JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Deep Learning and Optimization for Automated Skin Melanoma Detection: A Focused Review

G Rajasekhar Reddy
Research Scholar,
Department of CSE,
Annamalai University, Chidambaram.
rajasekhar9440467437@gmail.com

Meena Abarna K.T
Assistant Professor,
Department of CSE,
Annamalai University, Chidambaram.
abarnakt@yahoo.com

K Subba Reddy
Professor,
Department of CSE,
RGM College of Engg & Tech, Nandyal.
mrsubbareddy@yahoo.com

Abstract—Melanoma is a serious type of skin cancer and early detection is important to treatment. While dermoscopy is a valuable tool for information in pigmented lesion diagnosis, it requires clinical experience, could be influenced by the quality of the image, lesion variability, and complexity of the diagnosis. Deep learning can learn discriminative features from images automatically which saves the effort of hand-crafted features. As a result, convolutional neural networks (CNNs) have emerged as a significant research area in computer assisted skin lesion analysis. This review deals with automated melanoma analysis which involves image pre-processing, segmentation, feature representation, classification, dataset selection, and optimization of the model. Compares basic machine learning with deep learning and introduces CNN architectures, transfer learning, hyper-parameter optimization (HPO). Special emphasis is placed on publicly available dermoscopic databases and optimization techniques, such as Random Search, Swarm Intelligence algorithm. The review highlights continuing challenges, including dataset diversity, class imbalance, sensitivity to image artifacts, limited generalization, computational cost, and model interpretability. Finally, it is suggested that a CNN framework inspired by optimization could be a potential direction for future research.

Keywords: Melanoma detection, dermoscopy, deep learning, convolutional neural networks, image segmentation, hyperparameter optimization.

I. INTRODUCTION

Melanoma is a type of skin cancer that develops from cells called melanocytes, which make the skin's pigment. It may have substantial variation in presentation, which makes the classification of melanoma based on images a difficult pattern recognition problem. Considerable research has focused on computer-assisted skin cancer analysis because deep neural networks have demonstrated promising classification performance in comparison with dermatologists under specific experimental conditions [1].

Clinical evaluation generally involves visual examination and dermoscopy, with histopathological examination performed when necessary. When interpreted by a trained and experienced dermatoscopist, dermoscopy can reveal patterns that may not be apparent during visual examination and can assist clinical assessment. Computer-aided systems are therefore being developed to provide reproducible image analysis and support, rather than replace, clinical decision-making. Systematic review of CNN-based skin lesion classification also reveals that it is difficult to compare the results from the different studies as the datasets, pre-processing methods, training methods and evaluation methods vary [2]-[3].

The number of public dermoscopic databases has increased and the use of common benchmarks for the evaluation of algorithms. Datasets, however, are not all alike in the appearance of their lesions, in the acquisition of their images, and in the categories of diagnosis. As such, a model that works well on one set of images might not work as well on images taken in a different clinical context. Hence this review not only focuses on classification architectures but the entire processing pipeline as well as the optimization decisions that could impact model reliability.

This review is to summarize the key technical aspects of automated analysis of melanoma. It covers skin and lesion characteristics, traditional machine learning techniques, deep learning techniques, public datasets, optimization techniques, research gaps and a suggested research direction of hybrid CNN optimization.

II. BACKGROUND ON SKIN MELANOMA DETECTION

A. Skin Structure and Melanoma Development

The skin is a multi-layered organ that acts in several physiological functions such as protection, sensation and thermoregulation. The epidermis is significant for image-based melanoma analysis because melanocytes are in the basal layer of this tissue. The dermis is the layer of tissue that is underneath the epidermis and has blood vessels, hair follicles, and glands. The subcutaneous layer is the layer beneath the dermis layer and acts as a cushioning and energy storage.

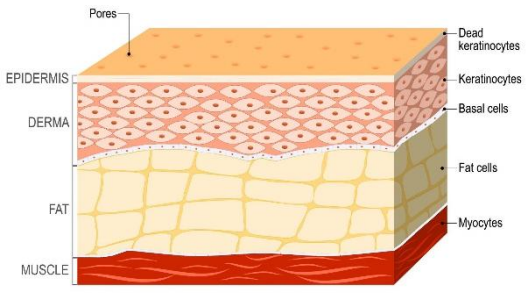

Figure 1: Simplified structure of human skin.

