS (InterPlanetary File System) environment, with its cryptographic hash registered on the Ethereum Blockchain via smart contracts.

The decentralized interface is developed with React and MetaMask, enabling secure verification and traceability of medical records. The process culminates in a Standardized Clinical Report Registered on the Blockchain, ensuring transparency, auditability, and compliance with data security and authenticity standards.

5 Results and Discussion

This section presents the results obtained in the detection, segmentation, classification, and blockchain configuration stages. The evaluation metrics used for detection and segmentation in this study were: Accuracy, Sensitivity, Specificity, Dice Coefficient, and Jaccard Index. The classification evaluation metrics included: Accuracy, Precision, Recall, F1-Score, and Matthews Correlation Coefficient [Müller *et al*., 2022]. Additionally, Bartlett's statistical test was applied to the segmentation results.

Figure 3 presents two examples of output from the application throughout the stages of the proposed method, using two different input images. In the detection stage, the output generated by the YOLO framework consists of a binary detection mask along with a bounding box. In the segmentation stage, the output is a binary segmentation mask, which represents an enhanced version of the detection mask refined using fine-tuned models. The classification stage determines whether the melanoma is benign or malignant, and additionally classifies attributes such as lesion size, location on the skin, and patient age. Finally, the output of the pre-diagnosis stage consists of a technical textual report generated by a Large Language Model (LLM), enhanced with domain-specific knowledge related to the pathology and the classified attributes.

5.1 Melanoma Detection

This section presents and discusses the results obtained in the detection step of the proposed methodology. In this step, different models from the YOLO framework are compared

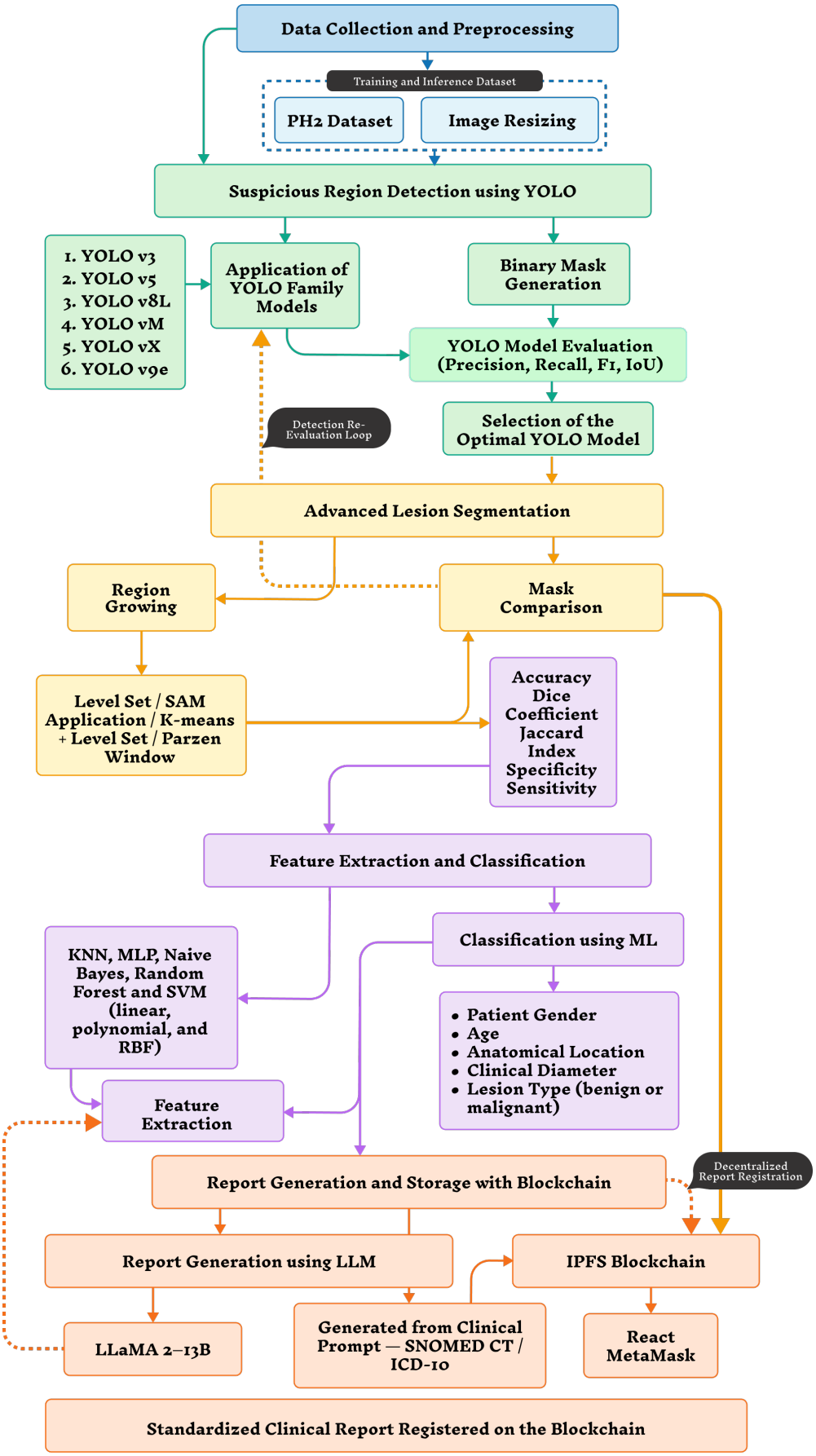

Figure 2. Detailed flowchart of the proposed methodology for automated dermoscopic image analysis using Artificial Intelligence, Large Language Models, and Blockchain.

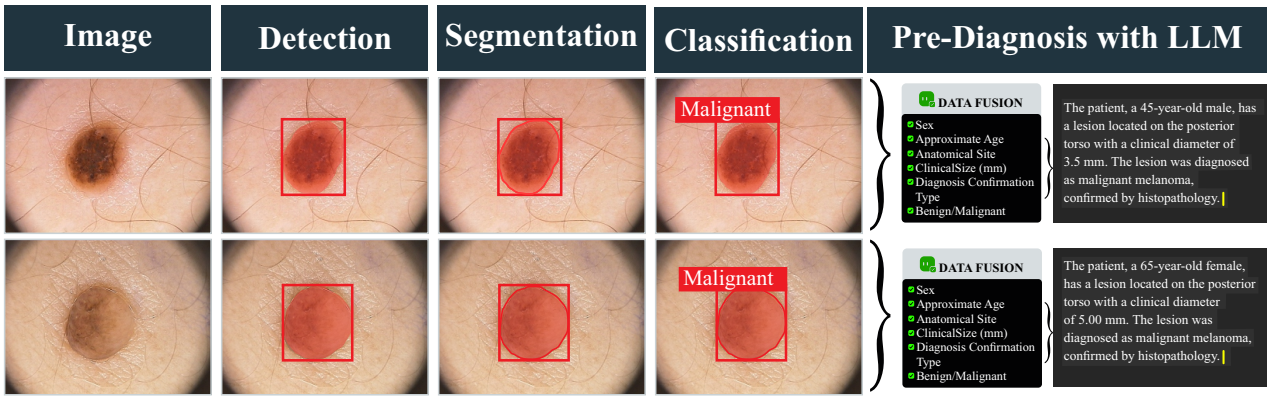

Figure 3. Results of the different stages of the proposed methodology for two distinct input images. The result of the detection stage consists of the demarcation of the melanoma region with a bounding box and a binary detection mask. During segmentation, the edges of the binary mask are refined through the enhancement of the binary mask via fine-tuning methods, generating an enhanced segmentation binary mask. Classification refers to the detection of malignancy and other medical attributes present in the dataset. The pre-diagnosis with LLM generates a technical written report based on the information obtained from classification and enhancement with medical and pathological context.

Table 3. Metric results for melanoma detection in the PH2 Dataset with different versions of the YOLO framework.

