

MULTI-INPUT MELANOMA CLASSIFICATION USING MOBILENET-V3-LARGE ARCHITECTURE

Submitted: 24th May 2024; accepted: 9th July 2024

Serra Aksoy

DOI: 10.14313/jamris-2025-008

Abstract:

All over the globe there exists a serious problem with skin cancer, most especially melanoma; a malignancy which is known to behave aggressively and can metastasize. Detecting this early is the key to saving lives. This study introduces a new method of classifying melanoma using an advanced model known as MobileNet-V3-Large. This technique differs from others in that it considers both the images of the skin lesion and tabular data including factors consisting of the patient's approximate age, gender and the location of the lesion on the body of the patient. Such an approach empowers the predictions of whether the skin lesion may be malignant or benign. Tested on a huge collection consisting of skin images combined with tabular data, it was established that this method outperforms others already existing. The results of this study showed that a high accuracy of 99.56% was achieved using the proposed model. This study indicates that utilizing the multi-input method will substantially enhance diagnosis for melanoma hence reducing mortalities in the future.

Keywords: melanoma cancer, benign, malignant, convolution neural network (CNN), deep learning (DL), pre-trained model

1. Introduction

In 2023, it's estimated that nearly 97,610 individuals in the US received the daunting diagnosis of invasive malignant melanoma, which was slightly lower than the 99,780 cases reported in 2022. Tragically, the number of lives lost to this disease increased from 7,650 in 2022 to an estimated 7,990 in 2023 [1–3]. Projections indicated that approximately 89,070 cases of early in situ melanoma would be identified in 2023. The silver lining is that when melanomas are detected and treated at this early stage, there's a chance for complete recovery. Since most melanomas manifest visibly on the skin, it underscores the crucial importance of early detection and treatment [4]. Skin cancer, increasingly prevalent today, affects humans widely due to the skin's status as the body's largest organ [5]. When facing skin cancer, it's important to be proactive and utilize all available resources to detect it early and enhance treatment outcomes. This type of cancer encompasses two main categories: melanoma and nonmelanoma skin cancer [6]. Melanoma, though rare, is perilous and often fatal, accounting for a mere

1% of cases yet yielding a higher death toll. Originating from melanocytes, which regulate skin pigmentation, melanoma can emerge anywhere on the body but commonly appears on sun-exposed areas like the hands, face, and neck. Early detection is crucial for effective treatment; otherwise, melanoma can metastasize, leading to dire consequences. Various subtypes exist, including nodular melanoma, superficial spreading melanoma, acral lentiginous, and lentigo maligna. Conversely, nonmelanoma skin cancers like basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and sebaceous gland carcinoma (SGC) are more prevalent. These cancers typically arise in the outer layers of the skin and have a lower tendency to metastasize, making them relatively easier to treat compared to melanoma [6–12].

Fortunately, technology plays a crucial role in aiding dermatologists. Artificial intelligence acts as a reliable assistant, swiftly analyzing images of skin lesions to identify potential signs of cancer with greater speed and accuracy. Additionally, advanced imaging techniques like reflectance confocal microscopy and optical coherence tomography provide valuable insights into skin abnormalities without invasive procedures [3, 13–15]. In the field of medical imaging, experts are hands-on in assembling datasets and guiding machine learning researchers on what aspects to consider. Consider the HAM10000 dataset, for instance; it's a collection of images showing different skin lesions [17]. Since 2016, it has played a crucial role in the International Skin Imaging Collaboration (ISIC) challenge, assisting in addressing various skin-related issues [17–19]. There are other datasets like PH², featuring 200 dermoscopic images sorted into various diagnosis groups, and the Interactive Atlas of Dermoscopy, which offers over 1000 clinical cases with thorough annotations and pathology findings. These resources have proven invaluable in advancing computer-aided diagnosis (CAD) research [21].

Segmentation is a foundational and intricate task in the automated analysis of skin lesions. Within clinical contexts, rule-based diagnostic systems heavily rely on precise lesion segmentation to estimate critical diagnostic parameters such as asymmetry, border irregularity, and lesion size accurately. These metrics serve as fundamental components for the application of established algorithms like the ABCD algorithm and its derivatives, such as ABCDE and ABCDEF.

Table 1. Skin cancer detection methodologies and their respective datasets

Reference	Methodology	Dataset(s)	Evaluation Metrics
[22]	The preprocessing images and fine-tuning convolutional neural networks with transfer learning, with EfficientNet B4 identified as the top-performing model.	HAM10000 dataset	F1 Score: 87%, Accuracy: 87.91%
[23]	Automated Skin-Melanoma Detection (ASMD) system using image processing and SVM-based classification, proposing a Melanoma-Index (MI) for clinical use.	DD image dataset	Accuracy: 97.50%
[24]	Automatic skin cancer diagnosis system including Histogram of Gradients (HG) and Histogram of Lines (HL), combined with other features.	HPH dermoscopy database and the Dermofit standard database	Accuracy: 98.79% (HPH) and 92.96% (the standard Dermofit)
[25]	Skin cancer detection system utilizing Genetic Programming (GP) for evolving a classifier and feature selection.	PH ² dataset	Accuracy: 97.92%
[26]	Image processing and deep learning techniques, including Convolutional Neural Networks (CNNs), for skin cancer detection and classification.	MNIST HAM10000 dataset	Weighted Average Accuracy: 0.88, Weighted Average Recall: 0.74, Weighted F1-score: 0.77
[27]	Classification of skin lesions, utilizing dynamic-sized kernels and both ReLU and leakyReLU activation functions.	HAM10000 dataset	Overall accuracy: 97.85%
[28]	Soft-Attention mechanism in deep neural architectures for skin lesion classification.	HAM10000 dataset and ISIC-2017 dataset	Precision: 93.7% (HAM10000), sensitivity: 91.6% (ISIC-2017)
[29]	MobileNetV3 introducing the Improved Artificial Rabbits Optimizer (IARO) algorithm to enhance feature selection	PH ² , ISIC-2016, and HAM10000 datasets	Accuracy: 87.17% (ISIC-2016), 96.79% (PH ² dataset), and 88.71% (HAM10000)
[30]	SkinTrans, an improved transformer network, for skin cancer classification, utilizing vision transformers (ViT) with self-attention mechanism.	HAM10000 and clinical datasets	Accuracy: 94.3% (HAM10000) and 94.1% (Clinical)

The ABCD algorithm provides a structured framework for evaluating skin lesions, with each letter representing key morphological features. Accurate lesion segmentation is imperative for enabling reliable and effective automated diagnostic assessments, thereby enhancing clinical decision-making processes. As such, robust segmentation methodologies are essential for ensuring the integrity and clinical utility of automated skin lesion analysis systems [30–33].

In the world of academic research, three-field plots (Figure 1) and co-occurrence networks (Figure 2) serve as invaluable tools for understanding the intricate web of connections within scholarly literature (93 reviews). These visualizations provide insights into how keywords, university affiliations, and authors' countries of origin intersect. These visual tools provide a snapshot of how ideas move and evolve within academia. Conceptualize these tools as architectural schematics, opening the complex interrelations among these elements. The height of the boxes within the three-field plot signifies the volume of publications associated with each affiliation, offering a straightforward gauge of research output. A taller box indicates a higher number of publications from that affiliation, a key metric for evaluating research productivity.

In research publications, a graphical representation (see Figure 2) called the co-occurrence network of author keywords shows the connections between different author keywords. In the same research documents, individual keywords are nodes within these networks and edges that connect these nodes indicate the repeated co-occurrence of such words often. Through the research field or a specific set of publications, these networks highlight essential information by presenting an elastic picture of the underlying relationships. Upon closer examination of Figure 1, it is revealed that Germany, the United States, and Australia prominently emerge as the forerunners in academic endeavors that chronicle the steps made within the artificial-intelligence driven melanoma.