Normally, melanocytes make melanin, a protective barrier for the skin against ultraviolet (UV) radiation. When these cells grow abnormally, they can form melanoma. Suspicious lesions may be discolored, asymmetric, have irregular borders and different texture. The following visual features may be helpful for assessments, but

cannot be used as a sole diagnostic parameter. Automated systems should thus be viewed as clinical decision support systems, not as a substitute for clinical diagnosis.

Melanoma cells are found in the basal layer of the epidermis. In normal physiological conditions they control the production of melanin that protects skin from the harmful effect of ultraviolet radiation. Excessive exposure to the sun's ultraviolet radiation and genetic factors may play a role in the abnormal growth of melanocytes.

Melanoma spreads to other parts of the body or into tissue can be life-threatening. Therefore, early recognition is an important goal of screening and diagnostic processes. One of the big challenges in computation is the ability to differentiate malignant from benign lesions in the presence of different illumination, acquisition device, lesion size, and patient characteristics.

B. Clinical Diagnosis of Skin Melanoma

Clinical evaluation of melanoma may involve visual examination, dermoscopy, and histopathological assessment. Dermoscopy gives a magnified and illuminated view of lesion structures which may not be evident in the normal eye photographs. Clinical diagnosis involves more than any single checklist; however, the ABCDE warning characteristics (asymmetry, border irregularity, color variation, diameter, and evolution) are commonly considered during assessment. In automated image analysis, these visual characteristics are represented using numerical features and computational models.

These clinical characteristics are commonly considered by dermatologists during the evaluation of suspicious skin lesions.

The ABCDE warning properties that are normally thought of. The following is a summary of activities during the assessment of melanoma:

A – Asymmetry: shape of two halves of the lesion is different appearance.

B – Border irregularity:The lesion may be irregularly shaped, Poorly defined (irregular) boundaries.

C – Color variation: The lesion may be several different colors or non-uniform pigmentation.

D – Diameter: The diameter of the lesion is taken into account along with, other clinical characteristics.

E – Evolution: changes in the size, shape, colour of the lesion or changes in the number of lesions. other characteristics over time may indicate the need for further clinical evaluation.

Although dermoscopy provides valuable information about Its interpretation can be time consuming and, Needs special clinical skills Image artifacts, All of the above, such as hair, ruler marks, shadows, bubbles, and irregularities. The automated processing may be further complicated by illumination. Artifacts can be created by the methods of hair removal that have been researched. impair boundary detection and lesion analysis; Therefore, handling of artefacts is an important pre-processing task. consideration [6]. These observations are used to validate the use of standardized preprocessing, robust segmentation, and Evaluation of various data sets.

III. TRADITIONAL MACHINE LEARNING AND DEEP LEARNING APPROACHES

A. Machine Learning Approaches

Traditionally, a computer-aided diagnosis pipeline takes a dermoscopic image and changes it into a compact feature representation. This gets used by the classifier after that. The workflow has four stages. Image preprocessing happens first. Then lesion segmentation and feature extraction before classification. Traditional methods work when there are not enough resources or training data. It feels like that part is easy to miss sometimes.

i. Image preprocessing

The pre-analysis stage of image preprocessing works to eliminate unneeded variations which helps enhance image quality for following analysis procedures. The standard image processing steps consist of three main operations which include color normalization and contrast enhancement and artifact reduction. The segmentation process divides the lesion area from the surrounding skin tissue to reduce the impact of non-essential background pixels during the following processing steps. The traditional segmentation methods consist of threshold techniques and edge-based approaches and clustering algorithms and deformable model systems. The methods stop working properly when the lesion boundaries become difficult to identify because of uneven shapes or lighting distortions. The U-Net architecture for deep segmentation operates through an encoder-decoder system which includes skip connections to extract contextual data and maintain spatial information [7].

ii. Lesion segmentation

Melanoma analysis requires precise segmentation because any mistakes in boundary detection will affect the entire process of feature extraction and classification determination. The International Skin Imaging Collaboration (ISIC) challenges have proved benchmark evaluation to be useful but they show that test set results do not indicate how well systems will work with different image collections [5]. Numerical performance measures need to be combined with reliable segmentation and independent validation for complete evaluation.

iii. Feature extraction

The current systems extract descriptors through direct derivation from lesions which have been segmented during the process. The combination of different characteristics appears through color statistics and texture measures and shape descriptors and boundary irregularity information. The features show their value through proper design but their performance depends on feature selection procedures and their capacity to generalize to new data patterns. Deep learning solves this problem through its ability to create hierarchical representations which it learns from raw image data.