Detection Model	Accuracy(%)	Sensitivity(%)	Specificity(%)	Dice(%)	Jaccard(%)
YOLOv8L	98.40 ± 2.75	91.37 ± 18.82	99.33 ± 0.95	92.33 ± 18.87	88.95 ± 18.23
YOLOv8M	99.02 ± 0.45	95.52 ± 2.55	99.51 ± 0.68	96.46 ± 0.77	93.18 ± 1.45
YOLOv8N	**99.11 ± 0.42**	95.23 ± 2.41	**99.66 ± 0.53**	**96.70 ± 0.89**	**93.63 ± 1.68**
YOLOv8S	99.08 ± 0.40	**95.50 ± 2.62**	99.57 ± 0.59	96.59 ± 0.79	93.43 ± 1.48
YOLOv8X	98.42 ± 2.50	90.60 ± 18.74	99.46 ± 0.82	92.17 ± 18.84	88.66 ± 18.20
YOLOv9c	98.82 ± 0.51	94.66 ± 3.27	99.42 ± 0.74	95.67 ± 1.19	91.73 ± 2.19
YOLOv9e	**98.07 ± 3.06**	**89.94 ± 18.71**	**99.13 ± 1.46**	**91.26 ± 18.71**	**87.01 ± 18.03**

using the metrics Accuracy, Sensitivity, Specificity, Dice, and Jaccard.

Table 3 presents the results of the respective metrics for each model. Regarding the similarity metrics: Accuracies above 98% indicate, in a general way, the potential of the models after proper training; there was a good variation in sensitivity, with the lowest value obtained being 89% by the YOLOv9e model, which is a strong indicator of positive melanoma detection in the image, considering the obvious context of the dataset with dermatoscopic images; Finally, the specificity was generally good, with no values below 99%, which suggests that all models managed to easily differentiate the pixels of the background region.

Regarding the overlap metrics: The lowest DICE value, obtained by the YOLOv9e model, was 91.26%, with the highest value being from the YOLOv8N model, which was 96.70%. This indicates the good ability of the models to overlap the detected region of interest with the annotation from the medical expert. The Jaccard index, which measures the intersection between pixel sets, also showed good variation, with the best value being 93.63%, from the YOLOv8N model.

Overall, the YOLO models showed stable performance on the PH2 dataset, accurately delineating melanoma borders, as evidenced by the high accuracy, specificity, and sensitivity values obtained across various configurations. However, some variation in contour positioning was observed, reflected in the fluctuations in the Jaccard and Dice coefficients. This effect is possibly linked to the inherent functionality of the YOLO architecture, in which the bounding box generation phase limits the internal region later used to form the binary mask. Consequently, small spatial misalignments may occur at the lesion borders, especially at the corners of the region of interest, causing slight differences between the predicted masks and the GT annotations.

The YOLOv8N model achieved the best detection results, exhibiting the highest accuracy, specificity, Dice, and Jaccard values. This outcome indicates that lightweight configurations can achieve high discriminative performance even when compared with more complex architectures such as YOLOv8X and YOLOv9. This behavior may be associated with the optimization strategy adopted in the N model, which prioritizes essential features detection and reduces sensitivity to redundant variations observed in deeper versions. Given that the N variant was designed for low-latency applications, this efficiency-oriented configuration plausibly contributed to its superior generalization performance.

Graph 4 presents the visual difference between the metrics of the YOLO models. It is noticeable that the YOLOv8N shows greater stability, with smaller deviations compared to the other models.

5.2 Melanoma Segmentation

This section presents and discusses the results obtained in the Segmentation step of the proposed methodology. In this stage, different fine-tuning models applied to the YOLO framework results are compared using the metrics Accuracy, Sensitivity, Specificity, Dice, and Jaccard. The Bartlett test is applied to the results to compare the statistical equivalence of the models.

Table 4 presents the metric results for the segmentation models developed in this study. The binary segmentation

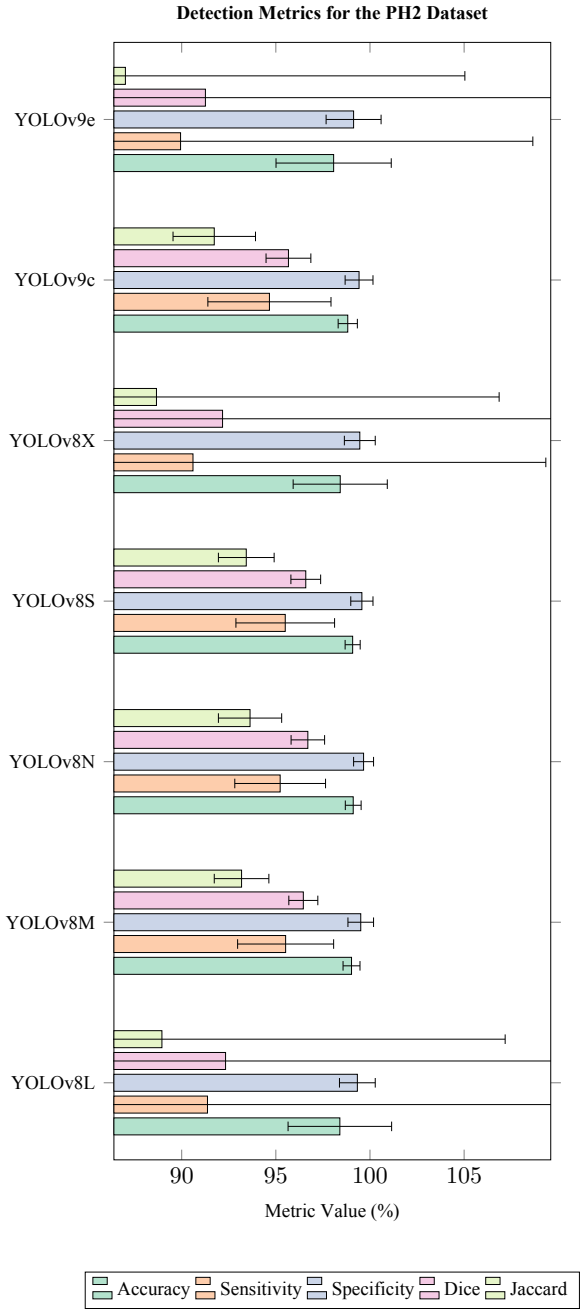

Figure 4. Comparison of the metrics presented in table 3.

mask resulting from the YOLO output goes through a fine-tuning process using different methods based on image processing and even CNN.

In general, the accuracies worsened with the segmentation models based on image processing, whether Parzen, Parzen_Clustering, or Parzen_Region_Growing. The result showed a decrease of about 2% to 1% compared to the metric value obtained with the binary detection mask. The standard deviation of the models also reflects this variation. Overall, the models with SAM as fine-tuning had a lower deviation by about 2% to 1%.

However, for the fine-tuning based on the concatenation of YOLO mask detections and SAM, the results generally showed a satisfactory improvement, especially for the sensitivity metric. A small improvement was also observed in accuracy, dice, and Jaccard, with a slight loss in the specificity metric. But overall, the proportion between the models mainly fluctuated with the YOLO mask, maintaining the superiority of the YOLOv8N-based models, especially for the segmentation method aimed at SAM, achieving the highest metrics of: Accuracy, with 99.13%, Specificity, with 99.57%, Dice, with 96.80%, and Jaccard, with 93.81%.

The segmentation models based on Parzen with region growing, in general, seem to have worsened their results, with the worst sensitivity being from the YOLOv8X + Parzen_Region_Growing model, at a value of 94.06, and the worst specificity being from the YOLOv9c + Parzen_Region_Growing model. This phenomenon could probably have been improved by obtaining an adaptive threshold value for region growing and dynamic kernel size control. Both mechanisms were not implemented in this study, with these parameters being selected through empirical testing for fine-tuning the models.

The fine-tuning models based on clustering also seem to not have had the expected effect on the image, aiming to homogenize the grayscale levels to facilitate segmentation. The worst accuracy in the comparison was from the YOLOv8S + Parzen_Clustering model, with a value of 96.83%. The same model also obtained the worst DICE and Jaccard, with 90.91% and 84.58%, respectively. What could be done to improve the applicability of clustering would be the use of a dynamic method to select the ideal cluster value for the grayscale level; this value was selected through empirical study based on the grayscale levels of the images. An adaptive value could assist the method in achieving the ideal fine-tuning effect.

Considering that the best fine-tuning methods for each YOLO model were those with SAM, the Bartlett statistical test was performed on the metric distribution of these models to compare the homogeneity of variances between the metrics of the fine-tuning models with SAM.

Thus, table 5 presents the result of the Bartlett statistical test for a significance level of 1%, to estimate the statistical equivalence of variance between the metrics obtained from the best model, YOLOv8N + SAM, and the other SAM-based fine-tuning models. The ”=” indicates that the calculated p-value reached the variance threshold, thus confirming the equivalence of variances. The ”x” indicates that the calculated p-value did not reach significance, rejecting the equivalence of the variances.

