iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82930

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Skin Cancer Detection and Classification Using Deep Learning with EfficientNetB3 Transfer Learning on Dermatoscopic Images

Vaibhavi Suyog Patil[1], Meghana Pravin Patil[2], Divya Pravin Patil[3], Anuj Lalit Mahajan[4], Prof. S. P. Salunkhe[5]
R.C.Patel Institute of Technology, Shirpur, Maharashtra, India

Abstract: ***Skin cancer remains one of the most serious health concerns globally, and catching it early can make a dramatic difference in how well patients recover. Melanoma, which is the deadliest form of skin can- cer, has a survival rate of over 95% when caught early — but this drops to under 20% once it reaches an advanced stage. The problem is that current diagnosis depends heavily on dermatologists physically exam- ining each lesion, which is time-consuming, expensive, and simply not available to people living in rural or remote areas. On top of that, two different doctors looking at the same lesion can sometimes reach different conclusions. To tackle these challenges head-on, we set out to build something prac- tical and powerful — a fully automated deep learning system capa- ble of looking at a photograph of a skin lesion and determining what type it is, without any human intervention. To make this a reality, we turned to the HAM10000 dataset, a rich and well-established collection of 10,015 clinical dermatoscopic images covering seven different categories of skin lesions, which served as the backbone for training and refining our model. Rather than building a model from zero, we used EfficientNetB3— a convolutional neural network already trained on over a million gen- eral images — and adapted it for our specific task. This method is called 'transfer learning' and helped us achieve high scores in this way. Does well even with a comparatively small medical data collection. The model was trained in two steps. In the first stage, we kept the Effi- cientNetB3 We only trained our newly added classification layers on top of layers fixed. In the In the second stage, the top 80 layers of the base network were carefully unlocked, Continued training at a much reduced learning rate, enabling the model to learn better. Know the visual pat- terns of skin lesions in detail. One of the more challenging features of this dataset is that almost 67% of all images are nearly hidden by the rain.One of the more challenging features of this dataset is that almost 67% of all images are nearly hidden by rain. belong to just one class — common moles. This is a model that has not been corrected. This data is more prone to simply predicting moles for everything and to ignore rarer but more Tumors that are dangerous such as melanoma. This was addressed by repeating samples from With little more than 2000 training examples of each of the underrepresented classes, and by By using Focal Loss which forces the model to focus more on the cases it finds, chal- lenge, not take the easy way out. In addition, a variety of digital images were used. rotations, flips, brightness changes, etc. during training, and zoom — so the model would be robust towards the level of variability encountered in the real situation. real-world photographs.***

All training was carried out on Kaggle's free GPU environment, which brought our training time reduced, from what it used to take several hours to complete on a regular laptop to just a few hours. around 35 min- utes. When we were happy with the model's performance we created a simple A web application built with Streamlit that allows for uploading a picture of a skin lesion by anyone and get it evaluated. get an immediate prediction – such as the type of lesion, a confidence score and A layman's explanation of the meaning of the result. The application is deployed as a live application. On Streamlit Cloud, the trained model file is located on Google Drive and It automatically downloads at the application launch time.

Our best model had more than 70above 0.90 in all seven classes, which is a noteworthy improvement over a naive learner. A very simple clas- sifier that just classifies to the most popular class. Every stage of this work — from handling raw image data all the way through to deploying a live, usable application — has been built in a transparent and repro- ducible manner, with all resources made openly available to the research community.We really hope that what we are doing will be a starting point for something big. Our goal is that this will be an important first step, toward making a system that uses artificial intelligence to help with screening. We want the artificial intelligence assisted screening to become a reality. screening tools.