The distribution of lesion colors together with their intensity levels becomes the focus of color features which developers analyze through statistical approaches and color histogram methods. The surface patterns together with local intensity variations become the focus of texture features which researchers study through Local Binary Patterns (LBP). The geometric features of lesions become shape features which include their total area and perimeter length and their degree of shape distortion.

Boundary features represent the way lesions show their edges and their uneven shapes which medical professionals use to identify suspicious skin lesions.

Classical classifiers can use interpretable descriptors such as color histograms, Local Binary Patterns (LBP), and shape or structural measures based on image gradients. Their main limitation is that the representation is designed before model training and may not capture all discriminative structures. Modern systems may therefore combine conventional preprocessing with learned deep features.

iv. Classification

Once a feature vector has been constructed, lesion categories can be classified by training a suitable model. Common classifiers include k-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM). These models analyze compact feature sets to produce baseline results which help evaluate deep learning performance improvements.

Some common classifiers are traditional such as:

- Support Vector Machines (SVM)
- Random Forest
- k-Nearest Neighbors (KNN)

Traditional pipelines encounter a main operational challenge because they depend on systems which need human operators to determine features and perform initial processing steps procedures.If the selected descriptors fail to represent important lesion characteristics, the classifier cannot recover information that was not encoded in the feature vector. Researchers developed end-to-end deep learning models which solve this problem by learning feature representations together with classification functions in a single system.

B. Deep Learning Approaches

Deep learning models learn hierarchical representations directly from training data. In image analysis, initial layers Initially, image analysis systems detect local features through their first processing stage which later stages use to develop advanced identification systems.combine these features to generate increasingly abstract and discriminative representations. Dermoscopic images because convolutional operations exploit spatial structure and can learn reusable filters.

Convolutional neural networks include convolutional layers together with nonlinear activation functions and pooling operations and classification layers which can be multiple in number. The training process of neural networks uses backpropagation to adjust network parameters through weight updates which minimize the loss function values. The training process of neural networks uses backpropagation to update network parameters through weight adjustments which minimize the loss function values.Training aims to reduce the loss function through network parameter optimization which occurs during supervised learning with labeled data samples.The system organizes components into a hierarchy which makes it simpler to define all visual elements through manual work.The training process for CNNs demands large amounts of data which creates a risk for overfitting when medical data sets contain few samples or show highly uneven class distribution.The medical data set contains limited samples which leads to possible CNN training overfitting because of its data requirement for training

Skin lesion representation and classification require the following CNN architecture families for their application.

AlexNet: AlexNet is a deep convolutional neural network architecture that demonstrated the effectiveness of CNNs forlarge-scale image classification and established an important foundation for subsequent deep learning research [8].

VGGNet operates with small convolutional kernels which it combines with deeper network architecture to develop hierarchical image representations.The architectural design of this structure has served as a model which various transfer-learning applications have implemented for their development.

Residual Network (ResNet): ResNet employs residual mappings and shortcut connections to facilitate the training of substantially deeper neural networks while addressing optimization difficulties associated with increasing network depth [9].

EfficientNet: EfficientNet uses a compound scaling strategy to balance network depth, width, and input resolution, providing a trade-off between classification performance and computational efficiency [10].

They should not be measured only by reported accuracy as the training sets, augmentation methods, class definition, resolution of images, and validation process can be quite different. The target data set may have only a small number of training samples, in which case the pre-trained model may be a good initialization for the target model.

C. Publicly Available Datasets for Skin Lesion Analysis

Public datasets allow repeatability, with developers being able to conclude results based on the public and for the user to benchmark. Their usefulness depends not only on the value of research datasets depends on both their image count and the quality of their annotations and their diagnostic variety and their acquisition environment and their training and testing set independence, diversity, acquisition conditions, and the independence of training and test sets.

The International Skin Imaging Collaboration (ISIC) has ISIC facilitates dermoscopic image dataset development through its support for data collection and its organization of lesion segmentation and disease classification challenges which include attribute analysis.These benchmarks have led to standardized evaluation and research of model generalization [5].