For the best model, that is, YOLOv8N + SAM, the metrics showed statistical equivalence for the models YOLOv8M + SAM, YOLOv8S + SAM, and YOLOv9C + SAM. This indicates parity between the models for the PH2 dataset addressed. An interesting factor is that these models have some degree of optimization for real-time applications, as the variations between the ”S”, ”M”, and ”C” versions generally reflect adjustments in the depth and width of the network, further balancing computational efficiency with performance, as the application requires.

The graph in Figure 5 presents the visual comparison of the fine-tuning methods for the YOLOv8N, YOLOv8S, and YOLOv8X models from table 4.

5.3 Melanoma Classification

This section presents and discusses the results obtained in the Classification stage of the proposed methodology. In this stage, different extraction/classification models are compared

Table 4. Results of YOLO + Fine-Tuning Segmentation for melanoma detection in the PH2 Dataset.

Detection Model	Fine-Tuning	Accuracy(%)	Sensitivity(%)	Specificity(%)	Dice(%)	Jaccard(%)
YOLO + Fine-Tuning Segmentation Results						
YOLOv8L	Parzen	97.63 ± 1.21	96.36 ± 2.99	97.76 ± 1.86	93.00 ± 4.10	87.55 ± 7.01
	Parzen_Clustering	97.71 ± 1.36	97.12 ± 2.62	97.42 ± 1.98	92.68 ± 4.63	87.01 ± 7.83
	Parzen_Region_G	97.87 ± 1.06	95.50 ± 3.55	97.65 ± 1.69	93.58 ± 3.59	88.10 ± 6.26
	SAM	98.91 ± 0.52	95.95 ± 2.36	99.29 ± 0.84	96.02 ± 1.21	92.36 ± 2.19
YOLOv8M	Parzen	97.52 ± 1.66	95.90 ± 3.88	97.69 ± 2.35	93.48 ± 4.28	88.00 ± 7.27
	Parzen_Clustering	97.51 ± 1.78	96.82 ± 3.28	97.57 ± 2.46	92.99 ± 4.78	87.62 ± 7.99
	Parzen_Region_G	97.92 ± 1.50	94.62 ± 4.75	97.70 ± 2.16	93.54 ± 3.96	88.47 ± 6.84
	SAM	98.98 ± 0.50	96.14 ± 2.30	99.35 ± 0.83	96.31 ± 1.19	92.91 ± 2.16
YOLOv8N	Parzen	97.53 ± 1.28	97.93 ± 1.90	97.42 ± 2.04	93.58 ± 3.76	88.33 ± 6.38
	Parzen_Clustering	97.63 ± 1.42	98.15 ± 1.41	97.39 ± 2.21	92.96 ± 3.99	87.48 ± 6.71
	Parzen_Region_G	98.12 ± 1.14	97.00 ± 2.44	97.67 ± 1.89	93.95 ± 3.29	89.54 ± 5.68
	SAM	**99.13 ± 0.40**	95.89 ± 2.21	**99.57 ± 0.63**	**96.80 ± 0.79**	**93.81 ± 1.48**
YOLOv8S	Parzen	97.17 ± 1.76	98.28 ± 0.76	96.91 ± 2.62	91.60 ± 5.27	85.64 ± 8.61
	Parzen_Clustering	**96.83 ± 1.86**	**98.95 ± 0.48**	96.64 ± 2.74	**90.91 ± 5.39**	**84.58 ± 8.76**
	Parzen_Region_G	97.45 ± 1.62	97.91 ± 1.25	97.32 ± 2.44	92.51 ± 4.79	86.78 ± 8.00
	SAM	99.03 ± 0.43	96.15 ± 2.37	99.42 ± 0.76	96.45 ± 1.11	93.17 ± 2.02
YOLOv8X	Parzen	97.87 ± 0.85	95.90 ± 3.66	98.27 ± 1.27	93.81 ± 3.23	88.75 ± 5.54
	Parzen_Clustering	98.17 ± 0.85	96.10 ± 3.67	97.90 ± 1.27	93.65 ± 3.20	88.98 ± 5.50
	Parzen_Region_G	98.00 ± 0.75	**94.06 ± 4.27**	98.17 ± 0.99	94.01 ± 3.13	89.24 ± 5.41
	SAM	98.91 ± 0.54	95.25 ± 2.82	99.39 ± 0.89	96.00 ± 1.40	92.35 ± 2.52
YOLOv9c	Parzen	97.64 ± 1.19	95.43 ± 2.99	97.40 ± 1.56	92.97 ± 4.10	87.16 ± 7.01
	Parzen_Clustering	97.44 ± 1.30	96.22 ± 2.62	97.45 ± 1.78	91.57 ± 4.64	87.11 ± 7.82
	Parzen_Region_G	98.04 ± 1.00	94.52 ± 2.86	**94.44 ± 1.65**	93.34 ± 3.59	88.41 ± 6.26
	SAM	98.78 ± 0.64	95.33 ± 3.01	99.27 ± 0.95	95.58 ± 1.88	91.59 ± 3.31
YOLOv9e	Parzen	97.53 ± 1.28	97.94 ± 1.90	97.42 ± 2.04	93.58 ± 3.76	88.33 ± 6.38
	Parzen_Clustering	97.63 ± 1.42	98.15 ± 1.41	97.39 ± 2.21	92.96 ± 3.99	87.48 ± 6.71
	Parzen_Region_G	98.12 ± 1.14	97.00 ± 2.44	97.67 ± 1.89	93.95 ± 3.29	89.54 ± 5.68
	SAM	98.65 ± 0.81	94.55 ± 3.40	99.17 ± 1.23	94.99 ± 2.43	90.55 ± 4.18

Table 5. Bartlett's statistical test for segmentation results (YOLO + SAM).

Comparison (YOLOv8n + SAM)	Accuracy	Sensitivity	Specificity	Dice	Jaccard
YOLOv8M + SAM	=	=	=	=	=
YOLOv8L + SAM	×	×	×	×	×
YOLOv8S + SAM	=	=	=	=	=
YOLOv8X + SAM	×	×	=	×	×
YOLOv9C + SAM	=	=	=	=	=
YOLOv9E + SAM	×	×	×	×	×

based on the metrics Accuracy, Precision, Recall, F1-Score, and Matthews. Additionally, the training and prediction time were also calculated.

Table 6 shows the hyperparameters selected for each model based on the extractor/classifier combination during the execution that generated the metrics for this study. It is interesting to note that the hyperparameters of the randomForest were generally quite similar across the combinations, with n_estimators ranging from 100 to 300. Additionally, the linear SVMs used the same value for C. These similarities might occur due to the characteristics of the classification dataset, ISIC 2017.

Table 7 presents the classification results obtained for the set of models formed by the different extractor/classifier combinations discussed in this study, for application on the ISIC 2017 dataset. The classified attribute was benign_malignant, one of the most important attributes in the dataset, and used in the classification of various methods in the literature.

It can be observed from Table 7 that, overall, the accuracies range between 70% and 85%, which is a good margin of accuracy considering the conditions of the dataset, with a large number of values. The precision values were low, ranging from 45% to 66%. Recall, in general, ranged between 47% and 55%. The Matthews correlation coefficient values were also low. These results suggest that the models achieved satisfactory results, but fell short in some more specific cases. This could likely be improved with further study on the dataset preprocessing, such as using fine-tuning, removing outliers, or even clustering the dataset to reduce the number of samples while maintaining representativeness.

The highest classification accuracy was 86.0%, achieved by the LBP + MLP model combination, which also had the highest precision value, 74%. The best recall, 64%, was obtained with the DenseNet201 + NaiveBayes combination. The same feature extractor, when combined with MLP, also achieved the highest F1-score (60%) and Matthews coefficient (23%). This indicates that MLP stood out as the classifier for the adopted dataset and hyperparameters, appearing in the best