In Figure 2, specifically tailored analysis of keywords associated with artificial intelligence is offered, providing a comprehensive breakdown with a total of 30 nodes.

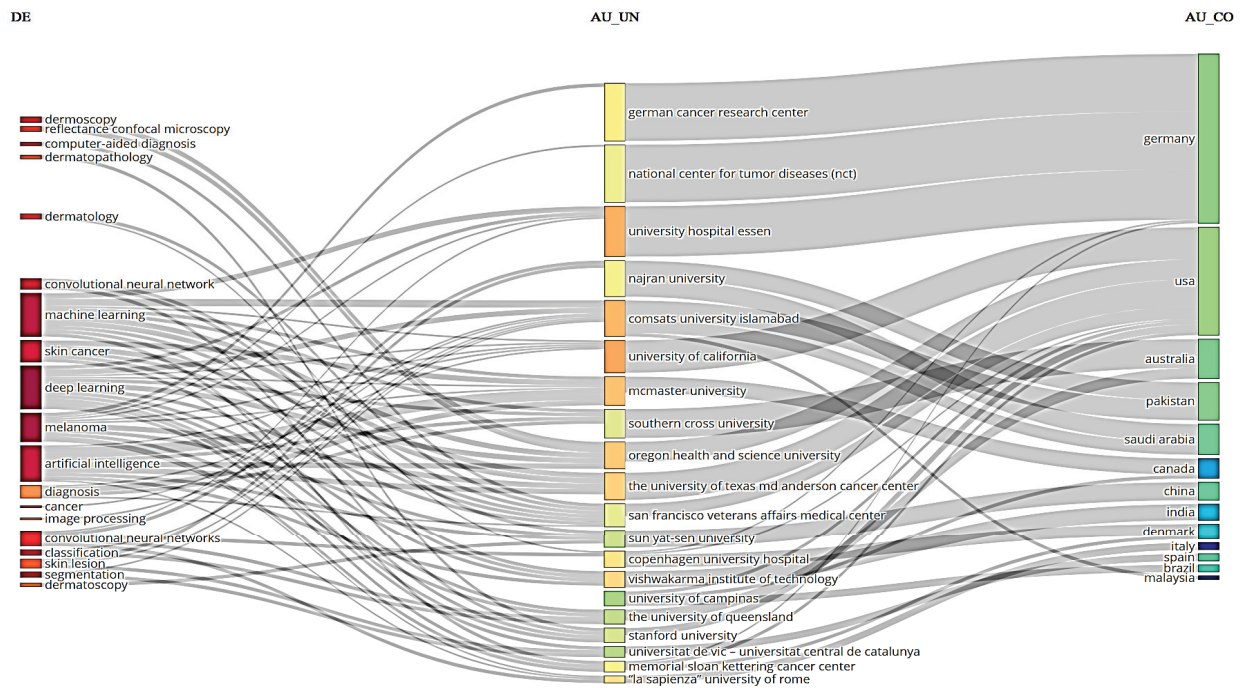


Figure 1. Three-field plot analysis (AU_UN—ID—AU_CO)

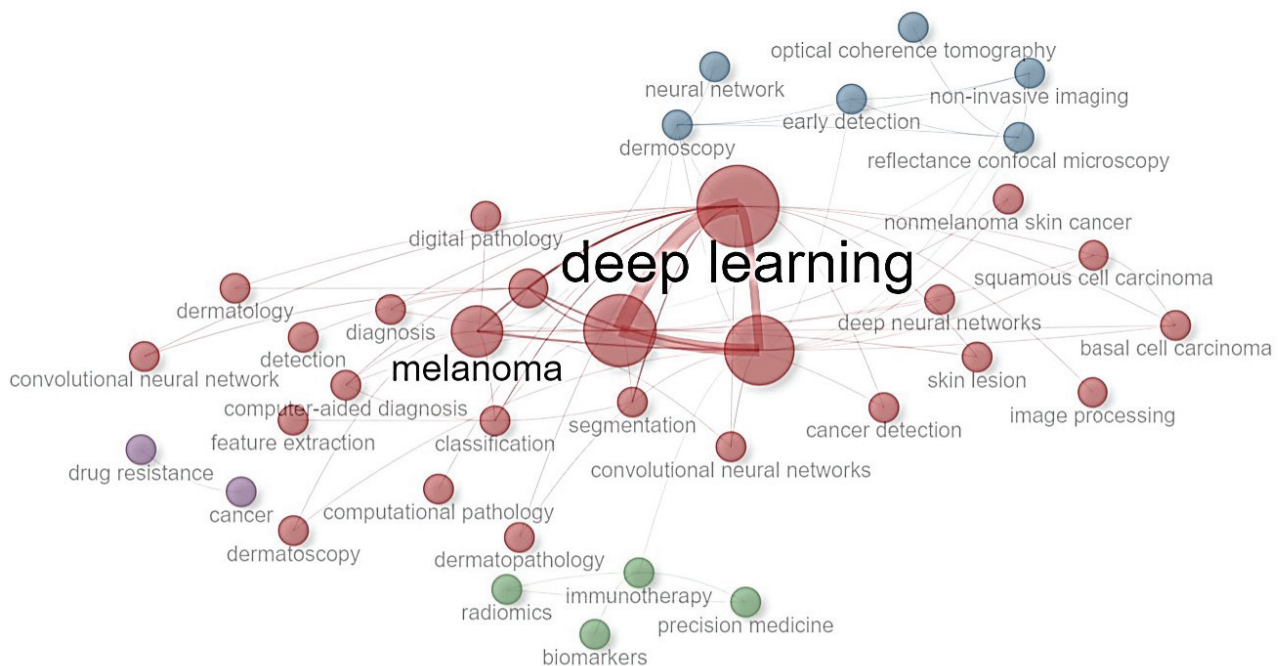


Figure 2. Co-occurrence network of author keywords

2. Different Pretrained Convolutional Neural Networks

In this study the feature blocks of these pretrained Convolutional Neural Networks (CNNs) [35] have been used as the feature extractor of the given image data, and they have been compared according to the classification metrics of the whole proposed architecture, which consists of the feature extractor block from the pretrained model, with that the best metrics have been obtained and a classification block for the image data, as well as another classification block for

the tabular data, and a last classification block from which the combined image and tabular data pass.

2.1. ConvNeXt Base

ConvNeXt Base, an innovative CNN architecture, represents a significant advancement in the field of computer vision. While not as widely recognized as some mainstream architectures, ConvNeXt Base has garnered attention for its robust feature extraction capabilities and computational efficiency.

Notable for its utilization of grouped convolutions, layer normalization, and stochastic depth regularization, ConvNeXt Base aims to strike a delicate balance between model complexity and performance. In terms of performance, ConvNeXt Base has demonstrated promising results in various image classification tasks. For instance, researchers have employed ConvNeXt Base to classify complex medical images, such as those related to tumor detection in mammography and histopathological analysis. Its ability to capture intricate visual patterns while maintaining computational efficiency makes ConvNeXt Base a compelling choice for diverse applications. Architecturally, ConvNeXt Base comprises a series of CNBlocks, each featuring grouped convolutions, linear transformations, and permutations. These blocks, augmented with stochastic depth layers, facilitate enhanced model robustness and generalization capabilities. Furthermore, ConvNeXt Base leverages layer normalization and various activation functions to capture complex patterns effectively.