The HAM10000 dataset consists of dermoscopic images of 7 different diagnostic categories corresponding to 10,015 cases.The images were collected from multiple sources, making the dataset a widely used benchmark for skin lesion classification research[4].

A publicly available data-set should not be viewed as encompassing the entire clinical population, however. Patient-level separation of the training and test sets should be used, the distributions of different classes should be reported, preprocessing should be standardized, and more than just accuracy should be reported. External/cross-dataset validation is helpful for assessing generalization.

D. Evaluation Metrics for Imbalanced Skin Lesion Datasets

The performance of automated melanoma detection systems should be assessed in different measures due to the fact that dermoscopic data sets available for the public are not generally distributed equally amongst all classes. If the dataset is imbalanced, a high overall accuracy could be obtained by a classifier that performs well on the majority class and poorly on the minority classes such as melanoma. So, it is not enough to be accurate to make an automated melanoma detection system reliable. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is complementary to other measures of classification performance, such as sensitivity, specificity, precision, F1-score, and balanced accuracy.

The accuracy is the number of samples correctly classified divided by all samples evaluated. It can give a general idea of how well classifiers are performing, but might be misleading if the data set has an unequal number of benign and malignant cases. A classifier with a higher accuracy but that misclassifies a significant number of melanomas.

The sensitivity (recall) is the percentage of the actual melanomas that are properly classified. It is defined as

$$TP/(TP + FN)$$

and is a ratio of true positive to false negative cases. Sensitivity is important in detection of melanoma as false negative predictions can lead to the failure to identify suspicious lesions and to a lack of clinical evaluation. When designing automated detection systems, then, it is important to consider how to make the system more sensitive to detecting melanomas.

Specificity is the percentage of people who don't have melanoma who are correctly classified by the classifier. It is defined as

$$TN/(TN + FP)$$

where TN is true negative and FP is false positive. Specificity provides the means to limit false positive predictions and avoid additional clinical follow-up. Specificity, however, should be taken into account in combination with sensitivity and not optimized alone.

Precision is the percentage of the predicted cases that are melanoma cases. It is defined as

$$TP/(TP + FP)$$

It is measure of the reliability of predictions of positive instances and a gauge of the impact of false positive classifications.

The F1 score is the harmonic mean of the precision and sensitivity, and it is defined as

$$\frac{2 \times (\text{Precision} \times \text{Sensitivity})}{(\text{Precision} + \text{Sensitivity})}$$

It combines the accuracy of the classifier to correctly identify positive cases with the false-positive predictions of the classifier. If the data is imbalanced, Class-wise F1 Score and macro-averaged F1 Score can be helpful to evaluate the performance of the minority classes without domination by the majority classes.

AUC-ROC shows how well the classification model separates the two classes in classification at various classification thresholds. An AUC score that is higher represents a better discriminatory ability between melanoma and non-melanoma cases. AUC should be interpreted along with sensitivity and specificity as well as the distribution of the classes.

Balanced accuracy can also be used as a good measure for imbalanced datasets. It is the mean of the sensitivity and specificity in binary classification. If the dataset is multi-class, such as HAM10000, then macro-averaging and reporting class-wise sensitivity is recommended to prevent a performance drop for less prevalent diagnostic classes being hidden by the majority classes.

Therefore a good melanoma detection system needs to provide several other complementary measures besides overall accuracy. Confusion matrices, independent validation (if possible), and particular focus on melanoma sensitivity, minority-class F1-score, balanced accuracy, and AUC-ROC should be considered. These measures offer a more holistic evaluation of the classification performance and possible clinical usefulness of automated systems in a clinical decision-support role.

E. Optimization Techniques in Deep Learning

CNN performance can vary based on hyperparameters such as learning rate, batch size, optimizer, regularization techniques, data augmentation, and training time. Manual tuning is challenging if the search space is large. Hyperparameter Optimization (HPO) is a systematic method of testing various configurations according to a well-defined objective and validation process.

i. Grid Search and Random Search

Grid search exhaustively tests all possible combinations of the hyperparameters, which can become costly if there are many parameters and values. Random search is able to sample configurations from given distributions, and can search more efficiently. Bergstra and Bengio have shown that random search can be a good starting point for HPO in the presence of a small number of hyperparameters that have a strong effect on performance [11].

ii.Swarm Intelligence Algorithms

Another method for traversing large and/or non-convex configuration spaces is through the use of metaheuristic optimization. The swarm intelligence techniques are inspired from collective search behaviour and can be applied to optimize a continuous parameter or a discrete parameter. They work best under certain conditions with respect to the objective function, design of the search space, computational budget and stability of the resulting model. As they should be compared against simpler optimizations baselines and not taken as better for certain. Some popular optimization techniques based on swarm intelligence are listed below.