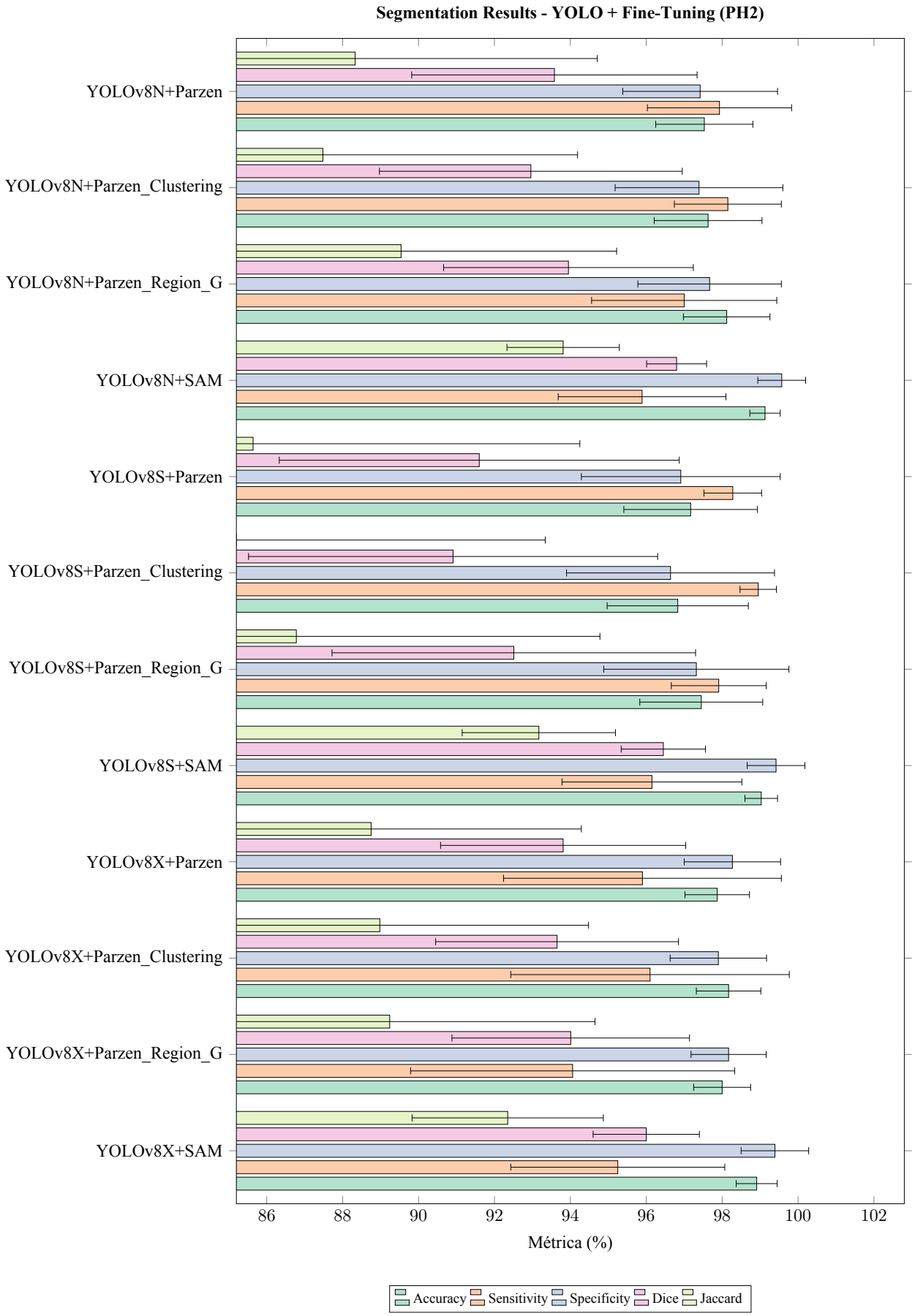

Figure 5. Comparative graph of the segmentation results for the fine-tuning models YOLOv8N, YOLOv8S, and YOLOv8X, from Table 4.

Table 6. Table 3 – Model Parameters

Feature Extractor	Classifier	Best Params
DenseNet121	KNN	{'n_neighbors': 9}
	MLP	{}
	NaiveBayes	{}
	RandomForest	{'n_estimators': 200, 'max_depth': None}
	SVM_Linear	{'C': 0.1}
	SVM_Polynomial	{'degree': 4, 'C': 0.1}
	SVM_RBF	{'gamma': 0.01, 'C': 10}
DenseNet169	KNN	{'n_neighbors': 7}
	MLP	{}
	NaiveBayes	{}
	RandomForest	{'n_estimators': 100, 'max_depth': None}
	SVM_Linear	{'C': 0.1}
	SVM_Polynomial	{'degree': 2, 'C': 0.1}
	SVM_RBF	{'gamma': 0.001, 'C': 0.1}
DenseNet201	KNN	{'n_neighbors': 9}
	MLP	{}
	NaiveBayes	{}
	RandomForest	{'n_estimators': 300, 'max_depth': None}
	SVM_Linear	{'C': 0.1}
	SVM_Polynomial	{'degree': 2, 'C': 0.1}
	SVM_RBF	{'gamma': 0.01, 'C': 10}
LBP	KNN	{'n_neighbors': 9}
	MLP	{}
	NaiveBayes	{}
	RandomForest	{'n_estimators': 200, 'max_depth': 20}
	SVM_Linear	{'C': 0.1}
	SVM_Polynomial	{'degree': 2, 'C': 0.1}
	SVM_RBF	{'gamma': 0.001, 'C': 0.1}
VGG16	KNN	{'n_neighbors': 9}
	MLP	{}
	NaiveBayes	{}
	RandomForest	{'n_estimators': 100, 'max_depth': None}
	SVM_Linear	{'C': 0.1}
	SVM_Polynomial	{'degree': 2, 'C': 0.1}
	SVM_RBF	{'gamma': 0.001, 'C': 0.1}
VGG19	KNN	{'n_neighbors': 7}
	MLP	{}
	NaiveBayes	{}
	RandomForest	{'n_estimators': 200, 'max_depth': 10}
	SVM_Linear	{'C': 0.1}
	SVM_Polynomial	{'degree': 2, 'C': 0.1}
	SVM_RBF	{'gamma': 0.001, 'C': 0.1}

combinations. It can be said that the best combination was LBP with MLP, but DenseNet201 also performed well.

LBP may have stood out as a texture extractor since the dermoscopic image dataset tends to separate the melanoma region from the patient's skin well, which may have facilitated the work of LBP extraction. On the other hand, DenseNet201 has a robust architecture and performs well even with a large dataset. The VGG, in contrast, had mediocre results, with some of the worst metrics.

The polynomial SVM, in general, had poor results in terms of metrics, as highlighted in the table. This could be due to the difficulty of modeling with this classifier for a complex dataset with a considerable number of attributes.

It is important to emphasize that the classification stage using the ISIC 2017 dataset is inherently affected by a strong class imbalance, where benign lesions considerably outnumber malignant cases. This imbalance tends to inflate overall accuracy while reducing recall, F1-Score, and MCC for the minority (malignant) class, as the model becomes biased toward predicting the majority class. The observed values — accuracy around 86% but lower F1 and MCC scores — are therefore consistent with the data distribution rather than indicative of poor model generalization.

In this context, the reduced performance for the attributes *anatom_site_general* and *age_approx* can also be explained by the limited contextual information available in the segmented images, which restricts the model's ability to capture relational or anatomical cues. Future versions of this work may incorporate rebalancing strategies, such as weighted loss functions or synthetic oversampling (e.g., SMOTE), as well as multimodal metadata fusion to mitigate bias and improve sensitivity in minority classes.

Table 8 presents the average classification accuracies for the considered best feature extractor/classifier combination, LBP + MLP. The ISIC 2017 dataset includes the attributes diagnosis_confirm_type, benign_malignant, clin_size_long_diam_mm, anatom_site_general, and age_approx.

The highest value was for diagnosis_confirm_type, with 90% accuracy. This class only indicates the presence or absence of melanoma in the image. In the most important class, benign_malignant, which indicates whether the melanoma is benign or malignant, the best combination achieved an average accuracy of 86%, a satisfactory value considering the dataset. The clin_size_long_diam_mm class, which refers to the melanoma's diameter, achieved 79% accuracy in classifying this attribute. The other metrics were below 50%, and in general, they are naturally more difficult to classify without the background region. It is highly likely that anatom_site_general, which indicates the melanoma's location on the patient's body, and age_approx, which indicates the patient's age, would have had better results if the input image for classification were the whole image and not just the segmented melanoma, as it currently is. However, this process would increase the initial attributes, representing a trade-off in performance.

5.4 Comparison with the state of the art

In this section, the results of studies found in the literature that used the PH2 and ISIC 2017 datasets, whose proposals are similar to the proposed study, are compared. The study compares both the segmentation dataset, using the metrics Accuracy, Sensitivity, Specificity, DICE, and Jaccard. Classification is compared using Accuracy.

All comparative results reported in Tables 8 and 9 were obtained exclusively from the test set, ensuring that no image from the training or validation stages was reused during evaluation. This procedure guarantees the fairness and reproducibility of the comparison with state-of-the-art methods.

5.4.1 Comparison Segmentation

Table 9 presents a comparison with the state of the art for the proposed model and methods found in the literature that also used the PH2 dataset for melanoma segmentation.