2.2. MobileNet V3 Large

MobileNet V3 Large stands as a testament to the evolution of lightweight CNN architectures optimized for mobile and embedded devices. Built upon the success of its predecessors, MobileNet V3 Large prioritizes computational efficiency without sacrificing classification accuracy. Its architectural design, characterized by depth-wise separable convolutions, inverted residual blocks, and squeeze-and-excitation modules, enables efficient feature extraction across diverse datasets.

In practical applications, MobileNet V3 Large has found widespread adoption, particularly in scenarios where resource constraints pose challenges. For example, it has been instrumental in various mobile applications, ranging from real-time object detection to image classification in low-power devices. The incorporation of hard-swish activation functions and dropout layers further enhances its performance and regularization capabilities. Architecturally, MobileNet V3 Large comprises a hierarchical structure of depth-wise separable convolutions and inverted residual blocks, facilitating efficient feature extraction and spatial downsampling. Additionally, squeeze-and-excitation modules enable adaptive feature recalibration, enhancing the discriminative power of the model across different input domains.

2.3. VGG16

VGG16, a seminal CNN architecture, has left an indelible mark on the landscape of image classification. Renowned for its simplicity and effectiveness, VGG16 remains a cornerstone in the field despite the emergence of newer architectures. Its architectural design, characterized by repeated blocks of convolutional layers followed by rectified linear unit (ReLU) activation functions and max-pooling operations, prioritizes feature extraction and spatial downsampling.

In practice, VGG16 has been extensively utilized in various computer vision tasks, ranging from image recognition to object localization [6]. Its straightforward design and strong performance make it a popular choice for benchmarking and experimentation in research settings. Architecturally, VGG16 comprises a hierarchical structure of convolutional layers, augmented by ReLU activation functions and max-pooling operations. This design fosters hierarchical feature extraction, enabling the model to capture increasingly abstract representations as information traverses deeper into the network. Additionally, the incorporation of max-pooling layers facilitates spatial downsampling, reducing computational complexity while preserving discriminative information.

2.4. EfficientNet V2 S

EfficientNet V2 S emerges as a pinnacle in the realm of scalable and efficient CNN architectures, embodying state-of-the-art advancements in model design and optimization. With a focus on achieving optimal performance across varying computational budgets, EfficientNet V2 S has garnered widespread acclaim for its adaptability to diverse deployment scenarios. Its architectural design, characterized by a standard convolutional layer followed by fused MobileNetV2 (MBConv) blocks, epitomizes efficiency without compromising classification accuracy.

In practical applications, EfficientNet V2 S has demonstrated remarkable efficacy across a spectrum of tasks, from image classification to object detection. Its hierarchical structure and strategic incorporation of stochastic depth layers contribute to enhanced model robustness and generalization capabilities. Architecturally, EfficientNet V2 S comprises a cascade of MBConv blocks, each featuring depth-wise separable convolutions and efficient channel attention mechanisms. This design fosters efficient feature extraction and aggregation, enabling the model to capture complex visual patterns while minimizing computational overhead. Additionally, the utilization of stochastic depth layers enhances model regularization, contributing to improved performance on diverse datasets.

2.5. DenseNet161

DenseNet161 stands as a paradigmatic shift in CNN architectures, distinguished by its dense connectivity pattern that promotes feature reuse and gradient flow propagation throughout the network. Renowned for its superior performance in capturing intricate visual representations, DenseNet161 has become a cornerstone in image classification tasks, particularly in scenarios requiring robust feature extraction. In practical applications, DenseNet161 has demonstrated remarkable efficacy across various domains, including medical image analysis and remote sensing. Its dense connectivity pattern and hierarchical structure facilitate information flow, enabling the model to capture increasingly abstract representations as information traverses deeper into the network.

Architecturally, DenseNet161 comprises a series of dense blocks, each featuring densely connected layers that receive direct input from all preceding layers within the block. This design promotes feature reuse and gradient flow propagation, facilitating efficient training and improved model expressiveness. Additionally, transition layers and global average pooling operations further enhance feature aggregation and classification performance, making DenseNet161 a formidable contender in the realm of image classification architectures.

3. Material and Method

3.1. Dataset Acquisition and Preprocessing

The primary dataset utilized in this study is the SIIM-ISIC Melanoma Classification Competition Dataset sourced from Kaggle. This dataset comprises a comprehensive collection of medical images and corresponding metadata essential for melanoma classification research. The images are primarily dermoscopic images of skin lesions, while the metadata includes critical patient information such as identifiers, sex, age, anatomical site, diagnosis, and malignancy status.

Upon acquisition, the dataset underwent extensive preprocessing to ensure uniformity and compatibility with the model architecture. The images were initially available in multiple formats, including DICOM, JPEG, and TFRecord. To facilitate ease of processing and analysis, all images were converted to a standard format, namely JPEG. Additionally, the images were resized to a uniform resolution of 224x224 pixels to ensure consistency across the dataset.

3.2. Data Augmentation and Transformation

To enhance the diversity and robustness of the training dataset, various data augmentation techniques were applied. Leveraging the transforms module from the PyTorch library, augmentations such as random vertical and horizontal flips were performed on the images. These augmentations help the model generalize better by introducing variations in the training data, thereby reducing overfitting tendencies.

Furthermore, normalization was applied to the image data to standardize the pixel values. This normalization process involved scaling the pixel values to a standard range, typically between 0 and 1, to stabilize the training process and accelerate convergence.

3.3. Proposed Deep Learning Model

Convolutional Neural Networks (CNNs) have become integral in various domains, particularly in computer vision tasks, due to their efficacy in learning and extracting features from visual data. CNNs are adept at discerning intricate patterns and structures within images, making them suitable for tasks like image classification, object detection, and segmentation.

Table 2. Number of tabular and image data taken from SIIM- ISIC Dataset

Class	Total	Training	Testing
Malignant	571	461	110
Benign	579	459	120
Total	1150	920	320

In this study, CNNs are utilized to tackle the crucial task of classifying benign and malignant skin lesions in dermatology. The objective is to leverage CNNs to identify underlying patterns and structures within skin lesion images accurately distinguishing between benign and malignant cases.

The dataset (Table 2) incorporates both tabular data, which consists of patient demographics (sex, approximate age) and lesion location, and image data characterized by three channels and dimensions of 224x224 pixels. The image data undergoes a sequential process, starting with convolutional layers, followed by batch normalization layers, and then Hardwish activation layers from the MobileNet V3 Large Model's Feature Block, to extract pertinent features. Subsequently, it passes through linear, batch normalization, and dropout layers to prepare it for classification. Meanwhile, the tabular data undergoes transformations via multiple linear and ReLU layers. These two sets of data are combined and processed through a linear layer. For classification, the combined data then proceeds through a sigmoid activation function and is rounded as final classification.

3.4. Proposed Deep Learning Model Architecture

Incorporating MobileNetV3's feature extraction mechanism, this model harnesses the efficiency and sophistication of modern CNN architectures. MobileNetV3's feature extractor is pivotal in distilling intricate patterns from skin lesion images. Comprising multiple layers, as shown in Figure 3 MobileNetV3 initiates with a convolutional layer, followed by batch normalization for stabilization and Hardswish activation for non-linearity. This is particularly effective in capturing diverse features across various skin lesion types. Subsequently, MobileNetV3's architecture integrates Inverted Residual blocks, each composed of depthwise separable convolutions and pointwise convolutions. These intricate structures enable the model to efficiently capture spatial hierarchies and semantic information within the image data.