Particle Swarm Optimization (PSO)

Grey Wolf Optimizer (GWO)

Ant Colony Optimization (ACO)

Other population-based meta-heuristics.

A population based metaheuristic algorithm inspired by the hierarchical leadership structure and hunting behaviour of grey wolves is one of these methods and is known as Grey Wolf Optimizer (GWO) [12]. During CNN development, hyperparameters such as learning rate, number of filters, dropout rate, and batch size may be optimized. The optimization process needs to achieve target validation results through rapid execution while preventing the optimization target from developing overfitting issues.

Research Gap Analysis

The major challenges and research gaps associated with automated melanoma detection are discussed below.

Although significant progress has been achieved, automated melanoma detection continues to face technical and clinical challenges. First, public datasets do not fully represent the diversity of clinical images encountered in practice. The performance across different datasets becomes affected by variations which exist between acquisition devices and patient populations and lesion types and annotation methods used. The second problem of class imbalance leads to better classifier performance on common categories but it produces incorrect classifications for rare yet medically essential categories.

The system encounters a major obstacle because it fails to correctly identify images which contain artifacts. The preprocessing methods succeed in removing hair and other artifacts but they fail to eliminate the effects of shadows and uneven lighting and ruler markings which impact the detection of lesion boundaries and their texture details. Another difficulty is model selection and tuning: a CNN can have high training/validation accuracy but low test (unseen) accuracy. Reported performance also depends on evaluation design and dataset composition.

Table 1 summarizes the major research gaps in automated melanoma detection, their effects on classification performance, and corresponding research directions. Table 2 presents comparative overview of existing deep learning approaches for skin lesion classification.

Common classes can overwhelm optimization, obscuring bad minority class sensitivity. Hair and other illumination changes can change artifacts in the image, which can be confusing when trying to determine the boundary or texture of lesions.

Implementation of proven pretreatment and test artifact hardening is most important. Boundary errors can propagate for feature learning and classification. Manual choices can lead to non-reproducible and/or non-stable outcomes. Large networks can lead to over-fitting small datasets and may be more expensive to train and make inferences with. Table 1: Consolidated Research Gaps and Directions

Research gap	Effect on melanoma detection	Recommended direction
Dataset diversity	A model may perform well on one benchmark but degrade on images from another source.	Use multi-source data and external or cross-dataset validation.
Class imbalance	Majority classes may dominate classification performance, resulting in poor melanoma sensitivity and increased false-negative predictions.	Report melanoma-specific sensitivity, class-wise precision and F1-score, macro-averaged metrics, and balanced accuracy. Apply suitable class-balancing strategies and independent validation.
Image artifacts	Hair, ruler marks and illumination changes can distort lesion boundaries and texture.	Apply validated preprocessing and test artifact robustness.
Segmentation quality	Boundary errors can propagate into feature learning and classification.	Compare segmentation quality and evaluate end-to-end impact.
Hyperparameter sensitivity	Manual choices can produce unstable or non-reproducible results.	Use controlled search methods and report search spaces.
Model complexity	Large networks may increase training and inference cost and overfit small datasets.	Compare accuracy with parameter count, training cost and inference cost.
Generalization	Internal validation may not represent real clinical deployment conditions.	Use patient-level splits and independent validation.

Table 2: Quantitative Comparison of Existing Deep Learning Approaches for Skin Lesion Classification.

