Generalised Skin Cancer Detection using Transfer Learning for Real-world Scenarios

Mohamed Dwedar, Fatima Mammadova, Daniel Onwuchekwa, Roman Obermaisser
Chair for Embedded Systems, University of Siegen
Siegen, Germany

***Abstract*—Skin cancer has been recognised as a significant global health concern, with millions of cases diagnosed annually. Early detection, particularly for melanoma, has improved survival rates. In recent years, artificial intelligence (AI) techniques have gained attention in skin cancer classification and have shown potential for automated skin cancer diagnosis. However, the research mainly uses dermoscopic or macroscopic images for AI model training. This research addresses these limitations by developing a generalised deep learning model that integrates dermoscopic and macroscopic image datasets, utilising transfer learning with VGG16 to enhance accuracy and accessibility. The model is designed to generalise features from the macroscopic and dermoscopic perspective, making skin cancer diagnosis more scalable. The model was trained and tested on two diverse datasets, including HAM10000 [15] and PAD-UFES-20 [14], ensuring adaptability to clinical settings and patient populations. We achieved an average Receiver Operating Characteristic Area Under the Curve (ROC AUC) score of 0.97, which indicates a good performance of the classifier. Furthermore, we also test the model in the real-time data application by testing the model's ability to classify the lesion from a web-camera stream for real-world scenarios.**

***Index Terms*—deep learning, embedded systems, transfer learning, artificial intelligence, generalisation, skin cancer**

I. Introduction

Skin cancer is one of the most prevalent forms of cancer globally, significantly impacting public health systems and individuals. The World Health Organization reports that between 2 to 3 million non-melanoma skin cancers and approximately 132,000 melanoma skin cancers occur worldwide annually [1], [2].

In recent years, skin tumour detection and diagnosis have increasingly relied on data processing, mainly through analysing images from existing datasets. These images are fed into Artificial Intelligence (AI) models trained to identify and diagnose diseases based on the provided data. The image types used for the classification are usually macroscopic or dermoscopic. Macroscopic images (or clinical images) are images taken with a digital camera or a smartphone camera. On the other hand, dermoscopic images are captured using a dermatoscope attached to a smartphone or a camera. Macroscopic images provide a broad view of skin lesions, capturing the general surface characteristics. In contrast, dermoscopic images offer a detailed view of the lesion's structures to detect any early signs of cancer. These images allow clinicians to see features like pigmentation, vascular structures, and patterns not visible to the naked eye.

Dermoscopic images require specialised equipment - dermatoscopes and trained professionals to interpret the detailed patterns visible at high magnification. This reliance on specialised tools and expertise makes it challenging to use these methods in regions lacking medical infrastructure. On the other hand, macroscopic images are easier to obtain using standard cameras or smartphones, but they lack the detailed information needed for an accurate diagnosis on their own. As a result, much of the existing research focuses on one type of image, either dermoscopic or macroscopic, limiting diagnostic models' effectiveness. This single-image approach fails to take advantage of the complementary information that both types of images could provide together, which could improve diagnostic accuracy.

In this work, we develop a concept to generalise the diagnostic process. This concept enables using any camera under varying conditions with or without a dermatoscope, thereby expanding the applicability of AI-driven diagnosis. In this study, we explore the enhancement of skin cancer classification by combining dermoscopic and macroscopic image datasets to improve diagnostic accuracy. We employed VGG16, a pre-trained Convolutional Neural Network (CNN), to process these diverse datasets effectively, leveraging its strong feature extraction capabilities. The model is initially trained on the ImageNet dataset and fine-tuned on the macroscopic and dermoscopic datasets. Since each dataset varied and was collected under different conditions, preprocessing was required to standardise the data. This standardisation involved resizing all images to a uniform dimension.

Additionally, each dataset contained class imbalances, where some classes were over-represented compared to others. This imbalance leads to certain classes dominating the model's learning process, resulting in biased performance [12]. Random Over-Sampling technique was employed to address class imbalances in datasets, including HAM10000 [15] and PAD-UFES [14]. An accuracy of 79%, an F1 Score of 79%, and an ROC AUC score of 97% were achieved in classifying skin lesions, demonstrating strong potential in real-time applications.

The rest of this paper is structured as follows. Section II discusses the related works for analysing the state-of-the-art for skin cancer classification. Section III presents the methodology of the work. Section IV and Section V present the AI model's setup and evaluation. Finally, Section VI gives the conclusion of this research work.

II. Related Work

Azhar Imran et al. developed an ensemble Deep Learning (DL) model combining VGG, CapsNet, and ResNet for skin cancer detection [6]. Using the ISIC public dataset, their model achieved a significant predictive accuracy, reaching 93.5%, and outperformed individual models in sensitivity, accuracy, specificity, F-score, and precision.