The image features extracted by MobileNetV3 are then further refined through the image classifier component. This classifier comprises multiple layers, starting with a flattening operation to convert the multi-dimensional feature maps into a one-dimensional tensor. Subsequently, the flattened features are processed through a sequence of linear transformations, ReLU activations, batch normalization for feature scaling, and dropout regularization to mitigate overfitting risks.

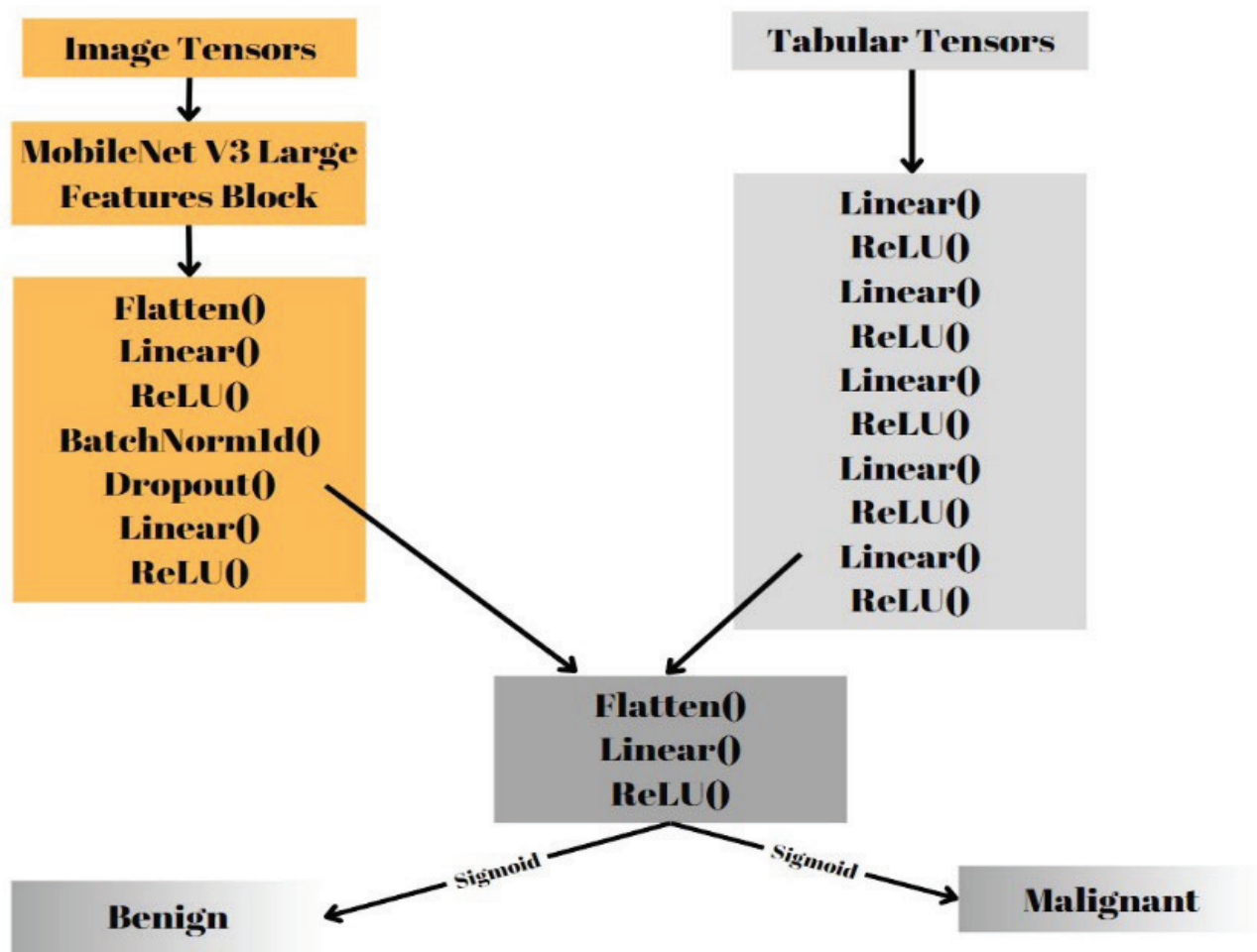


Figure 3. Proposed model architecture

These operations collectively facilitate the learning of discriminative features essential for accurate lesion classification. Concurrently, the tabular data undergoes processing through a dedicated tabular classifier component. Similar to the image classifier, this component comprises multiple fully connected layers augmented with ReLU activation functions. These layers enable the model to capture intricate relationships within the tabular data, leveraging patient demographics and medical history for improved lesion classification accuracy.

Finally, the fused features from both the image and tabular classifiers are concatenated and passed through the fusion layer. This layer consists of a flattening operation followed by a linear transformation. While simpler in structure compared to the classifiers, the fusion layer synthesizes the extracted features from both modalities, facilitating a holistic analysis of the input data and ultimately contributing to enhanced classification performance. By integrating both image and tabular data sources and leveraging sophisticated feature extraction mechanisms and classification components, this model presents a comprehensive approach to skin lesion classification.

3.5. Experimental Setup

For conducting the experiments, a computational setup consisting of an Intel Core i9 processor and an NVIDIA GeForce RTX GPU was utilized. The experiments were executed within the VS Code Jupyter Notebook interface leveraging the PyTorch framework for model implementation. The training process spanned over 35 epochs and took approximately 28 minutes to complete on the aforementioned hardware configuration. The architecture of the proposed methodology is shown in Figure 4.

The experimental setup encompassed the evaluation of multiple neural network architectures. Loss and accuracy curves were generated for each architecture, as depicted in the following figures.

In the study of skin lesion classification, the experimental process is initiated with the collection of two distinct types of data. Image Data, which consist of skin lesion images, and Tabular Data, which encompass patient demographics and clinical information. The data then undergoes a preprocessing stage. This stage is multifaceted and includes the removal of any null values present in the data, the elimination of zero values from the 'age_approx' column, and the custom encoding of the 'sex' and 'anatomical_site_general_challenge' columns.