Study / method	Metrics to extract	Advantages to evaluate	Limitations to evaluate
Esteva et al. (2017), deep CNN	Accuracy and other reported evaluation metrics	Demonstrates deep-learning-based skin cancer classification	Dataset and evaluation setting must be considered when assessing generalizability
Brinker et al. (2018), CNN systematic review	Metrics reported by individual studies	Synthesizes evidence on CNN-based skin cancer classification	Differences between study designs may limit direct comparisons

HAM10000-based classification studies	Class-wise sensitivity, specificity, precision, F1-score, macro-F1, balanced accuracy, and AUC-ROC, where reported.	Enables evaluation on a publicly available, multi-source dataset	Class imbalance and dataset-specific performance
ResNet-based studies	Reported classification metrics and model complexity, where available	Residual connections facilitate training deeper networks	Performance depends on the dataset, training procedure, and model configuration
EfficientNet-based studies	Reported classification metrics, parameter count, and computational cost, where available	Balances network depth, width, and input resolution	Actual performance and efficiency depend on implementation and evaluation conditions
Proposed PSO/GWO-optimized CNN framework (future research)	Baseline versus optimized metrics and optimization cost (to be evaluated experimentally)	Systematic exploration of model configurations	Additional computational overhead and possible overfitting during optimization

Table 2 provides a comparative overview of selected deep learning approaches for skin lesion classification. The following discussion critically examines the contributions and limitations of the reviewed studies.

Critical Analysis of the Literature

This section critically reviews existing studies on deep learning-based skin lesion classification, highlighting their contributions, limitations, and potential research directions.

Recent years have seen impressive progress in automated skin lesion classification, driven largely by advances in deep learning, transfer learning, image segmentation, and hyperparameter tuning. The first papers to show the promise of deep neural networks for dermoscopic image analysis of skin cancer, such as the ones pioneered by Estéva et al., introduced seminal architectures like AlexNet, ResNet and EfficientNet and dermoscopic segmentation tools like the pioneering U-Net by Ronneberger et al. But converting these theoretical models to actual clinical environments presents its very own set of challenges. Models often fail to perform well on the data they have not encountered before and face challenges such as low performance on the minority class, execution time, high dimensionality, and class imbalance. Moreover, the introduction of physical artifacts (e.g. hair) in dermoscopic images may alter features and early segmentation mistakes tend to propagate to produce poor classification performance downstream. One potential approach to improving model performance and addressing these bottlenecks is hyperparameter optimization.

The methods include simple methods such as random search proposed by Bergstra and Bengio, as well as metaheuristic approaches such as the Grey Wolf Optimizer (GWO) proposed by Mirjalili et al. These optimization methods can explore complex hyperparameter spaces well, but they can also require a lot of computational power with multiple iterations of training runs – and these methods improve over traditional training cannot just be taken on faith without empirical evidence. It highlights an interesting lack in the existing literature. The need of this study is an empirical study that is accurately systematic to evaluate the performance of CNN based on optimization by Particle Swarm Optimization (PSO) and GWO with conventional training models and random-search models. Such research should include a comprehensive assessment of sensitivity, specificity, precision, F1-score, AUC-ROC, and computational complexity and independent validation [13]. Newer explainable methods, such as Skin-CAD, have incorporated feature selection and transfer learning, but the utility of these methods remains to be seen in a large and diverse test set. SkinNet-14 aims to classify skin images based on low-resolution dermoscopic images while keeping training time optimized, however it is not generalizable in all clinical settings yet and further evaluation is necessary[14]. A lightweight cross-attention transformer is used to learn multi-scale attention of the lesion, but transformer-based models can be faced with issues in terms of computational efficiency and stability when dealing with various datasets. Multiclass classification with visual explanations and lesion segmentation integrate to an explainable hybrid approach, but the explainability and performance using external datasets are yet to be evaluated. The findings of these studies reveal improvements in classification performance, efficiency and interpretability; however, these studies continue to emphasize the need for a more rigorous nature of generalization and clinical validation.

Future Research

The research, which is proposed, focuses on creating an optimization-based hybrid CNN approach for automated diagnosis of melanoma. This approach compares CNN setup as a structured optimization problem rather than adjusting all the hyper parameters. The entire pipeline involves image preprocessing, an optional stage for artifact removal, lesion-based feature representation, deep feature learning using CNNs and supervised classification. A baseline CNN and a transfer-learning model will be created with a public dermoscopic dataset during the first stage. Baseline benchmarks are to be the standard training techniques and random-search optimization. Additionally, in the second phase PSO and GWO will be studied for their ability to search for hyperparameter values that maximizes the validation accuracy of the model while avoiding the unnecessarily complex ones.