A study proposes an efficient skin cancer detection system, comparing traditional methods with CNN-based models [7]. They classified the HAM10000 dataset using ResNet50, MobileNet, and an SVM classifier, achieving a notable accuracy of 99.15% with their high-accuracy SVM model.

Mohammed Hasan et al. analysed CNNs to distinguish benign from malignant skin cancers [8]. Among various models, VGG16 achieved the highest accuracy of 93.18%, highlighting the potential of deep learning in improving skin cancer detection accuracy and reducing reliance on traditional biopsy methods.

Another study introduces an on-device CNN application for skin cancer detection, processing 10,015 images locally on mobile devices [9]. This approach enhanced privacy, reduced latency, and saved bandwidth, demonstrating the potential of mobile health (mHealth) applications in medical diagnostics.

The research introduces the E-VGG19 model, which combines VGG19 with traditional classifiers for real-time skin cancer detection [3]. Their model showed substantial improvements in classification metrics and provided valuable insights into integrating deep learning with traditional methods.

A study demonstrated that Deep Neural Networks (DNNs) outperform dermatologists in classifying melanoma from dermoscopic images [18]. The CNN achieved higher sensitivity (82.3%) and specificity (77.9%) compared to dermatologists, suggesting DNNs can enhance melanoma diagnosis.

R.C. Maron et al. evaluated CNNs for multiclass skin cancer classification, showing systematic outperformance compared to 112 dermatologists. The CNN trained on 11,444 dermoscopic images demonstrated higher specificity while maintaining sensitivity [10].

III. Methodology

This section provides a comprehensive overview of the methodologies used in this research, outlining the systematic approach from data acquisition to real-time deployment, depicted in Figure 1. The process begins with data collection from two key datasets, HAM10000 and PAD-UFES, which serve as the primary sources for analysis. Following data acquisition, detailed preprocessing steps are performed to prepare the data for feature extraction using the VGG16 model within a Transfer Learning framework.

The next stage involves training a multiclass classifier to optimise the extracted features for enhanced model performance. Each methodological step is outlined to ensure clarity and understanding of the employed techniques.

After training, the model undergoes rigorous performance evaluation against various metrics to validate its effectiveness and practical applicability. This evaluation is crucial for demonstrating the model's reliability in real-world scenarios.

Finally, the research includes deploying the model in a real-time application, testing its robustness in live environments, and showcasing its potential for practical impact.

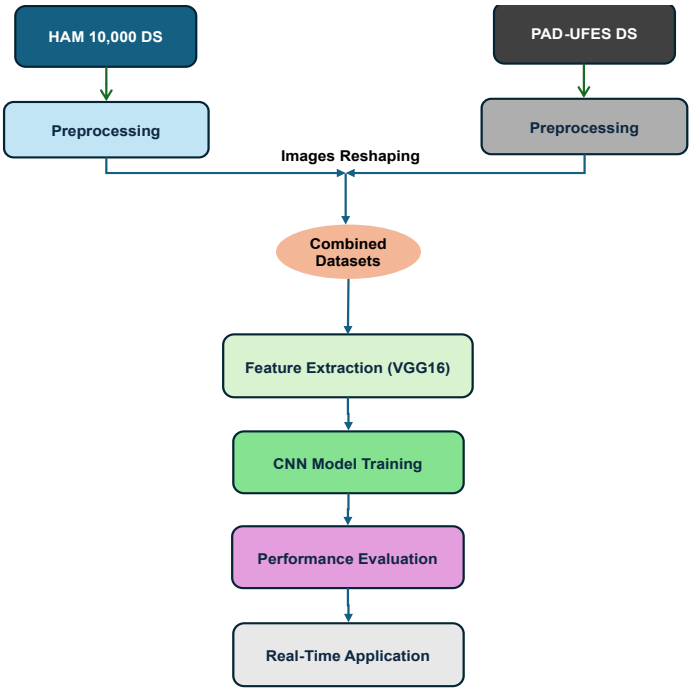

Figure 1. Generalizing Transfer Learning Model Methodology

TABLE I: Original and New Image Sizes for the Datasets.

Dataset Name	Original Size	New Size
HAM10000	Appx. 600x450 pixels	64x64 pixels
PAD-UFES	Appx. 735x734 pixels	64x64 pixels

A. Data Resizing

Data resizing is essential in image processing, especially for machine learning applications like skin cancer detection, where uniform input dimensions are crucial. This process is typically performed using functions like 'cv2.resize' from OpenCV, which scales images to smaller, consistent sizes, ensuring that all model inputs are uniform. Uniformity in image size improves model efficiency, reduces computational demand, and accelerates processing, which is particularly important when training complex models on large datasets [12].