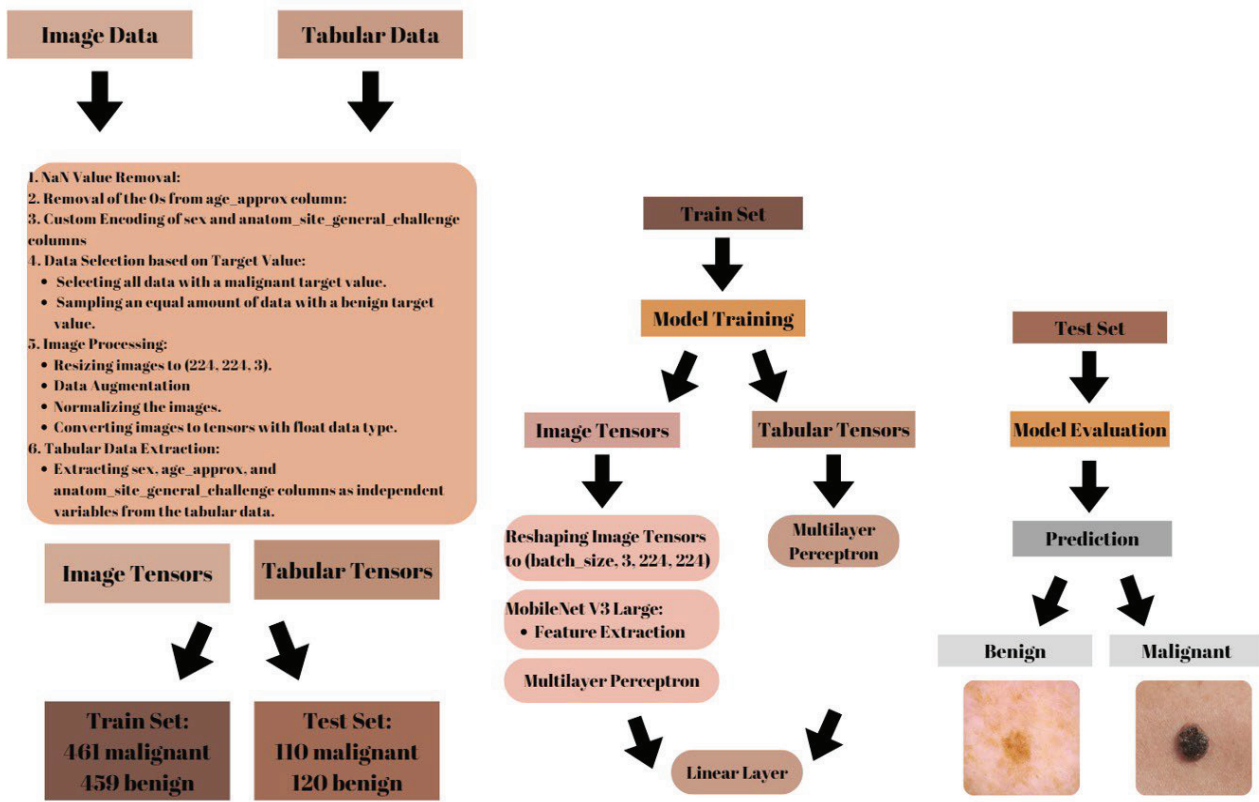


Figure 4. Proposed methodology

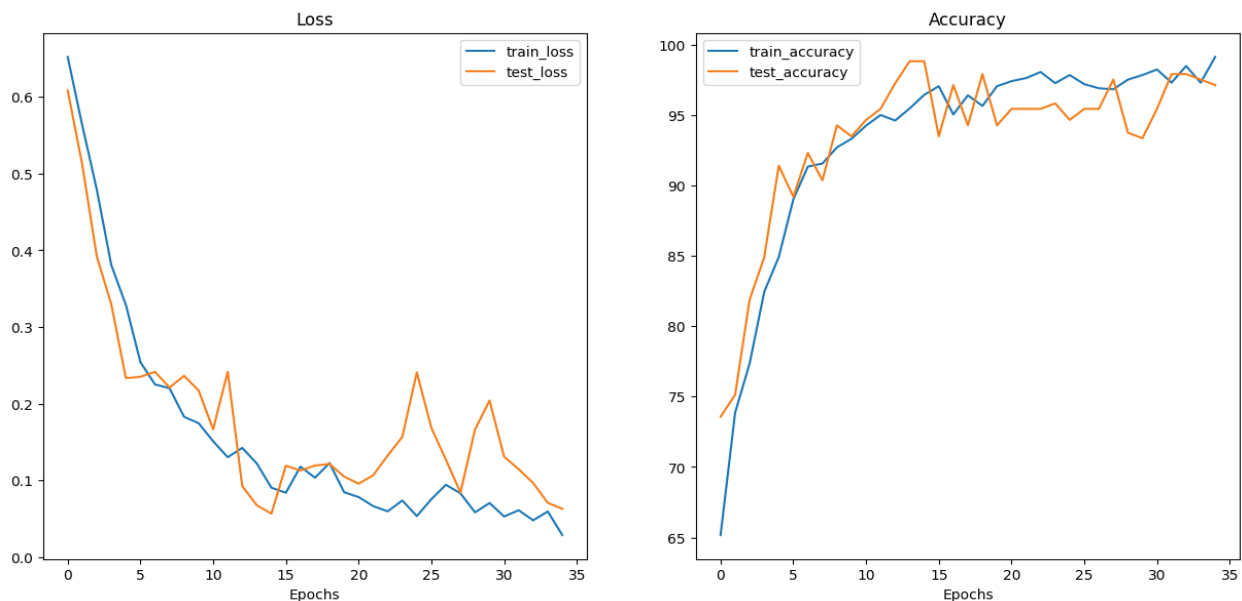


Figure 5. Densenet161 loss and accuracy curves

Additionally, data selection is performed based on the target value, selecting all data with a malignant target value and sampling an equal amount of data with a benign target value. The images are then processed, resized to a standard size of (224, 224, 3), augmented to increase the diversity of the dataset, normalized, and converted to tensors with a float data type.

Concurrently, 'sex', 'age_approx', and 'anatomical_site_general_challenge' are extracted as independent variables from the tabular data. The Image Tensors and Tabular Tensors are then utilized to train the model. The Image Tensors are processed through an MobileNet V3 Large model for feature extraction, while the Tabular Tensors are processed through a Multilayer Perceptron. The training process involved iteratively optimizing the model parameters using the Adam optimizer with a learning rate of 0.01.