Next, the different metaheuristic algorithms are going to be extensively tested to ensure that they produce some consistent and repeatable increase in performance when compared to simpler baseline algorithms, rather than achieving a lucky increase on a single experiment and floundering on others. The sensitivity (recall), specificity, precision, F1-score, balanced accuracy, and Area under the Receiver Operating Characteristic Curve Evaluation will be considered in the evaluation, with specific metrics depending on class distribution of the dataset. There is a particular focus on the sensitivity of melanoma detection and the accuracy rates for minority classes, to prevent false negatives for melanoma diagnosis. If there is multiclass data, class specific break down metrics as well as macro-averaged scores will be provided where applicable.Confusion matrices and independent and/or cross-dataset validation will also be evaluated to determine reliability and generalization. We should consider some pretrained models such as ResNet and EfficientNet as top candidates to use the transfer learning part. The use of interpretability techniques will be secondary to the main goal, used to indicate areas of the image that are most impacting predictions, but not necessarily enough to provide a definitive statement of the accuracy of a diagnosis. The overall objective is to have a framework that is efficient in computation and can be reproduced, and that can be well-tested for generalization performance in the real world.

IV. CONCLUSION

Automated skin lesion analysis has evolved from the field of automated skin lesion analysis has progressed from using traditional image features and basic classification methods to deep neural networks which extract important features directly from dermoscopic images. The scope of computer-assisted lesion analysis has been extended through CNNs, transfer learning, public datasets, and systematic optimization. The deployment process faces six main obstacles which include diverse datasets and unbalanced classes and image defects and hyperparameter tuning challenges and high computational demands and poor model adaptation to new data. The evaluation process needs to assess three essential factors which include robustness and reproducibility and external validity in addition to classification accuracy. A hybrid optimization-based system enables the evaluation of PSO and GWO algorithms through their ability to enhance CNN hyperparameter selection while keeping

computational requirements low and preserving both model generalization and classification accuracy.

REFERENCES

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017, doi: 10.1038/nature21056.
[2] M. Dildar, S. Akram, M. Irfan, H. U. Khan, M. Ramzan, A. R. Mahmood, et al., "Skin cancer detection: A review using deep learning techniques," Int. J. Environ. Res. Public Health, vol. 18, no. 10, p. 5479, 2021, doi: 10.3390/ijerph18105479.
[3] T. J. Brinker, A. Hekler, J. S. Utikal, N. Grabe, D. Schadendorf, J. Klode, et al., "Skin cancer classification using convolutional neural networks: Systematic review," J. Med. Internet Res., vol. 20, no. 10, e11936, 2018, doi: 10.2196/11936.
[4] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," Sci. Data, vol. 5, p. 180161, 2018, doi: 10.1038/sdata.2018.161.
[5] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC)," arXiv:1902.03368, 2019.
[6] Q. Abbas, M. E. Celebi, and I. Fondón García, "Hair removal methods: A comparative study for dermoscopy images," Biomed. Signal Process. Control, vol. 6, no. 4, pp. 395–404, 2011, doi: 10.1016/j.bspc.2011.01.003.
[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234–241.
[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1097–1105.
[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[10] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Machine Learning, PMLR, vol. 97, 2019, pp. 6105–6114.
[11] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," J. Mach. Learn. Res., vol. 13, pp. 281–305, 2012.
[12] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," Adv. Eng. Softw., vol. 69, pp. 46–61, 2014, doi: 10.1016/j.advengsoft.2013.12.007.
[13] O. Attallah, "Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning," Computers in Biology and Medicine, vol. 178, Art. no. 108798, 2024, doi: [10.1016/j.compbiomed.2024.108798]
[14] A. Al Mahmud, S. Azam, I. U. Khan, S. Montaha, A. Karim, A. Haque, M. Z. Hasan, M. Brady, and R. Biswas, "SkinNet-14: A deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time," Neural Computing and Applications, vol. 36, pp. 18935–18959, 2024, doi: 10.1007/s00521-024-10225-y.
[15] D. P. Yadav, B. Sharma, S. Chauhan, J. L. Webber, and A. Mehbodniya, "Dual scale light weight cross attention transformer for skin lesion classification," PLOS ONE, vol. 19, no. 12, Art. no. e0312598, 2024, doi: 10.1371/journal.pone.0312598.
[16] M. Fiaz, M. B. S. Khan, A. H. Khan, A. Bilal, M. Abdullah, A. A. Darem, and R. Sarwar, "An explainable hybrid deep learning framework for precise skin lesion segmentation and multi-class classification," Frontiers in Medicine, vol. 12, 2025, doi: 10.3389/fmed.2025.1681542