Table 9 provides a quantitative comparison of the results obtained by the proposed method compared to several approaches from the literature that used the same segmentation dataset, PH2. In general, the results of the proposed method stand out, especially in terms of accuracy, Jaccard index, Dice coefficient, sensitivity, and specificity. Almost all of these metrics were above 95% for the proposed method. These indices highlight the high capability of the method in precisely delineating the melanoma region and differentiating the affected tissue from healthy skin, showing a robustness that is desirable in clinical applications.

In comparison to classical methods based on deep neural networks, such as the Deep CNN pixel-wise segmentation [Badrinarayanan *et al.*, 2017] and the Full Resolution Convolution Networks approach [Al-Masni *et al.*, 2018], the

Table 7. Classification Results for the Extractor/Classifier Combinations Applied to the ISIC 2017 Dataset, with the Classified Attribute Being benign_malignant.

Feature Extractor	Classifier	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)	Matthews(%)	Train Time(s)	Predict Time(s)
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±0.00
	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	0.00±0.00
	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	0.00±0.00
	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±0.00
	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±0.00
	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
DenseNet169	KNN	85.0±1.0	66.0±5.0	53.0±1.0	52.0±2.0	13.0±5.0	0.57±0.01	0.04±0.01
	MLP	83.0±2.0	61.0±4.0	57.0±3.0	58.0±3.0	18.0±6.0	12.73±1.34	0.00±0.00
	NaiveBayes	68.0±2.0	56.0±1.0	61.0±2.0	55.0±2.0	17.0±3.0	0.06±0.00	0.00±0.00
	RandomForest	86.0±1.0	73.0±19.0	51.0±0.0	48.0±1.0	8.0±5.0	553.99±8.45	0.01±0.00
	SVM_Linear	81.0±2.0	60.0±4.0	58.0±3.0	58.0±3.0	17.0±7.0	22.89±2.09	0.09±0.01
	SVM_Polynomial	85.0±1.0	44.0±3.0	50.0±0.0	46.0±0.0	0.0±0.0	26.80±3.34	0.12±0.01
	SVM_RBF	86.0±1.0	53.0±16.0	50.0±1.0	47.0±1.0	4.0±6.0	79.11±5.14	1.48±0.54
DenseNet201	KNN	85.0±1.0	65.0±7.0	53.0±1.0	52.0±2.0	12.0±4.0	0.78±0.25	0.06±0.06
	MLP	84.0±1.0	65.0±3.0	59.0±1.0	**60.0±1.0**	**23.0±3.0**	14.24±2.28	0.00±0.00
	NaiveBayes	69.0±2.0	58.0±2.0	**64.0±2.0**	57.0±2.0	21.0±4.0	0.07±0.00	0.00±0.00
	RandomForest	86.0±1.0	70.0±19.0	51.0±1.0	48.0±2.0	7.0±6.0	**597.90±15.94**	0.01±0.01
	SVM_Linear	82.0±1.0	61.0±3.0	59.0±2.0	59.0±2.0	19.0±4.0	20.83±0.79	0.12±0.01
	SVM_Polynomial	85.0±1.0	**43.0±1.0**	**50.0±0.0**	**46.0±0.0**	**0.0±0.0**	34.85±2.75	0.17±0.00
	SVM_RBF	86.0±1.0	64.0±21.0	51.0±1.0	47.0±1.0	6.0±6.0	87.90±4.19	**1.72±0.58**
LBP	KNN	85.0±1.0	62.0±4.0	52.0±1.0	50.0±1.0	9.0±3.0	0.05±0.00	0.00±0.00
	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	0.00±0.00
	NaiveBayes	77.0±2.0	59.0±2.0	63.0±2.0	60.0±2.0	22.0±4.0	0.01±0.00	0.00±0.00
	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±0.00
	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±0.00
	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±0.00
	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±0.00
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±0.00
	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	0.00±0.00
	NaiveBayes	58.0±5.0	56.0±2.0	62.0±3.0	51.0±4.0	17.0±4.0	0.02±0.00	0.00±0.00
	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±0.00
	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±0.00
	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±0.00
	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
VGG19	KNN	85.0±2.0	66.0±11.0	53.0±1.0	52.0±2.0	13.0±6.0	0.18±0.01	0.01±0.00
	MLP	82.0±1.0	59.0±3.0	57.0±3.0	58.0±3.0	16.0±5.0	26.26±3.40	0.00±0.00
	NaiveBayes	65.0±7.0	57.0±2.0	62.0±2.0	55.0±4.0	18.0±4.0	0.02±0.00	0.00±0.00
	RandomForest	85.0±1.0	69.0±19.0	50.0±0.0	47.0±0.0	5.0±4.0	206.01±4.86	0.01±0.00
	SVM_Linear	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	0.0±0.0	16.14±1.39	0.03±0.00
	SVM_Polynomial	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	0.0±0.0	7.30±0.56	0.04±0.00
	SVM_RBF	85.0±1.0	43.0±1.0	50.0±0.0	46.0±0.0	0.0±0.0	10.63±0.28	0.08±0.02

Table 8. Average classification accuracy results for other attributes of the ISIC 2017 dataset for the LBP + MLP combination.

Feature Extractor	Classifier	Label	Accuracy(%)
LBP	MLP	diagnosis_confirm_type	90.0 ± 1.0
		benign_malignant	86.0 ± 1.0
		clin_size_long_diam_mm	79.0 ± 3.0
		anatom_site_general	41.0 ± 2.0
		age_approx	12.0 ± 1.0

Jaccard and Dice indices of the proposed method (93.81% and 96.80%, respectively) show a significant improvement over these methods. While these traditional approaches generally yielded values ranging from 80% to 91%, the proposed method achieved values approaching the gold standard, highlighting the effectiveness of the enhanced segmentation process through fine-tuning that aims to improve the boundaries of the melanoma binary segmentation mask.

Another relevant comparison is with geometric and clustering-based methods, such as Delaunay Triangulation [Pennisi *et al*., 2016] and the Clustering-based method [Patiño *et al*., 2018]. While these techniques offer relatively good accuracy, they achieved lower sensitivity and overlap coefficients compared to the proposed method. The sensitivity shown by the proposed method (95.89%) underscores its superior ability to detect the areas affected by melanoma, minimizing false negatives, which is crucial for accurate clinical analysis.

More recent approaches from the literature, such as the Deep class-specific learning features segmentation [Bi *et al*., 2019] and the Geodesic Active Contour [Vasconcelos *et al*., 2019], also achieved competitive results. However, they still fell short of the performance of the proposed method, particularly in the precision of the segmented boundaries and specificity. With an average specificity of 99.57%, the proposed method demonstrates refined control over the delimitation between the area of interest and the background, an essential aspect to reduce false positives and ensure the reliability of the system in clinical contexts.

5.4.2 Comparison Classification

Table 10 provides a comparative illustration of the results obtained in the melanoma classification stage, using the ISIC 2017 dataset. It is evident that the proposed method, combining the LBP + MLP extractor/classifier set, achieved an accuracy of 86.0 ± 1.0%, surpassing other approaches in the literature. This performance is notable, especially when compared to popular and current deep architectures such as Inception-v3, ResNet-50, and Inception-ResNet-v2, which achieved values in the range of 81% to 82%.

This superiority in terms of accuracy for the ISIC 2017 dataset demonstrates that the strategy of using LBP for tex-

Table 9. Comparison table with the state of the art for works that used the same PH2 segmentation dataset.