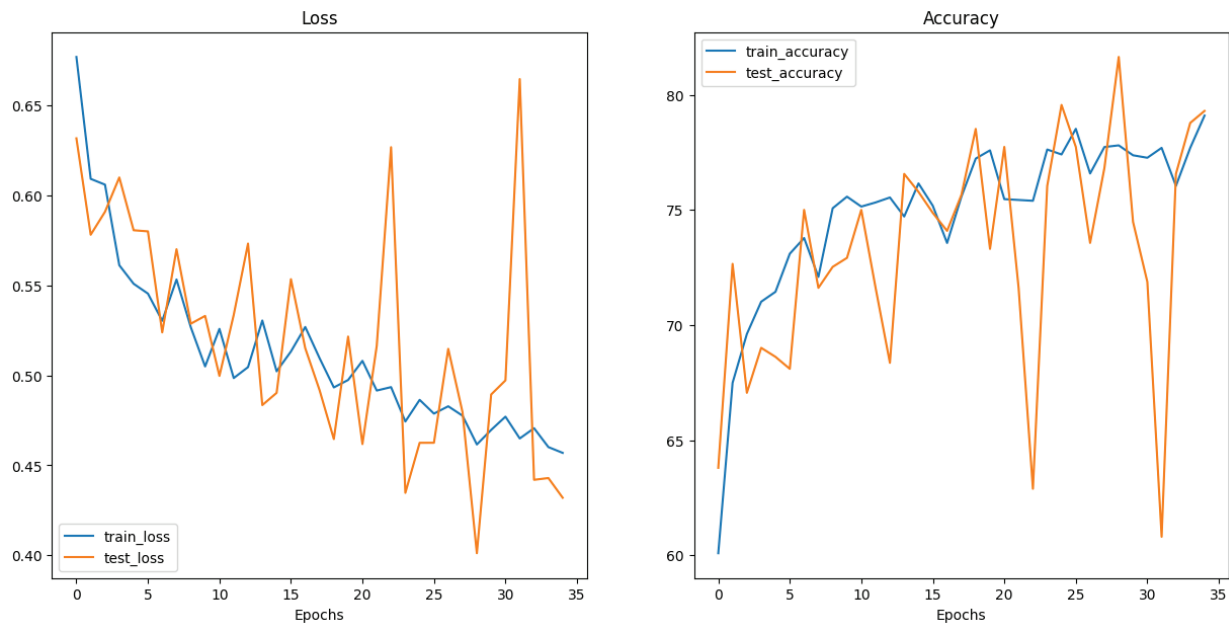


Figure 6. Convnext_base loss and accuracy curves

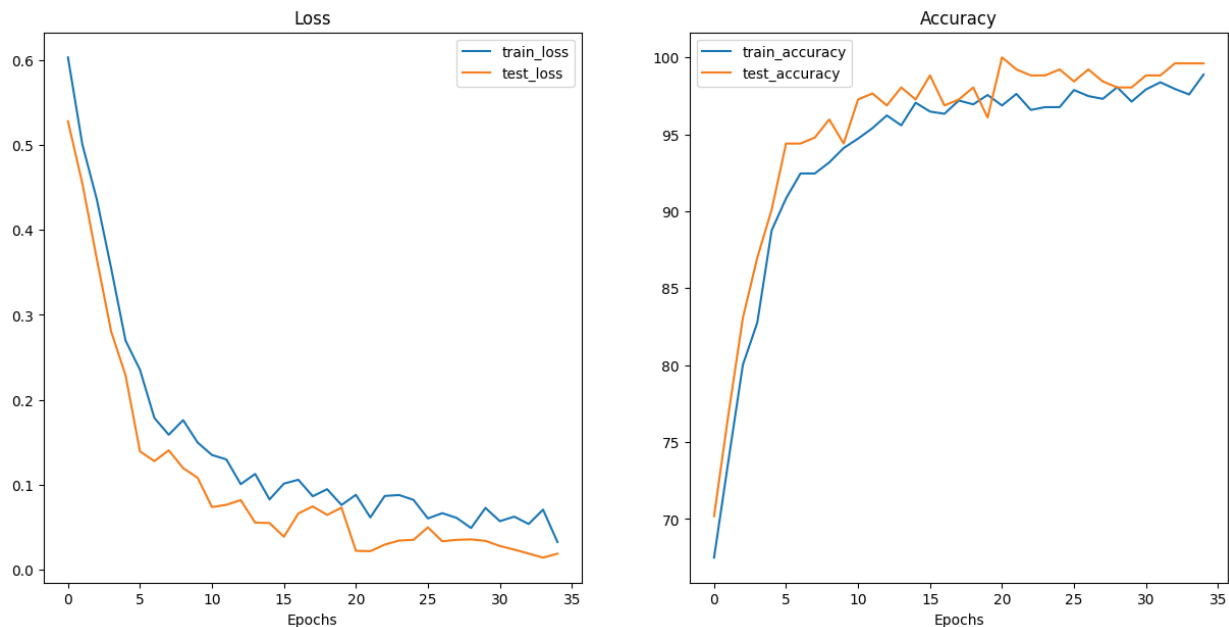


Figure 7. Mobilenet_v3_large loss and accuracy curves

The loss function employed was the Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss), which is well-suited for binary classification tasks such as melanoma classification.

Finally, the model is evaluated using a separate test set, outputting a prediction of either 'Benign' or 'Malignant' for each skin lesion. To evaluate the performance of the trained model, a comprehensive suite of evaluation metrics was employed. The primary metrics used were the confusion matrix and the classification report. The confusion matrix provides a detailed breakdown of the model's predictions, including true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). From the confusion matrix, metrics such as accuracy, precision, recall (sensitivity), and F1-score can be derived.

4. Results and Discussion

The results of this study highlight MobileNet V3 Large as the optimal pretrained model for skin lesion classification, offering superior accuracy and robustness. These findings underscore the potential of deep learning models in enhancing diagnostic accuracy and guiding clinical decision-making in melanoma detection. The results obtained from the evaluation on the test set are summarized in Table 3, while Table 4 provides the corresponding confusion matrix.