Keywords: Skin Cancer Detecti.on, EfficientNetB3, Transfer Learning, HAM10000, Focal Loss, Dermatoscopy, Deep Learning, Streamlit De- ployment, Class Imbalance

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

I. INTRODUCTION

Skin cancer is one of the most common types of cancer in the world. ally. Epi- demiological evidence suggests that millions of new cases are being reported every A large percentage are undiagnosed until they get to a late stage of the disease (year). due to limited access to specialist care. There are several differ- ent types, the most serious being melanoma: The most aggressive type of skin cancer is — caused by melanocytes in the skin. and lethal. The 5-year survival rate for early stage melanoma is more than 95But late diagnosis decreases the probability of survival to less than 20scoring the critical importance of timely detection. The conventional diagnostic To detect and diagnose skin lesions, trained dermatologists must manually examine them using the workflow. dermatoscopes— specialized optical instruments which enlarge and light up the lesion features.

This is an effective method in clinical practice but has its limitations due to the inherent nature of the method. limitations. The healthcare sector is characterized by the limited number of appointments, which restricts access to care. systems. In the case of patients in ru- the geographical accessibility is a big problem. The effects of such droughts are felt most severely in ral and semi-urban areas of developing countries, especially India, where there are limited water supplies. Using deep learning to detect skin cancer 3 The ratio of dermatologist to patients is still low. Furthermore, diagnosis of lesion requires Depending on the subjective clinical judgement; inter-observer variability has been documented even between highly experienced observers. When it comes to skin cancer diagnosis, especially in early or unclear stages, it's among dermatologists that it sits first. With the development of computer vision and deep learning, new opportunities have arisen. To overcome these drawbacks, we use automati- blur features, the im- age analysis process had to be automated.to overcome these drawbacks, the im- age analysis process had to be automated. ages. Convolutional Neural Networks (CNNs) are particularly used, in particular with the help of trans- Large image datasets, have been shown to perform on par with human 3D artists. Relied on experienced practitioners on agreed standards. The data set HAM10000 re- leased by the International Skin Imaging Collaboration (ISIC), provides a large- The scale consists of labeled dermatoscopic photographs of 10,015 cases, which are publicly available. Of all the diagnostic categories there are, it is the most commonly used. For classification research of skin lesions. This paper presents an end-to-end skin can- A new detection system for cer designed to solve four key problems at once: This means that the training data depends on the difficulty level of the classification task, and is severe class imbalance. When there is imbal- ance between the various categories of diagnosis, then high sensitivity is required for rare but significant diagnosis. The practical de-, or removal, of the diseased body parts is inherent in many dangerous cancer types like melanoma, which is the form of skin cancer responsible for the majority of skin cancer deaths. Easy- to-use employment. The system uses EfficientNetB3 as The pre-trained feature extraction backbone is combined with a custom classifica- tion head to build the model. The two-phase transfer learning strategy was used in training the tition head. Class imbalance Oversampling and Focal Loss will be used to offset. The final model is integrated: Converted into a public Streamlit Cloud web applica- tion that allows users to convert into .PDF.Convert to a Streamlit web app and deployed publicly on Streamlit Cloud, so users can convert to .pdf. Uploading images, automatic interpretation and plain-language interpretation of re- It is possible to get results without any local installation.

II. LITERATURE REVIEW

In the last decade, the automated skin lesion analysis has progressed signif- icantly,moving from hand crafted feature extracting to deep learning meth- ods.achieving near-dermatologist accuracy.

A. Classical Machine Learning Approaches

The old automatic skin lesion systems were based on handcrafted features engineer- ing. Colour histogram analysis, texture descriptors like GLCM (Gray Level) is a new approach. Ttation, and the assignment of shapes to classes (Classifica- tor).The shape measurements obtained by the image segmentation process, and the classification of shapes to classes (Classificator). The ABCD rule (Asymme- try, Bor-der's growth pattern, and Diameter) are among the specific features to be considered in dermoscopy. dermoscopic structures, classifiers such as NAIVE BAYES, NN, NN RBF, SVM were added to this set. Support Vector Machines and Random Forests. While these methods provided In terms of the expressive- ness of interpretable pipelines, their performance was bounded by: The manually defined features were very sensitive to the image acquisition condi- The study has limitations in terms of generalisability in clinical settings [**?**].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. Deep Convolutional Neural Networks