Methods	Accuracy(%)	Jaccard(%)	Dice(%)	Sensitivity(%)	Specificity(%)
Proposed Method	**99.13 ± 0.40**	**93.81 ± 1.48**	**96.80 ± 0.79**	**95.89 ± 2.21**	**99.57 ± 0.63**
SDA-Detaction Melanoma	96.50	87.22	92.60	88.38	99.53
Deep CNN pixel wise segmentation (2015) [Badrinarayanan *et al*., 2017]	93.36	80.77	89.36	86.53	96.61
Delaunay triangulation (2016) [Pennisi *et al*., 2016]	89.66	-	-	80.24	97.22
Clustering-based (2018) [Patiño *et al*., 2018]	90.39	-	-	91.04	89.73
Full resolution convolution networks (2018) [Al-Masni *et al*., 2018]	95.08	84.79	91.77	93.72	95.65
FCN encoder–decoder (2018) [Baghersalimi *et al*., 2019]	-	85.3	91.5	-	-
Deep class-specific learning features seg (2019) [Bi *et al*., 2019]	95.30	85.90	92.10	96.23	94.52
Geodesic active contour (2019) [Vasconcelos *et al*., 2019]	94.59	86.16	92.17	91.72	97.99
FLog Parzen level set (2020) [de Souza Rebouças *et al*., 2021]	94.25	85.09	92.49	93.02	93.21

Table 10. Comparison with the State of the Art for Melanoma Classification Regarding the ISIC 2017 Dataset

Author	Feature Extractor	Classifier	Accuracy(%)
Proposed	LBP	MLP	**86.0 ± 1.0**
al [Al-Masni *et al*., 2020]	Inception-v3	Inception-v3	81.29 ± -
al [Al-Masni *et al*., 2020]	ResNet-50	ResNet-50	81.57 ± -
al [Al-Masni *et al*., 2020]	Inception-ResNet-v2	Inception-ResNet-v2	81.34 ± -
al [Al-Masni *et al*., 2020]	DenseNet-201	DenseNet-201	73.44 ± -
yilmaz [Yilmaz *et al*., 2021]	NASNetMobile - 16	NASNetMobile - 16	82.00 ± -
yilmaz [Yilmaz *et al*., 2021]	MobileNetV2 - 16	MobileNetV2 - 16	81.45 ± -
yilmaz [Yilmaz *et al*., 2021]	MobileNet - 32	MobileNet - 32	80.73 ± -
budhiman [Budhiman *et al*., 2019]	ResNet 50	ResNet 50	78.50 ± -
budhiman [Budhiman *et al*., 2019]	ResNet 25	ResNet 25	80.70 ± -
budhiman [Budhiman *et al*., 2019]	ResNet 7	ResNet 7	82.40 ± -

ture feature extraction, which is important for differentiating between benign and malignant areas of the skin, combined with a well-parameterized MLP classifier, is effective even in the face of CNN-based models. While methods using CNNs, such as those described by [Al-Masni *et al*., 2020] and [Yilmaz *et al*., 2021], show formidable performance, the results obtained with the LBP + MLP combination indicate that less complex approaches can, in certain scenarios, offer a competitive level of accuracy in melanoma classification.

As for the more complex models, it is highly likely that their performance could be improved with studies on their parameterization in relation to the specific characteristics of the ISIC 2017 dataset, or even preprocessing methods, in combination with the robust models applied.

5.5 Integration with *Blockchain*

The developed application demonstrates that integrating *blockchain* technology allows for the secure and distributed recording of all information generated in the final stage of the process, including processed images, model predictions, and clinical reports. This type of storage ensures the integrity, traceability, and immutability of the data, avoiding reliance on centralized databases. Additionally, using *blockchain* promotes transparency in the proposed method and enables the secure and controlled sharing of information between different medical institutions. In this way, it strengthens trust, auditability, and accountability in the use of automated solutions as support for dermatological diagnosis.

Storing clinical images in IPFS (*InterPlanetary File System*) significantly contributes to the decentralization and security of the data generated during the diagnostic process. By utilizing a distributed file system, each processed image is fragmented and identified by a unique hash, ensuring its integrity and preventing unauthorized alterations. Furthermore, the links generated by IPFS are immutable and can be directly referenced in *blockchain* transactions, enhancing the traceability and authenticity of the records. This approach ensures a reliable and permanent repository of clinical evidence, allowing for future audits and promoting greater transparency and trust in the use of automated systems for dermatological diagnosis support.

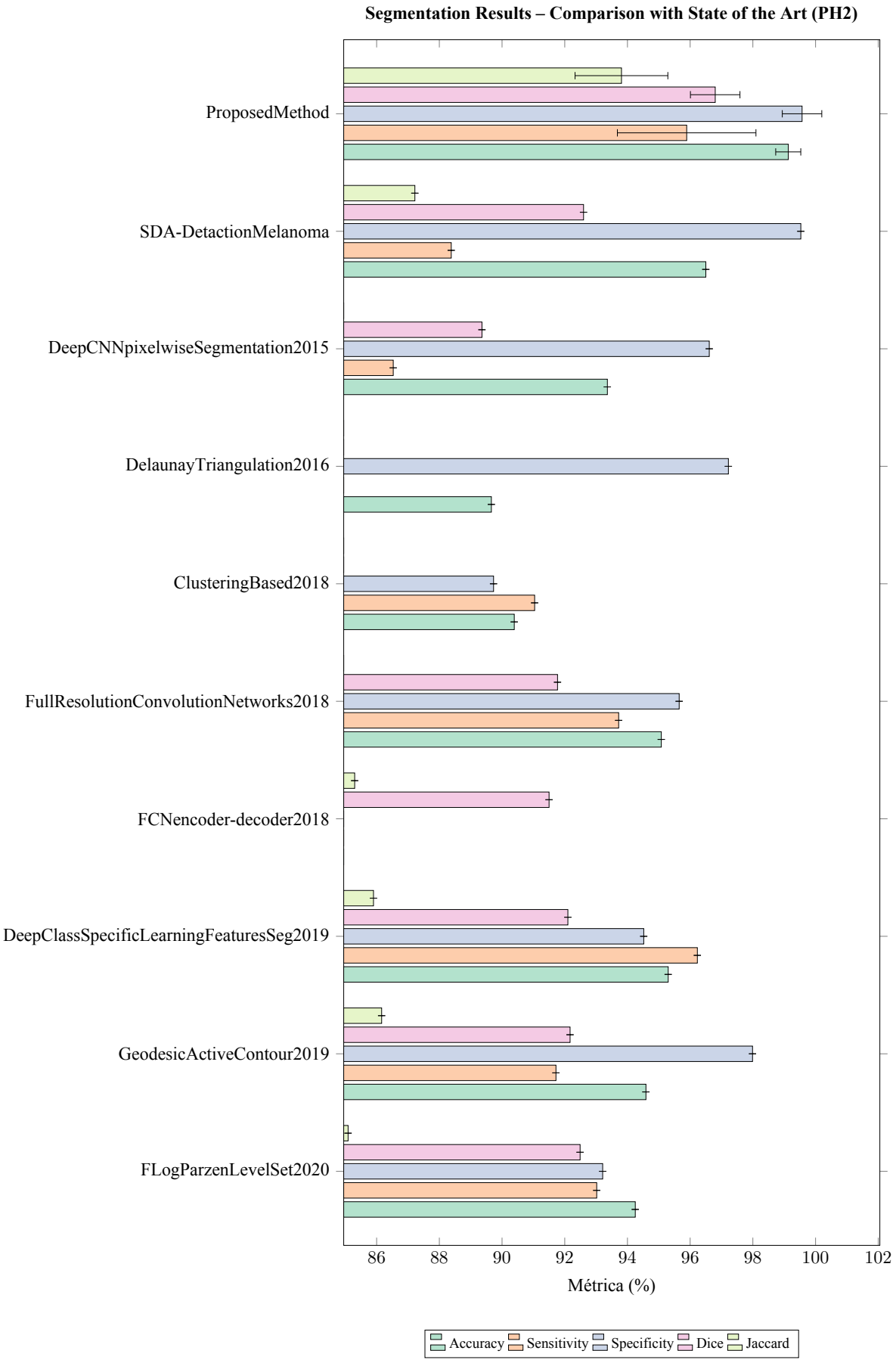

Figure 6. Comparative graph of segmentation results for works that used the PH2 dataset. ”nan” values correspond to data not provided in the table.

6 Conclusion

This study proposed an integrated framework that integrates different Artificial Intelligence techniques for the automated analysis of dermoscopic images, focusing on the identification of melanoma. The detection and segmentation processes were conducted using YOLO and SAM architectures, achieving an accuracy greater than 99%, achieved competitive performance benchmark methods in the current literature. For the classification stage, a structured pipeline was developed that combined feature extraction techniques with supervised learning methods, employing strategies such as grid search, transfer learning, and cross-validation. The best experimental configuration was obtained with the LBP-MLP combination, reaching an accuracy of 86%, highlighting the potential of the proposed approach to support clinical diagnosis.

The construction of the final model was carried out through the selection and combination of various methods for each step—detection, segmentation, and classification—evaluated based on empirical performance. The best-performing configuration was then compared to results from similar studies in the literature, demonstrating superior performance across different scenarios and datasets, thus validating the robustness of the adopted approach.

As a distinctive feature, the proposed pipeline incorporates a blockchain-based logging system at its final stage, responsible for the secure and immutable storage of results, including segmented images, predictions, and reports generated by generative AI. This additional layer enables full traceability of the process, ensuring data integrity while fostering transparency and reliability in the system's clinical use. Furthermore, the blockchain allows for secure and auditable data sharing among different institutions, enhancing interoperability across digital medical systems.