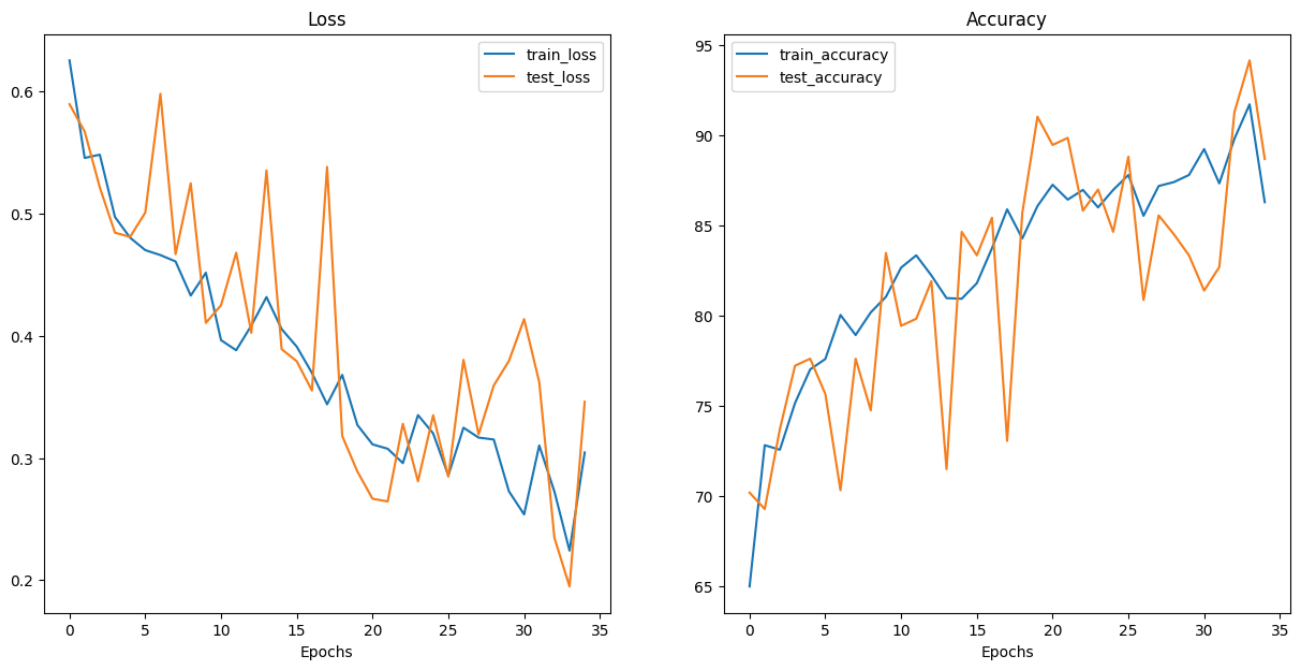


Figure 8. VGG16 loss and accuracy curves

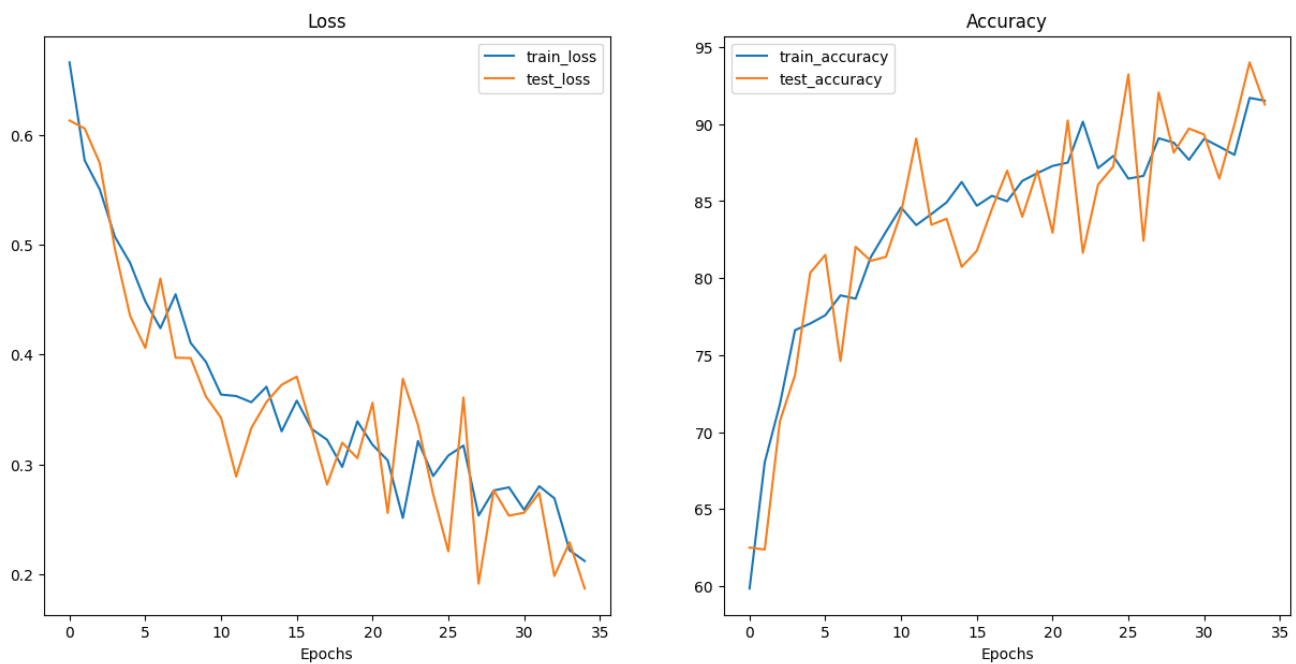


Figure 9. Efficientnet_v2_s loss and accuracy curves

Table 3. Performance on the test set

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Efficientnet_v2_s	91.74	96.33(B)/87.60(M)	87.50(B)/96.36(M)	91.70(B)/91.77(M)
Convnext_base	77.39	82.07(B)/73.38(M)	72.50(B)/82.72(M)	76.99(B)/77.77(M)
Densenet161	98.69	97.56(B)/1.00(M)	1.00(B)/97.27(M)	98.76 (B)/98.61 (M)
MobileNet_v3_large	99.56	1.00(B)/99.09(M)	99.16(B)/1.00(M)	99.58(B)/99.54(M)
VGG16	87.39	85.83(B)/90.83(M)	90.83(B)/85.83(M)	88.26(B)/88.26(M)

This study investigated the performance of different pretrained models as feature extractors for distinguishing benign and malignant skin lesions. Among

the models examined, MobileNet V3 Large demonstrated the highest accuracy, achieving an outstanding 99.56%.

Table 4. Confusion matrix on the test set

Model	TP (B)	TN (M)	FN	FP
Efficientnet_v2_s	105	106	15	4
Convnext_base	87	91	33	19
Densenet161	120	107	0	3
Mobilenet_v3_large	119	110	1	0
VGG16	109	92	11	18

A detailed comparison of the models' performance metrics, encompassing accuracy, precision, recall, and F1-score, is presented in Table 3. MobileNet V3 Large exhibited superior performance across all metrics, indicating its effectiveness in capturing discriminative features from skin lesion images. With precision and recall scores of 1.00 for benign and malignant lesions, respectively, MobileNet V3 Large demonstrated exceptional accuracy in identifying both classes.

EfficientNet V2 Small and VGG16 also yielded competitive results, achieving accuracies of 91.74% and 87.39%, respectively. However, they displayed slightly lower precision and recall compared to MobileNet V3 Large. Conversely, models such as ConvNext Base exhibited relatively lower performance, suggesting limitations in feature extraction from skin lesion images.

The confusion matrices provided in Table 4 offer insights into the models' classification performance, depicting TP, TN, FP, and FN. MobileNet V3 Large achieved the highest TP rate for both benign and malignant lesions, with only one FN and no FP instances. This underscores its robustness in accurate lesion classification, minimizing misclassifications, and mitigating the risk of false diagnoses.

In conclusion, the choice of a pretrained model as a feature extractor significantly impacts the performance of skin lesion classification tasks. MobileNet V3 Large emerged as the optimal model, offering superior accuracy and robustness in distinguishing between benign and malignant lesions. Future research could explore ensemble methods or fine-tuning strategies to further enhance model performance and generalization capabilities.

5. Conclusion

Concluding this study emphasizes the severity of the global skin cancer issue, particularly melanoma, due to its aggressive behavior and potential to spread. Early detection is pivotal in combatting this disease and saving lives. The introduction of a novel approach to melanoma classification utilizing the MobileNet-V3-Large model represents a significant advancement in this field.

By integrating skin lesion images with pertinent patient data such as age, gender, and lesion location, this method demonstrates enhanced predictive capabilities for distinguishing between malignant and benign lesions.

The comprehensive analysis of a vast dataset confirms the superiority of this approach over existing methods, as evidenced by an impressive accuracy rate of 99.56%.

These findings hold promise for revolutionizing melanoma diagnosis and subsequently reducing mortality rates associated with this disease. By harnessing the power of multi-input methodology, clinicians can anticipate more accurate and timely diagnoses, thereby facilitating prompt intervention and improved patient outcomes. This research heralds a new era in melanoma detection, underscoring the potential of innovative technologies to mitigate the impact of this deadly disease on global health.

Future research attempts should prioritize the refinement and optimization of multi-input models, the integration of emerging imaging technologies, and the development of user-friendly AI-driven diagnostic tools. Additionally, extensive validation studies and clinical trials are imperative to confirm the real-world effectiveness of novel diagnostic approaches. Furthermore, there is a pressing need to explore personalized medicine strategies tailored to individual patient characteristics, while simultaneously addressing global disparities in access to advanced diagnostic and treatment modalities. By following these paths, the field can strive towards more accurate diagnosis, personalized treatment regimens, and equitable healthcare access, ultimately leading to improved outcomes and reduced mortality rates for individuals affected by melanoma worldwide.

Author contribution

SA helped in conceptualization, software, validation, writing—original draft, resources and data curation. SA helped in methodology, writing—review & editing, supervision, revision, funding acquisition and formal analysis.

Data availability

All the datasets used in experimentation are available online on the Internet.

Conflict of interest

Authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work submitted for publication.

AUTHOR

Serra Aksoy* – Ludwig Maximilian University of Munich (LMU), Munich, Geschwister-Scholl-Platz 1, 80539, Germany, e-mail: Serra.Aksoy@campus.lmu.de.