As shown in Table I, resizing images standardises input sizes to 64x64 pixels across different datasets, ensuring consistency and improving the reliability of transfer learning models for skin cancer detection.

B. Data Normalization

Data normalisation in image processing involves scaling pixel intensity values to ensure uniform brightness and contrast across images. This step is crucial in machine learning for

skin cancer detection, as it enhances image quality consistency. Normalisation makes images more comparable, which not only aids in effective data processing by machine learning models but also boosts computational efficiency. By standardising images to a mean of zero and a standard deviation of one, the process aligns the pixel values to a scale optimal for numerical processing, speeding up the training of sophisticated algorithms [12].

C. Datasets Statistical Analyses and Preprocessing

The HAM10000 dataset contains 10,015 dermatoscopic images sourced from diverse populations, acquired and stored using various modalities. These images encompass a wide range of diagnostic categories crucial for training in the realm of pigmented lesions, including:

- Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec)
- Basal cell carcinoma (bcc)
- Benign keratosis-like lesions (solar lentigines, seborrheic keratoses, and lichen-planus-like keratoses, bkl)
- Dermatofibroma (df)
- Melanoma (mel)
- Melanocytic nevi (nv)
- Vascular lesions (angiomas, angiokeratomas, pyogenic granulomas, and haemorrhage, vasc)

An initial analysis of the HAM10000 dataset revealed a pronounced class imbalance. Most of the dataset is dominated by the 'nv' class, representing melanocytic nevi. In contrast, other classes, such as 'df' (dermatofibroma) and 'vasc' (pyogenic granulomas and haemorrhage), are underrepresented, as shown in Figure 2.

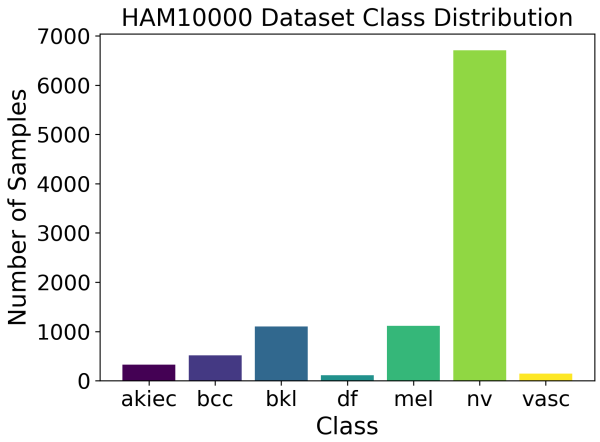

Figure 2. Distribution of the Classes in HAM10000 Dataset.

The PAD-UFES dataset comprises 2,298 samples of six types of skin lesions, each documented with a clinical image and accompanied by up to 22 descriptive features such as age, lesion location, Fitzpatrick skin type, and lesion diameter. This rich dataset includes the following skin lesions:

- Basal Cell Carcinoma (BCC)
- Squamous Cell Carcinoma (SCC), including Bowen's Disease (BOD)
- Actinic Keratosis (ACK)
- Seborrheic Keratosis (SEK)
- Melanoma (MEL)
- Nevus (NEV)

Bowen's disease is classified under SCC as it is considered an in-situ form of this carcinoma. The dataset categorises these lesions into three cancer types (BCC, SCC, and MEL) and three other dermatological conditions (ACK, NEV, SEK). Biopsy confirmation is available for all cancerous lesions, representing about 58% of the samples.

The initial analysis of the PAD-UFES dataset revealed a significant class imbalance, as illustrated in Figure 3. The distribution shows that certain classes are overrepresented, which leads to dataset unbalance. BCC (Basal Cell Carcinoma) class is the most dominant in the dataset, with around 800 instances. ACK (Actinic Keratoses) class also has a high representation, close to 700 instances. NEV (Melanocytic Nevi) and SEK (Seborrheic Keratoses) classes are moderately represented, with around 200-300 instances each. SCC (Squamous Cell Carcinoma) and MEL (Melanoma) are underrepresented classes, particularly MEL, which has fewer than 100 instances. This imbalance can lead to a biased model that performs well on overrepresented classes but poorly on underrepresented ones.

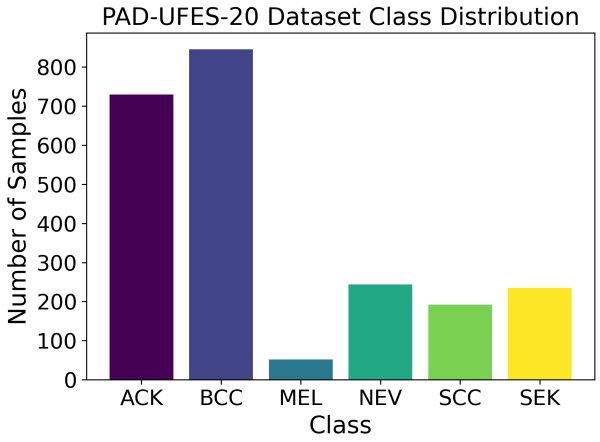

Figure 3. Distribution of the Classes in PAD-UFES-20.