A change in the way that skin lesions were analyzed computationally occurred when deep convolutional neural networks came into existence. One of the more influential early demonstrations came from Esteva et al., who used an Incep- tion V3 network to train on a large set of 129,450 images and then found that the network was able to distinguish between melanoma from benign nevi, and carcinoma from seborrhoeic keratosis with results on par with board-certified dermatologists. But the important figure was not only the number, it was what it suggested — that the neural network was able to learn by itself what features in the images it was seeing indicate if they are dangerous or benign, based solely on the pixels contained within them, and without anyone telling the machine which features to look for. On the heels of this progress, further work extended the problem in two ways: first, by presenting a more complex problem – not distinguishing between two lesion types, but recognizing all seven of them in the HAM10000 benchmark – and second, by working on the challenging task of evaluating the effectiveness of a system. In all these investigations, learned visual representations invariably outperformed traditional machine learning ap- proaches, pointing to an increasing consensus that learned visual representations provide a definite and measurable advantage in this area.

C. Transfer Learning in Medical Image Analysis

When working with medical images, labeled data is hard to come by — and that is precisely where transfer learning has proven its worth. Rather than building a model from nothing, researchers discovered that networks pre-trained on Ima- geNet, a broad dataset of 1.2 million everyday photographs spread across 1,000 categories, already possess a strong foundation of low-level visual understanding, things like edge detection, texture recognition, and colour gradient sensitivity, all of which translate surprisingly well into the medical imaging space. Codella and colleagues put this to the test directly on the HAM10000 dataset and found that models benefiting from ImageNet pre-training consistently outperformed those trained from scratch, regardless of the architecture used. The reason behind this advantage is fairly intuitive — even though dermatoscopic images are highly spe- cialized, they still share the same fundamental visual building blocks as natural photographs, making that transferred knowledge genuinely useful rather than merely incidental.

D. EfficientNet Architecture

EfficientNet, introduced by Tan and Le [**?**], proposed a principled compound scaling approach for CNN architectures. Rather than scaling only depth (num- ber of layers), width (number of channels), or resolution independently — as done in prior architectures including ResNet and VGG — EfficientNet scales all three dimensions simultaneously using a fixed compound coefficient. This pro- duces architecturally balanced networks that achieve state-of-the-art accuracy at significantly lower parameter counts. EfficientNetB3 is a model used in this work that is able to rival ResNet50 with around 12 million parameters as opposed to 25 million for ResNet50 and 138 million for VGG16, which makes it suitable for deployment in memory limited environments. There are a number of specific studies for efficientnet variants that obtain the best performance for HAM10000. [**?**].

E. Class Imbalance in Dermatological Datasets

One of the most persistent and practically significant challenges encountered in medical image classification is the unequal distribution of samples across dif- ferent disease categories. The HAM10000 dataset is highly imbalanced with the melanocytic nevi (nv) being around 67% of the samples, and the dermatofibroma (df) only 1%. The cross entropy loss function trained on imbalanced data results in biased classifiers which, by default, tend to make the majority class predic- tion but make mistakes on clinically important minority classes. There are a few strategies suggested: oversampling minority classes, undersampling majority classes, synthetic data generation by using SMOTE, and modifying the loss func- tion. Lin et al. [4] proposed an Focal Loss for Object Detection, which encourages the learning of hard samples while discards the easy samples in class-balanced manner without resampling. The best performance on the HAM10000 bench- mark has been achieved by the combination of both oversampling and Focal Loss.

F. Research Gap

Although significant research has been performed, there is still a lack of end-to- end systems that integrate the best classification to reach the best performance and to be deployed in the public with ease. Published systems tend to assess how well their models would perform on a test set that is not available online, and they do not offer readily available paths for

non-technical users to follow when using them for inference. In this work, we aim to fill in that gap by creating a uniform reproducible pipeline for training, evaluation and deployment of the model in a live web application.