Although the method developed in this study presents a conceptually concrete approach to aid medical diagnosis, integrated with blockchain, obeying security protocols in data movement, it is a fact that more in-depth studies on computational cost, communication flow bottlenecks and tests in a more comprehensive environment are of vital importance to have a better perspective of implementing the methodology as a production system.

As future work, the proposed methodology will be applied to additional datasets, such as brain tumors and hemorrhagic stroke. novel combinations will also be proposed through transfer learning. Additionally, the study will focus on security testing procedures related to diagnostic validation, traceability, medical auditing, secure data sharing, immutability, and trust in medical imaging data and sensitive patient information, as well as the use of various melanoma-related datasets.

Declarations

Authors' Contributions

Marcus Vinicius Candido de Figueiredo contributed to the theoretical studies related to blockchain. Adriell Gomes Marques contributed to the development of the code and the design of the core methodology. Francisco Italo Guilherme Da Silva contributed to the practical advancement of the blockchain studies, developing the methods under the guidance of Marcus Vinicius Candido de Figueiredo. José Jerovane Da Costa Nascimento contributed to image development, including methodological design representations and system visualization. Cidcley Teixeira de Souza reviewed the work and proposed improvements and adjustments to enhance the quality of the research. Luís Fabrício de Freitas Souza supervised and coordinated the research, secured funding, and provided guidance on the methodology. All authors also contributed to the writing of the manuscript, each providing input according to their respective areas of expertise. All authors read and approved the final version of the manuscript. The authors gratefully acknowledge the LISIA (Artificial Intelligence Systems Research and Innovation Group) for their support, resources, and collaboration in the development of this research.

Competing interests

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgements

The authors would like to thank all individuals and institutions that contributed to the development of this research, directly or indirectly. The authors would like to express their special thanks to the Research Group and Innovation Laboratory in Artificial Intelligence Systems - LISIA, and to Professor Dr. Fabrício Freitas for their guidance.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

AbdelSalam, F. (2023). A systematic review of blockchain technology benefits and threats. *PMC*. Available at: `https://pmc.ncbi.nlm.nih.gov/articles/PMC10701638/`.

Abedini, M. (2024). Skin cancer classification by leveraging segment anything model for semantic segmentation of skin lesion. DOI: 10.17148/IJARCCE.2024.13835.

AI, M. (2023). Llama 2 - 13b. `https://huggingface.co/meta-llama/Llama-2-13b`. Accessed: 2025-10-04.

Aishwarya, N., Kannaa, G. Y., and Seemakurthy, K. (2025). Yoloskin: A fusion framework for improved skin cancer diagnosis using yolo detectors on nvidia jetson nano. *Biomedical Signal Processing and Control*, 100:107093. DOI: 10.1016/j.bspc.2024.107093.

Aishwarya, N., Prabhakaran, K. M., Debebe, F. T., Reddy, M. S. S. A., and Pranavee, P. (2023). Skin cancer diag-

nosis with yolo deep neural network. *Procedia Computer Science*, 220:651–658. DOI: 10.1016/j.procs.2023.03.083.

Al-Masni, M. A., Al-Antari, M. A., Choi, M.-T., Han, S.-M., and Kim, T.-S. (2018). Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Computer methods and programs in biomedicine*, 162:221–231. DOI: 10.1016/j.cmpb.2018.05.027.

Al-Masni, M. A., Kim, D.-H., and Kim, T.-S. (2020). Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Computer methods and programs in biomedicine*, 190:105351. DOI: 10.1016/j.cmpb.2020.105351.

Alickovic, A. (2023). *React Application Architecture for Production: Learn best practices and expert tips to deliver enterprise-ready React web apps*. Packt Publishing. Available at: `https://ieeexplore.ieee.org/document/10162152`.

Alquran, H., Qasmieh, I. A., Alqudah, A. M., Alhammouri, S., Alawneh, E., Abughazaleh, A., and Hasayen, F. (2017). The melanoma skin cancer detection and classification using support vector machine. In *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pages 1–5. DOI: 10.1109/AEECT.2017.8257738.

Arnold, M., Singh, D., Laversanne, M., Vignat, J., Vaccarella, S., Meheus, F., Cust, A. E., De Vries, E., Whiteman, D. C., and Bray, F. (2022). Global burden of cutaneous melanoma in 2020 and projections to 2040. *JAMA dermatology*, 158(5):495–503. DOI: 10.1001/jamadermatol.2022.0160.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495. DOI: 10.1109/TPAMI.2016.2644615.

Baghersalimi, S., Bozorgtabar, B., Schmid-Saugeon, P., Ekenel, H. K., and Thiran, J.-P. (2019). Dermonet: densely linked convolutional neural network for efficient skin lesion segmentation. *EURASIP Journal on Image and Video Processing*, 2019(1):1–10. DOI: 10.1186/s13640-019-0467-y.

Berseth, M. (2017). Isic 2017-skin lesion analysis towards melanoma detection. *arXiv preprint arXiv:1703.00523*. DOI: 10.48550/arXiv.1703.00523.

Bhosale, K., Akbarabbas, K., Deepak, J., and Sankhe, A. (2019). Blockchain based secure data storage. *International Research Journal of Engineering and Technology (IRJET)*, 6(3):5058–5061. Available at: `https://www.irjet.net/archives/V6/i3/IRJET-V6I31286.pdf`.

Bi, L., Kim, J., Ahn, E., Kumar, A., Feng, D., and Fulham, M. (2019). Step-wise integration of deep class-specific learning for dermoscopic image segmentation. *Pattern recognition*, 85:78–89. DOI: 10.1016/j.patcog.2018.08.001.

Blaut, G., Ma, X., and Wolter, K. (2023). Explorando a aleatoriedade em blockchains. In *Conferência Internacional IEEE 2023 sobre Blockchain e Criptomoeda (ICBC)*. DOI: 10.1109/ICBC56567.2023.10174962.

Budhiman, A., Suyanto, S., and Arifianto, A. (2019). Melanoma cancer classification using resnet with data augmentation. In *2019 international seminar on research of information technology and intelligent systems (ISRITI)*, pages 17–20. IEEE. DOI: 10.1109/ISRITI48646.2019.9034624.

de Souza Rebouças, E., de Medeiros, F. N. S., Marques, R. C. P., Chagas, J. V. S., Guimaraes, M. T., Santos, L. O., Medeiros, A. G., and Peixoto, S. A. (2021). Level set approach based on parzen window and floor of log for edge computing object segmentation in digital images. *Applied Soft Computing*, 105:107273. DOI: 10.1016/j.asoc.2021.107273.

Deng, R., Cui, C., Liu, Q., Yao, T., Remedios, L. W., Bao, S., Landman, B. A., Wheless, L. E., Coburn, L. A., Wilson, K. T., *et al*. (2023). Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging. *arXiv preprint arXiv:2304.04155*. DOI: 10.2352/EI.2025.37.14.COIMG-132.

Dhanvardini, R., Martina, P., Vijay, R., Amirtharajan, R., and Pravinkumar, P. (2023). Development and integration of dapp with blockchain smart contract truffle framework for user interactive applications. In *2023 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–6. DOI: 10.1109/ICCCI56745.2023.10128406.

Duarte, A. F., Sousa-Pinto, B., Azevedo, L. F., Barros, A. M., Puig, S., Malvehy, J., Haneke, E., and Correia, O. (2021). Clinical abcde rule for early melanoma detection. *European Journal of Dermatology*, 31(6):771–778. DOI: 10.1684/ejd.2021.4171.

Ferrer, L. (2022). Analysis and comparison of classification metrics. *arXiv preprint arXiv:2209.05355*. DOI: 10.48550/arXiv.2209.05355.

Frange, V. M. N., Arruda, L. H. F., and Daldon, P. É. C. (2009). Dermatoscopia: importância para a prática clínica. *Revista de Ciências Médicas*, 18(4). Available at: `https://seer.sis.puc-campinas.edu.br/cienciasmedicas/article/view/637`.

Gül, S., Cetinel, G., Aydin, B. M., Akgün, D., and Öztaş Kara, R. (2025). Yolosamic: A hybrid approach to skin cancer segmentation with the segment anything model and yolov8. *Diagnostics*, 15(4):479. DOI: 10.3390/diagnostics15040479.

Hussain, M. (2023). Yolo-v1 to yolo-v8, the rise of yolo and its complementary nature toward digital manufacturing and industrial defect detection. *Machines*, 11(7):677. DOI: 10.3390/machines11070677.