To mitigate the impact of class imbalance on model performance, we employed Random Over-Sampling. This technique aims to balance the class distribution by increasing the number of instances in the minority classes. Specifically, it randomly selects instances from the minority class and duplicates them, thereby increasing the number of samples in those classes. This process is repeated until the class distribution is balanced [16].

Before applying the augmentation technique, we combine the samples from the PAD-UFES and HAM10000 datasets under the classes, which are present in both datasets: BCC & bcc - under the class bcc; ACK & akiec & SCC - under the

class akiec; SEK & bkl - under the class bkl; MEL & mel - under the class mel and NEV & nv - under the class nv.

Figure 4 overviews the combined dataset's class distribution. The figure shows that the 'nv' class included the most samples. Firstly, we undersampled this class until 3000 samples, and then by applying the Random Over-Sampling technique, we kept the same number of samples for each class.

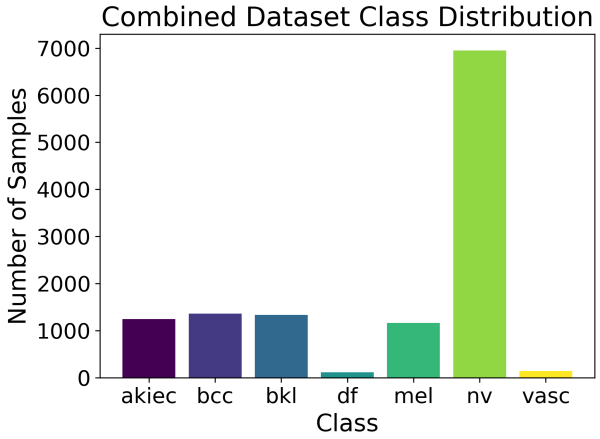

Figure 4. Distribution of the Classes after class-wise combination.

IV. VGG16 Transfer Learning Model Setup

Preparation of the datasets is imperative to effectively apply transfer learning techniques in skin cancer detection. Combining macroscopic and dermoscopic images provides a comprehensive understanding of different skin lesions, enhancing the model's diagnostic accuracy and generalisation. The HAM10000 and PAD-UFES datasets undergo comprehensive preprocessing, including resizing and normalising images to ensure consistent dimensions and pixel intensities. This harmonisation is crucial for maintaining uniformity across the data and enhancing the feature extraction capabilities of advanced models like VGG16.

The VGG16 model is employed as a feature extractor within a CNN pipeline designed for skin cancer detection, leveraging the combined strengths of both datasets to enhance diagnostic accuracy and model reliability.

The VGG16 model, shown in Figure 5, is a CNN initially trained on the ImageNet dataset, which contains millions of labelled images across thousands of categories. The VGG16 architecture comprises 16 layers, including convolutional, max-pooling, and fully connected layers. In this implementation, only the convolutional layers of VGG16 are used for feature extraction.

By setting `include_top=False`, the fully connected layers at the top of the VGG16 model are excluded, allowing the addition of custom fully connected layers suited to the specific task of skin cancer classification.

The base VGG16 model is frozen by setting `base_model.trainable = False`, ensuring that the weights of the pre-trained layers do not change during training. The pre-trained layers act as a fixed feature extractor.

TABLE II: Design Parameters of the Adapted VGG16-Classifier Model.

VGG16-(CNN) Model	**Details**
Input Shape	(64, 64, 3)
Base Model	VGG16 (pre-trained on ImageNet)
Base Model Trainable	No (Frozen)
Custom Layer 1	Global Average Pooling
Custom Layer 2	Dense, 512 units, ReLU activation
Custom Layer 3	Dense, 256 units, ReLU activation
Custom Layer 4	Dense, 128 units, ReLU activation
Output Layer	Dense, Softmax activation
Optimizer	Adam
Learning Rate	0.001
Loss Function	Sparse Categorical Crossentropy
Metrics	Accuracy
Epochs	100
Batch Size	32

The design parameters of the VGG16 model adapted for skin cancer detection are detailed in Table II.

The dataset is divided into training and validation sets with a ratio of 80:20; a learning rate of 0.001 was selected to finely tune the model adjustments during training, ensuring that the weights are updated in a controlled manner to prevent overshooting the minimum loss. With a batch size of 32, the model processes groups of 32 images simultaneously, allowing for efficient gradient estimation without overloading memory. The Adam optimiser was chosen for its effectiveness in handling sparse gradients and its adaptability in learning rates, making it ideal for deep learning tasks involving complex datasets.