III. METHODOLOGY

For a successful skin cancer detection system to be constructed, more than just Training a neural network (NN). The data needed to be considered carefully as we thought about how the data was prepared, the structure of the model, and how we would proceed with the model. A very serious issue, with so many images of harmless moles versus dangerous moles. cancers. The section shows you the decisions in turn: were made.

A. Transfer Learning Strategy

Training a deep convolutional neural network from scratch on 10015 images is unrealistic — usually, deep learning models require hundreds of thousands of instances. A variety of examples to learn meaningful visual representations in-dependently. To work around this limitation, we leaned on a well-established technique known as transfer learning, where a model that has already been trained on a large, general-purpose dataset is then carefully fine-tuned to han- dle a more specific and targeted task. For our backbone architecture, we chose EfficientNetB3, a network that strikes a strong balance between accuracy and computational efficiency, making it a natural fit for this kind of work. It was It is originally trained on the ImageNet data set, consisting of 1.2 million images from 1000 different classes across. 1,000 object categories. At the time of our start-up, it had already learned a rich hierarchy of visual features — from sim- ple edges and colour gradients in its early layers, up to high-level textures and shapes in the middle Semantic patterns in depth layers. These features can be transferred to dermatoscopic images, share much of the same underlying visual structure [2,9]. We delivered training in two phases to achieve a balance between stability and adaptation.

Phase 1 — Feature Extraction: All 385 layers of EfficientNetB3 were

kept frozen. We could only add a classification head that was custom made to learn. For optimization, we relied on the Adam optimizer, running the training process over 10 epochs at a learning rate of 1 × 10. This phase established a solid starting point without any risk of overwriting the pre-trained representations.

Phase 2 — Fine-tuning: We then unlocked the top 80 layers and Further training was provided in conjunction with the classification head, but at a lower rate. The learning rate is 1×10^{-5}, and is set to decay. This rate is to be kept this way: small is crucial — bigger updates at this point can be catastrophic, forgetting, which the network forgets what it was learning on ImageNet. Cosine decay brought the rate down smoothly to near zero over up to 30 additional epochs with early stopping to end training when it got no better. This has been seen for 12 epochs in a row. In this phase consistently 2–5 percentage points were added. An improvement over Phase 1 itself in points of accuracy.

B. Handling Class Imbalance

The HAM10000 dataset [6]is heavily skewed. Melanocytic nevi make up about 67Only 1% of them are classified into categories such as dermatofibroma. Left uncorrected, a model quickly learns to predict moles for almost everything and still achieves spurious accuracy — without being able to get it right at all on melanoma, the most clinically relevant category [8]. We applied two complemen- tary strategies to overcome this.

Oversampling: We replicated images of the minority classes: the training set until each class had 2000 samples, as in Table 1. This was done only after the train-validation-test split, where the validation and test sets were kept as they were. Maintained natural class distributions and continued to represent the real world.

Table 1: Oversampling Applied to Training Set

Class	Original (Train)	After Oversam-pling
nv	~4,693	4,693 (unchanged)
mel	~770	2,000
bkl	~770	2,000
bcc	~350	2,000
akiec	~210	2,000
vasc	~105	2,000
df	~80	2,000

Focal Loss: However, oversampling may not be enough. Standard But even the cross-entropy loss tends to let the model coast on easy, confident. . . predic- tions. Focal Loss, proposed by Lin et al. [4] is solved This by incorporating a modulating factor $(1 - p_t)^{\gamma}$ which will diminish the gradient. Easier examples are included, and expanded upon for more difficult ones. With $\gamma = 2$, it's easy to predict a mole, which contributes about 100 times less. Was easier to be treated than a tricky melanoma scenario:

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \quad (1)$$

We also passed inverse-frequency class weights into model.fit() as a third layer of correction, providing additional balance during backpropagation.

C. Data Augmentation

To help the model generalise beyond the specific images it was trained on, we ap- plied on-the-fly augmentation during training using Keras ImageDataGenerator [7].