Islayem, R., Gebreab, S. A., AlKhader, W., Musamih, A., Salah, K., Jayaraman, R., and Khan, M. K. (2025). Using large language models for enhanced fraud analysis and detection in blockchain based health insurance claims. *Scientific Reports*, 15(1):29763. DOI: 10.1038/s41598-025-15676-4.

Jiang, P., Ergu, D., Liu, F., Cai, Y., and Ma, B. (2022). A review of yolo algorithm developments. *Procedia computer science*, 199:1066–1073. DOI: 10.1016/j.procs.2022.01.135.

Kotter, E., Marti-Bonmati, L., Brady, A. P., and Desouza, N. M. (2021). Esr white paper: blockchain and medical imaging. *Insights into Imaging*, 12(1):82. DOI: 10.1186/s13244-021-01029-y.

Kumar, R. and Tripathi, R. (2019). Implementation of distributed file storage and access framework using ipfs and blockchain. In *2019 Fifth International Conference on Image Information Processing (ICIIP)*, pages 246–251. DOI: 10.1109/ICIIP47207.2019.8985677.

Lee, W.-M. (2023). Using the metamask crypto-wallet. In *Beginning Ethereum Smart Contracts Programming: With Examples in Python, Solidity, and JavaScript*, pages 111–144. Apress, Berkeley, CA. DOI: 10.1007/978-1-4842-9271-6$_5$.

Lyakhova, U. and Lyakhov, P. (2024). Systematic review of approaches to detection and classification of skin cancer using artificial intelligence: Development and prospects. *Computers in Biology and Medicine*, 178:108742. DOI: 10.1016/j.compbiomed.2024.108742.

Marchisotti, G. (2025). Tecnologia blockchain para compartilhamento de dados: projeto b-cpf da receita federal do brasil. *Revista Brasileira de Casos de Ensino em Administração*, 15:c6. DOI: 10.12660/gvcasosv15n1c6.

Marques, A. G., de Figueiredo, M. V. C., da Costa Nascimento, J. J., de Souza, C. T., de Mattos Dourado, C. M. J., de Albuquerque, V. H. C., de Freitas Souzal, L. F., *et al*. (2024). New approach generative ai melanoma data fusion for classification in dermoscopic images with large language model. In *2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 1–6. IEEE. DOI: 10.1109/SIBGRAPI62404.2024.10716298.

Mendonça, T., Celebi, M., Mendonca, T., and Marques, J. (2015). Ph2: A public database for the analysis of dermoscopic images. *Dermoscopy image analysis*, 2. DOI: 10.1201/b19107.

Mhaske, H. R. e Phalke, D. A. (2013). Detecção e classificação do câncer de pele melanoma com base no aprendizado supervisionado e não supervisionado. In *Conferência Internacional de Circuitos, Controles e Comunicações de 2013 (CCUBE)*, pages 1–5. DOI: 10.1109/CCUBE.2013.6718539.

Müller, D., Soto-Rey, I., and Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1):210. DOI: 10.1186/s13104-022-06096-y.

Nagaoka, T. (2025). Improved skin lesion segmentation in dermoscopic images using object detection and semantic segmentation. *Clinical, Cosmetic and Investigational Dermatology*, pages 1191–1198. DOI: 10.2147/CCID.S518751.

Nayeem, M. J., Rana, S., Alam, F., and Rahman, M. A. (2021). 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)*, pages 280–284. IEEE. DOI: 10.1109/ICICT4SD50815.2021.9397013.

Patil, Shankar M. e Rajguru, B. S. e. M. R. S. e. P. O. P. (2022). Detecção de doença de câncer de pele melanoma usando rede neural convolucional. In *2022 3ª Conferência Internacional de Tecnologia Emergente (INCET)*, pages 1–5. DOI: 10.1109/INCET54531.2022.9825381.

Patiño, D., Avendaño, J., and Branch, J. W. (2018). Automatic skin lesion segmentation on dermoscopic images by the means of superpixel merging. In *International conference on medical image computing and computer-assisted intervention*, pages 728–736. Springer. DOI: 10.1007/978-3-030-00937-3$_8$3.

Pennisi, A., Bloisi, D. D., Nardi, D., Giampetruzzi, A. R., Mondino, C., and Facchiano, A. (2016). Skin lesion image segmentation using delaunay triangulation for melanoma detection. *Computerized Medical Imaging and Graphics*, 52:89–103. DOI: 10.1016/j.compmedimag.2016.05.002.

Phoon, K.-K., Quek, S.-T., and An, P. (2003). Identification of statistically homogeneous soil layers using modified bartlett statistics. *Journal of Geotechnical and Geoenvironmental Engineering*, 129(7):649–659. DOI: 10.1061/(ASCE)1090-0241(2003)129:7(649).

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788. DOI: 10.1109/CVPR.2016.91.

Rezk, E., Haggag, M., Eltorki, M., and El-Dakhakhni, W. (2023). A comprehensive review of artificial intelligence methods and applications in skin cancer diagnosis and treatment: Emerging trends and challenges. *Healthcare Analytics*, page 100259. DOI: 10.1016/j.health.2023.100259.

Ring, C., Cox, N., and Lee, J. B. (2021). Dermatoscopy. *Clinics in Dermatology*, 39(4):635–642. DOI: 10.1016/j.clindermatol.2021.03.009.

Sachdeva, S. (2009). Fitzpatrick skin typing: Applications in dermatology. *Indian Journal of Dermatology, Venereology and Leprology*, 75(1):93–96. DOI: 10.4103/0378-6323.45238.

Singh, G. and Sachan, M. (2014). Multi-layer perceptron (mlp) neural network technique for offline handwritten gurmukhi character recognition. In *2014 IEEE International Conference on Computational Intelligence and Computing Research*, pages 1–5. DOI: 10.1109/ICCIC.2014.7238334.

Thapar, P., Rakhra, M., Alsaadi, M., Quraishi, A., Deka, A., and Ramesh, J. V. N. (2024). A hybrid grasshopper optimization algorithm for skin lesion segmentation and melanoma classification using deep learning. *Healthcare Analytics*, 5:100326. DOI: 10.1016/j.health.2024.100326.

Topsakal, O. and Akinci, T. C. (2023). Creating large language model applications utilizing langchain: A primer on developing llm apps fast. In *International Conference on Applied Engineering and Natural Sciences*, volume 1, pages 1050–1056. DOI: 10.59287/icaens.1127.

Vasconcelos, F. F. X., Medeiros, A. G., Peixoto, S. A., and Reboucas Filho, P. P. (2019). Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour. *Cognitive Systems Research*, 55:44–59. DOI: 10.1016/j.cogsys.2018.12.008.

Wainstein, A. J. and Belfort, F. A. (2004). Conduta para o melanoma cutâneo. *Revista do Colégio Brasileiro de Cirurgiões*, 31:204–214. DOI: 10.1590/S0100-69912004000300011.

Wang, H., Zhao, S., Qiang, Z., Xi, N., Qin, B., and Liu, T. (2024a). Beyond direct diagnosis: Llm-based multi-specialist agent consultation for automatic

diagnosis. *arXiv preprint arXiv:2401.16107*. DOI: 10.48550/arXiv.2401.16107.

Wang, J., Yang, Z., Yao, Z., and Yu, H. (2024b). Jmlr: Joint medical llm and retrieval training for enhancing reasoning and professional question answering capability. *arXiv preprint arXiv:2402.17887*. DOI: 10.48550/arXiv.2402.17887.

Xiao, H., Li, L., Liu, Q., Zhu, X., and Zhang, Q. (2023). Transformers in medical image segmentation: A review. *Biomedical Signal Processing and Control*, 84:104791. DOI: 10.1016/j.bspc.2023.104791.

Xu, Y., Dos Santos, M. A., F. Souza, L. F., Marques, A. G., Zhang, L., da Costa Nascimento, J. J., de Albuquerque, V. H. C., and Rebouças Filho, P. P. (2022). New fully automatic approach for tissue identification in histopathological examinations using transfer learning. *IET Image Processing*, 16(11):2875–2889. DOI: 10.1049/ipr2.12449.

Yilmaz, A., Kalebasi, M., Samoylenko, Y., Guvenilir, M. E., and Uvet, H. (2021). Benchmarking of lightweight deep learning architectures for skin cancer classification using isic 2017 dataset. *arXiv preprint arXiv:2110.12270*. DOI: 10.48550/arXiv.2110.12270.

Zheng, Z., Xie, S., Dai, H., Chen, X., and Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. In *2017 IEEE International Congress on Big Data (BigData Congress)*, pages 557–564. DOI: 10.1109/BigDataCongress.2017.85.