The architecture includes layers such as Global Average Pooling and multiple dense layers with ReLU activation functions, strategically incorporated to enhance non-linear learning capabilities and reduce parameter counts, thus minimising overfitting while maintaining model depth for accurate classification. These components and the configuration of the VGG16 model frozen to leverage pre-trained weights without further adjustments during training were all integral in formulating a robust experimental setup to maximise accuracy in skin lesion classification.

V. VGG16 Transfer Learning Model Evaluation

Various metrics were employed to thoroughly assess the VGG16 transfer learning model's performance on the combined HAM10000 and PAD-UFES datasets. Beyond traditional accuracy metrics, the evaluation included F1 Score, ROC AUC (Receiver Operating Characteristic Area Under the Curve), and confusion matrices to provide a comprehensive overview of the model's effectiveness.

A. Evaluation Metrics

A confusion matrix was employed for the model evaluation, which segregates predicted and actual class labels into four

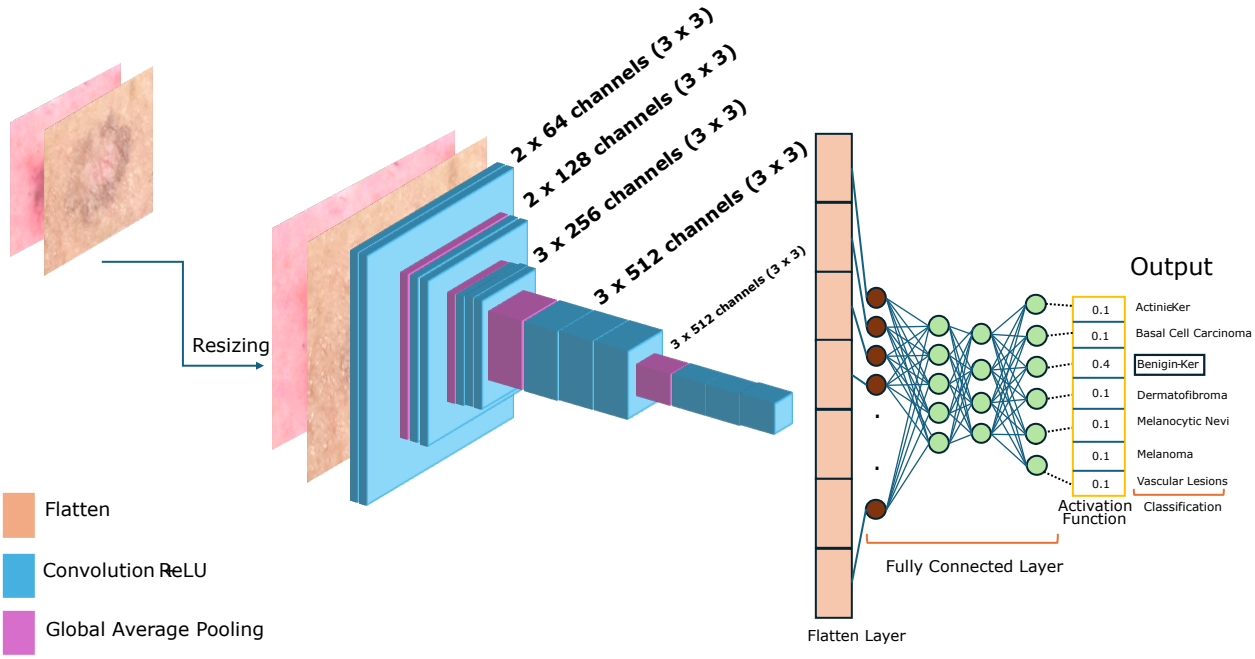

Figure 5. VGG16 Transfer Learning Architecture

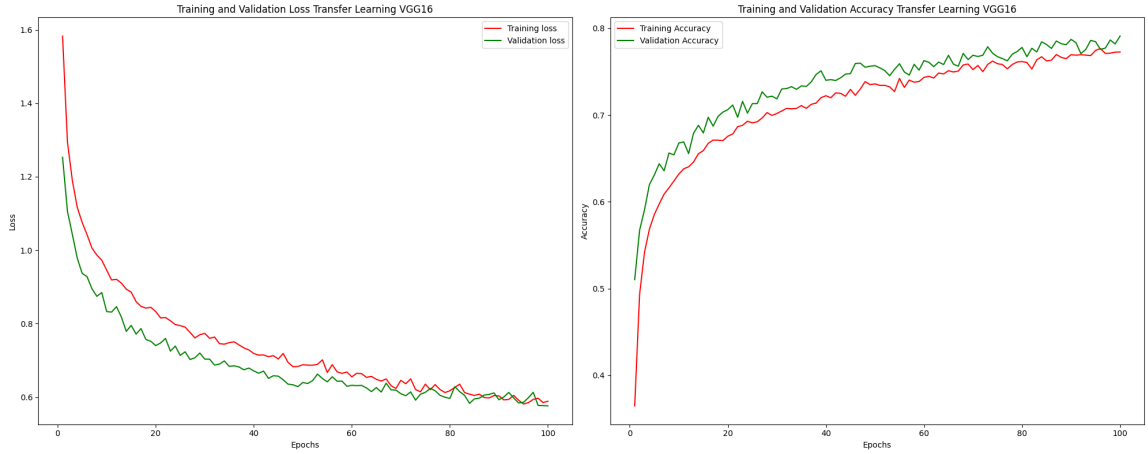

Figure 6. Training and validation loss and accuracy over the epochs for the combined datasets.

critical categories. These categories include:

TP: Instances correctly predicted as positive.
TN: Instances correctly predicted as negative.
FP: Instances incorrectly predicted as positive.
FN: Instances incorrectly predicted as negative.

Accuracy, a primary metric, measures the overall correctness of the model by calculating the ratio of correctly predicted instances to total instances, as shown in Equation 1. Beyond accuracy, other metrics like precision, recall, and the F1 score delve into more nuanced aspects of model evaluation, especially pertinent in datasets with class imbalances [16] [13].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Precision measures the proportion of actual positives among the predicted positives, computed as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Recall assesses the ability of the model to identify all actual positives, as shown here:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