*Corresponding author

References

- [1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA A Cancer J Clinicians*, vol. 72, no. 1, pp. 7–33, Jan. 2022, doi: 10.3322/caac.21708.
- [2] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA A Cancer J Clinicians*, vol. 73, no. 1, pp. 17–48, Jan. 2023, doi: 10.3322/caac.21763.
- [3] S. Aksoy, P. Demircioglu, and I. Bogrekci, "Enhancing Melanoma Diagnosis with Advanced Deep Learning Models Focusing on Vision Transformer, Swin Transformer, and ConvNeXt," *Dermatopathology*, vol. 11, no. 3, pp. 239–252, Aug. 2024, doi: 10.3390/dermatopathology11030026.
- [4] A. K. Nambisan et al., "Improving Automatic Melanoma Diagnosis Using Deep Learning-Based Segmentation of Irregular Networks," *Cancers*, vol. 15, no. 4, p. 1259, Feb. 2023, doi: 10.3390/cancers15041259.
- [5] R. Ashraf et al., "Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection," *IEEE Access*, vol. 8, pp. 147858–147871, 2020, doi: 10.1109/ACCESS.2020.3014701.
- [6] M. Elgamal, "Automatic Skin Cancer Images Classification," *IJACSA*, vol. 4, no. 3, 2013, doi: 10.14569/IJACSA.2013.040342.
- [7] F. Xiang, R. Lucas, S. Hales, and R. Neale, "Incidence of Nonmelanoma Skin Cancer in Relation to Ambient UV Radiation in White Populations, 1978-2012: Empirical Relationships," *JAMA Dermatol*, vol. 150, no. 10, p. 1063, Oct. 2014, doi: 10.1001/jamadermatol.2014.762.
- [8] A. Stratigos et al., "Diagnosis and treatment of invasive squamous cell carcinoma of the skin: European consensus-based interdisciplinary guideline," *European Journal of Cancer*, vol. 51, no. 14, pp. 1989–2007, Sep. 2015, doi: 10.1016/j.ejca.2015.06.110.
- [9] D. Koh, H. Wang, J. Lee, K. S. Chia, H. P. Lee, and C. L. Goh, "Basal cell carcinoma, squamous cell carcinoma and melanoma of the skin: analysis of the Singapore Cancer Registry data 1968-97," *Br J Dermatol*, vol. 148, no. 6, pp. 1161–1166, Jun. 2003, doi: 10.1046/j.1365-2133.2003.05223.x.
- [10] L. Fania et al., "Basal Cell Carcinoma: From Pathophysiology to Novel Therapeutic Approaches," *Biomedicines*, vol. 8, no. 11, p. 449, Oct. 2020, doi: 10.3390/biomedicines8110449.
- [11] L. Fania et al., "Cutaneous Squamous Cell Carcinoma: From Pathophysiology to Novel Therapeutic Approaches," *Biomedicines*, vol. 9, no. 2, p. 171, Feb. 2021, doi: 10.3390/biomedicines9020171.
- [12] S. K. T. Que, F. O. Zwald, and C. D. Schmults, "Cutaneous squamous cell carcinoma," *Journal of the American Academy of Dermatology*, vol. 78, no. 2, pp. 237–247, Feb. 2018, doi: 10.1016/j.jaad.2017.08.059.
- [13] M. Rubatto et al., "Classic and new strategies for the treatment of advanced melanoma and non-melanoma skin cancer," *Front. Med.*, vol. 9, p. 959289, Feb. 2023, doi: 10.3389/fmed.2022.959289.
- [14] S. R. Waheed, N. M. Suaib, M. S. Mohd Rahim, M. Mundher Adnan, and A. A. Salim, "Deep Learning Algorithms-based Object Detection and Localization Revisited," *J. Phys.: Conf. Ser.*, vol. 1892, no. 1, p. 012001, Apr. 2021, doi: 10.1088/1742-6596/1892/1/012001.
- [15] S. R. Waheed et al., "Melanoma Skin Cancer Classification based on CNN Deep Learning Algorithms," *Mal. J. Fund. Appl. Sci.*, vol. 19, no. 3, pp. 299–305, May 2023, doi: 10.11113/mjfas.v19n3.2900.
- [16] F. Pereira Dos Santos and M. Antonelli Ponti, "Robust Feature Spaces from Pre-Trained Deep Network Layers for Skin Lesion Classification," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Parana: IEEE, Oct. 2018, pp. 189–196. doi: 10.1109/SIBGRAPI.2018.00031.
- [17] 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, no. 1, p. 180161, Aug. 2018, doi: 10.1038/sdata.2018.161.
- [18] D. Gutman et al., "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)," 2016, doi: 10.48550/ARXIV.1605.01397.
- [19] N. Codella et al., "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)," 2019, doi: 10.48550/ARXIV.1902.03368.
- [20] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and M. H. Yap, "Analysis of the ISIC image datasets: Usage, benchmarks and recommendations," *Medical Image Analysis*, vol. 75, p. 102305, Jan. 2022, doi: 10.1016/j.media.2021.102305.
- [21] M. Combalia et al., "BCN20000: Dermoscopic Lesions in the Wild," 2019, doi: 10.48550/ARXIV.1908.02288.
- [22] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, p. 100034, Dec. 2022, doi: 10.1016/j.neuri.2021.100034.

- [23] K. H. Cheong et al., "An automated skin melanoma detection system with melanoma-index based on entropy features," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 3, pp. 997–1012, Jul. 2021, doi: 10.1016/j.bbe.2021.05.010.
- [24] N. Alfed and F. Khelifi, "Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images," *Expert Systems with Applications*, vol. 90, pp. 101–110, Dec. 2017, doi: 10.1016/j.eswa.2017.08.010.
- [25] Q. Ul Ain, Bing Xue, H. Al-Sahaf, and M. Zhang, "Genetic programming for skin cancer detection in dermoscopic images," in *2017 IEEE Congress on Evolutionary Computation (CEC)*, Donostia, San Sebastián, Spain: IEEE, Jun. 2017, pp. 2420–2427. doi: 10.1109/CEC.2017.7969598.
- [26] R. Garg, S. Maheshwari, and A. Shukla, "Decision Support System for Detection and Classification of Skin Cancer using CNN," 2019, doi: 10.48550/ARXIV.1912.03798.
- [27] T. H. H. Aldhyani, A. Verma, M. H. Al-Adhaileh, and D. Koundal, "Multi-Class Skin Lesion Classification Using a Lightweight Dynamic Kernel Deep-Learning-Based Convolutional Neural Network," *Diagnostics*, vol. 12, no. 9, p. 2048, Aug. 2022, doi: 10.3390/diagnostics12092048.
- [28] S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft-Attention Improves Skin Cancer Classification Performance," 2021, doi: 10.48550/ARXIV.2105.03358.
- [29] M. Abd Elaziz, A. Dahou, A. Mabrouk, S. El-Sappagh, and A. O. Aseeri, "An Efficient Artificial Rabbits Optimization Based on Mutation Strategy For Skin Cancer Prediction," *Computers in Biology and Medicine*, vol. 163, p. 107154, Sep. 2023, doi: 10.1016/j.combiomed.2023.107154.
- [30] C. Xin et al., "An improved transformer network for skin cancer classification," *Computers in Biology and Medicine*, vol. 149, p. 105939, Oct. 2022, doi: 10.1016/j.combiomed.2022.105939.
- [31] S. Gupta, N. Sharma, R. Tyagi, P. Singh, A. Aggarwal, and S. Chawla, "Cognitive Inspired & Computationally-Intelligent Early Melanoma Detection Using Feature Analysis Techniques," *JAIT*, Oct. 2023, doi: 10.37965/jait.2023.0334.
- [32] K. Kachare, N. Bhagat, and P. Raundale, "Advancements in Melanoma Skin Cancer Detection Using Deep Learning: A Comprehensive Review," in *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, Pune, India: IEEE, Aug. 2023, pp. 1–6. doi: 10.1109/ICCUBEA58933.2023.10392174.
- [33] K. Tomar, U. Panwar, and R. Gupta, "A melanoma skin cancer detection and analysis using dermoscopy image processing," presented at the MACHINE LEARNING AND INFORMATION PROCESSING: PROCEEDINGS OF ICMLIP 2023, Ranchi, India, 2023, p. 080007. doi: 10.1063/5.0168906.
- [34] R. Mittal, F. Jeribi, R. J. Martin, V. Malik, S. J. Menachery, and J. Singh, "DermCDSM: Clinical Decision Support Model for Dermatositis Using Systematic Approaches of Machine Learning and Deep Learning," *IEEE Access*, vol. 12, pp. 47319–47337, 2024, doi: 10.1109/ACCESS.2024.3373539.
- [35] S. Aksoy, P. Demircioglu, and I. Bogrekci, "Advanced Artificial Intelligence Techniques for Comprehensive Dermatological Image Analysis and Diagnosis," *Dermato 2024*, vol. 4, no. 4, pp. 173–186, doi: 10.3390/dermato4040015.