The F1 score, integrating both precision and recall, is a balanced metric that is particularly effective in scenarios of class imbalance. It adjusts for both the false positives and negatives:

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

The metrics of the overall model calculations utilize macro-averaging due to the focus on classifying multiple skin lesions. The metrics for each class were calculated and averaged using the arithmetic mean:

$$M_o = \frac{1}{N} \sum_{c=1}^{N_c} M_c \tag{5}$$

where M_o is the overall model metric, N_c is the number of classes and M_c is the metric for each class. Macro-averaging was chosen over weighted averaging because label imbalance was resolved using augmentation.

This structured approach thoroughly evaluates the model's capabilities, providing a multi-dimensional view of its effectiveness in classifying instances within an imbalanced dataset. These metrics collectively inform the model's strengths and weaknesses, making them essential tools for gauging performance in skin cancer detection models [13].

B. Results

Figure 6 presents the training and validation loss curves, showcasing the model's learning process and generalisation ability. The training loss consistently decreased, indicating effective learning, while the validation loss showed some fluctuations, reflecting the model's adaptability to new data.

The model achieved an accuracy of 79%, indicating that 79% of the predictions made by the model were correct. This accuracy reflects the overall effectiveness of the VGG16 model in correctly classifying the skin lesion images.

The F1 Score, the harmonic mean of precision and recall, also reached 0.79. An F1 Score of 0.79 suggests that the model performs well across different classes without favouring any specific class.

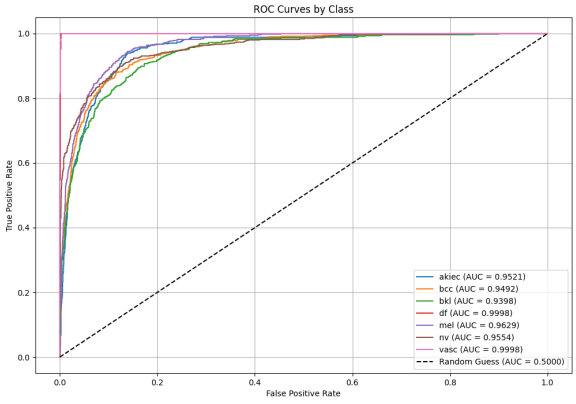

Figure 7. VGG16 Transfer Learning Model Evaluation Metrics

The ROC AUC (Receiver Operating Characteristic Area Under the Curve) score reached a value of 0.97. Figure 7 presents the ROC AUC score for each class predicted by the classifier. A ROC AUC score close to 1.0 indicates excellent discriminative ability, meaning the model is highly effective at differentiating between skin lesions.

The confusion matrix provides further insights into the model's performance for each class.

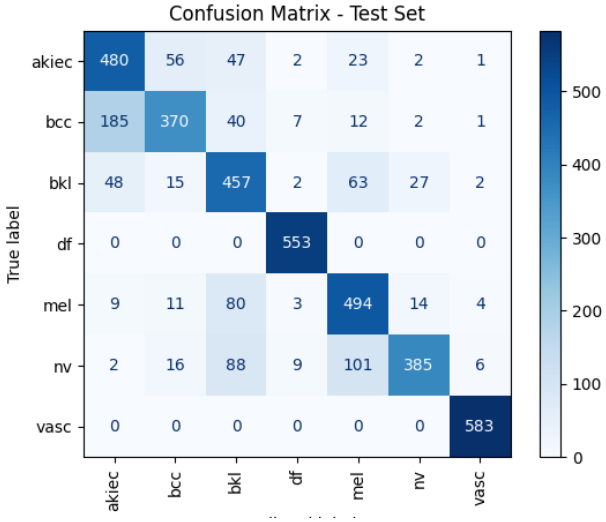

Figure 8. Confusion Matrix for Combined Datasets

Figure 8 shows that the model classified the vascular lesion and dermatofibroma classes with the highest precision. The classes including cancerous lesions - basal cell carcinoma (bcc), melanoma (mel) and squamous cell carcinoma cancer types (akiec) were classified with the F1-scores of 0.72, 0.68 and 0.76, respectively. The model class was classified with the highest precision among the three. The model most confused 'bcc' samples with 'akiec' samples, whereas 'nv' was confused with 'mel' 101 times. Benign keratosis-like lesions (bkl) were most confused with the class mel and" akiec", where the F1-score reached 0.69.

C. Discussion on Generalized Real-Time Application and Evaluation

To effectively apply real-time applications for skin cancer detection using live video streams, establishing a system setup and ensuring an AI model's robust implementation is essential. This real-time system utilises the classifier to process live video captured through a web camera.

The operation begins with initiating a web camera using the OpenCV library, which consistently captures video frames while the camera remains active. Each frame is subjected to a sequence of image processing techniques outlined below:

1) **Skin Detection:** Initially, the frame is converted into the HSV colour space, followed by applying colour segmentation to isolate regions of skin based on specific chromatic thresholds. This step focuses the processing on relevant areas likely to contain skin lesions.

2) **Contour Detection:** Post-segmentation, the algorithm searches for contours within the isolated skin regions. The most significant detected contour is presumed to be the region of interest (ROI), potentially encapsulating a lesion.
3) **Region of Interest (ROI) Processing:** The identified ROI is then cropped from the frame, resized to the CNN's required input resolution of 64x64 pixels, and normalised to ensure consistent lighting and colour distribution for accurate classification.

Upon processing the ROI, it is an input into the pre-trained CNN model tailored to classify captured skin lesions. Afterwards, the model categorises the captured lesion based on learned patterns from training data.

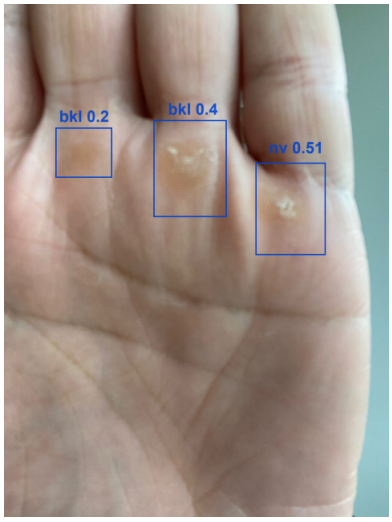

Figure 9. Real-time Detection using webcam

Figure 9 presents the preliminary results of the lesion detection and classification using an HD 1080p web camera and a DELL Latitude 5330 system. The application can detect the lesions and classify them. However, while capturing the skin, we noticed that lighting and camera focus variations significantly impact the accuracy of skin lesion classification.

VI. Conclusion

This work has analysed the application of the VGG16 Transfer Learning model for effectively detecting skin lesions across macroscopic and dermoscopic datasets. The primary objective was to enhance the skin lesion classification by investigating a generalised model trained on macroscopic and dermoscopic images. The approach overcomes the challenges of applying limited AI models with specific purposes, e.g., prediction only on dermoscopic images. A generalised model helps diagnose and adapt to different capture tools, clinical conditions, and environments.

The high ROC AUC score (0.96) underscored the models' ability to discriminate between different lesion classes, which is crucial for precise clinical diagnostics. The VGG16 Transfer Learning model combined features from both datasets and achieved an accuracy of 0.79, demonstrating its capacity to leverage pre-trained networks to improve learning from dermatological and macroscopic images. The model showed confidence in classifying lesions like vascular lesions and dermatofibromas while facing most challenges in classifying basal cell carcinomas. The results show the importance of using more various datasets in the future, which include classes with balanced samples and better model tuning.

The real-time application shows the initial results for skin lesion detection and the possible application of the clinical setup. However, inconsistent lighting and varying camera focal points can obscure or distort essential features of skin lesions, resulting in misclassification or false negatives. Differences in contour detection capabilities also influence how well the models delineate lesion boundaries, which is vital for accurate classification and effective treatment planning. The results will be used to improve the initial detection framework and to test on real cases for better assessment.

The real-time application of AI models for skin cancer diagnosis is a good tool for assessing professionals during clinical examinations and accelerating the diagnosis. However, these models can be heavy-weighted and require high computational resources. Field-programmable gate arrays (FPGAs) are widely used for AI model deployment, as they are highly energy efficient, consume less memory, and are faster than central processing units (CPUs) or graphical processing units (GPUs). We further plan to explore the possibilities of the model running on an FPGA, benefiting higher energy efficiency and faster diagnosis.

Acknowledgements

This research work was supported by KMU-innovativ, Germany, in the scope of the MoleVision project, with the funding code 13GW0588D. The Authors M. Dwedar and F. Mammadova contributed equally to this work.

References

[1] World Health Organization, "Skin Cancer Facts & Figures," *World Health Organization Reports*, 2023. Available online: https://www.who.int.
[2] American Cancer Society, "Cancer Facts and Statistics," *American Cancer Society Reports*, 2023. Available online:https://www.cancer.org.
[3] Kandhro, I., Manickam, S., Fatima, K., Uddin, M., Malik, U., Naz, A. & Dandoush, A. Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification. *Heliyon*. **10**, e31488 (2024),https://www.sciencedirect.com/science/article/pii/S2405844024075194
[4] N. C. F. Codella *et al.*, "Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging," in *Proceedings of the International Symposium on Biomedical Imaging*, 2017, pp. 1–5, doi: 10.1109/ISBI.2017.7950597.
[5] T. J. Brinker, A. Hekler, R. C. Enk, C. S. Berking, F. Haferkamp, and M. J. U. Klode, "Deep neural networks are superior to dermatologists in melanoma image classification," *European Journal of Cancer*, vol. 119, pp. 11–17, 2019, doi: 10.1016/j.ejca.2019.06.012.
[6] Imran, A., Nasir, A., Bilal, M., Sun, G., Alzahrani, A. & Almuhaimeed, A. Skin Cancer Detection Using Combined Decision of Deep Learners. *IEEE Access*. **10** pp. 118198-118212 (2022)

[7] L. I. Mampitiya, N. Rathnayake, and S. De Silva, "Efficient and Low-Cost Skin Cancer Detection System Implementation with a Comparative Study Between Traditional and CNN-Based Models," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 3, pp. 226–235, 2022. Available online: https://ojs.bonviewpress.com/index.php/JCCE/article/view/482, doi: 10.47852/bonviewJCCE2202482.

[8] Hasan, M., Fatemi, M., Monirujjaman Khan, M., Kaur, M. & Zaguia, A. Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks. *Journal Of Healthcare Engineering*. **2021**, 5895156 (2021)

[9] Dai, X., Spasic, I., Meyer, B., Chapman, S. & Andres, F. Machine Learning on Mobile: An On-device Inference App for Skin Cancer Detection. (2019,6)

[10] R. C. Maron *et al.*, "Evaluating Convolutional Neural Networks for Multiclass Skin Cancer Image Classification: Outperforming 112 Dermatologists," *Journal of Medical Imaging*, 2023. CNNs achieved higher specificity at equal sensitivity rates compared to dermatologists.

[11] G. Amaro *et al.*, "Comprehensive Evaluation of Multi-omics Data Integration Methods for Subtyping Cutaneous Melanoma Using the Cancer Genome Atlas Dataset," *Cancer Research Journal*, 2023. Found that multi-omics integration did not significantly outperform single-domain analyses.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016. Discusses methods for data preprocessing and standardisation in neural networks.

[13] P. I. Good, *Resampling Methods: A Practical Guide to Data Analysis*, 3rd ed., Birkhäuser, 2006. Available online: https://link.springer.com/book/10.1007/978-0-387-39582-9, doi: 10.1007/978-0-387-39582-9.

[14] Pacheco, A., Lima, G., Salomão, A., Krohling, B., Biral, I., De Angelo, G., Alves Jr, F., Esgario, J., Simora, A., Castro, P., Rodrigues, F., Frasson, P., Krohling, R., Knidel, H., Santos, M., Do Espírito Santo, R., Macedo, T., Canuto, T. & De Barros, L. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data In Brief*. **32** pp. 106221 (2020), https://www.sciencedirect.com/science/article/pii/S235234092031115X

[15] Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset is a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*. **5**, 1-9 (2018)

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. Discusses techniques for addressing class imbalance.

[17] Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. (2016), https://arxiv.org/abs/1506.02640

[18] Brinker, T., Hekler, A., Enk, A., Berking, C., Haferkamp, S., Hauschild, A., Weichenthal, M., Klode, J., Schadendorf, D., Holland-Letz, T., Von Kalle, C., Fröhling, S., Schilling, B. & Utikal, J. Deep neural networks are superior to dermatologists in melanoma image classification. *European Journal Of Cancer*. **119** pp. 11-17 (2019), https://www.sciencedirect.com/science/article/pii/S0959804919303491