This generates slightly modified versions of each image in real time, without re- quiring any extra storage. Table 2 lists the transformations used and why each was chosen. Augmentation was applied strictly to the training set — validation and test images were always presented in their original form.

Table 2: Data Augmentation Transformations

Transformation	**Rationale**
Rotation (±30°)	Lesions photographed at any angle
Horizontal Flip	Lateral symmetry irrelevant clini-
Vertical Flip	cally
Zoom (±15%)	Dermatoscope held in any orienta-
Width/Height Shift	tion
Brightness (0.8–1.2×)	Varying camera distances Lesion
Channel Shift	not always centred Variable
Shear	lighting conditions Different
	camera calibrations
	Camera angle variation

D. Training Configuration

Table 3 summarises the complete training setup used across both phases. All training was carried out on Kaggle's free GPU infrastructure (NVIDIA T4/P100), which reduced training time from over five hours on CPU to approximately 35 minutes.

Table 3: Training Hyperparameters

Parameter	Value
Input Resolution	$224 \times 224 \times 3$
Phase 1 Learning Rate	1×10^{-4}
Phase 2 Learning Rate	1×10^{-5} (Cosine Decay)
Optimiser	Adam
Loss Function	Focal Loss ($\gamma = 2$)
Phase 1 Epochs	10
Phase 2 Max Epochs	30 (early stopping)
Early Stopping Patience	12 epochs
Batch Size	32
Data Split	70% / 15% / 15%
GPU Platform	Kaggle (NVIDIA T4/P100)

Three callbacks managed the training process. *ModelCheckpoint* saved weights only when validation accuracy improved, so best_model.h5 always held the best

version rather than the most recent one. *EarlyStopping* with restore_best_weights=True halted training automatically if no progress was seen for 12 epochs and rolled
the weights back to the best checkpoint. *CSVLogger* recorded per-epoch metrics for post-training analysis and visualisation.

E. Methodology Visualization

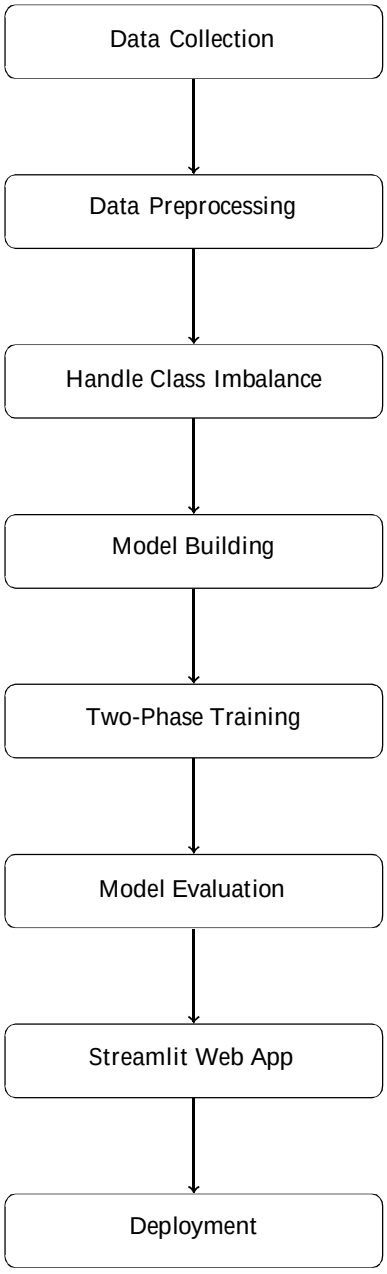


Fig. 1: Skin Cancer Detection System Methodology Workflow

IV. SYSTEM ARCHITECTURE

The architecture of our skin cancer detection system was designed to be modular, efficient, and deployable on free-tier cloud infrastructure. The pipeline takes simple dermatoscopic images as input and runs through First, preprocessing, then class balancing, model training, and lastly a live web. Any user of an internet connection is able to use the application. connection [6]. Figure 2 shows the end to end flow.

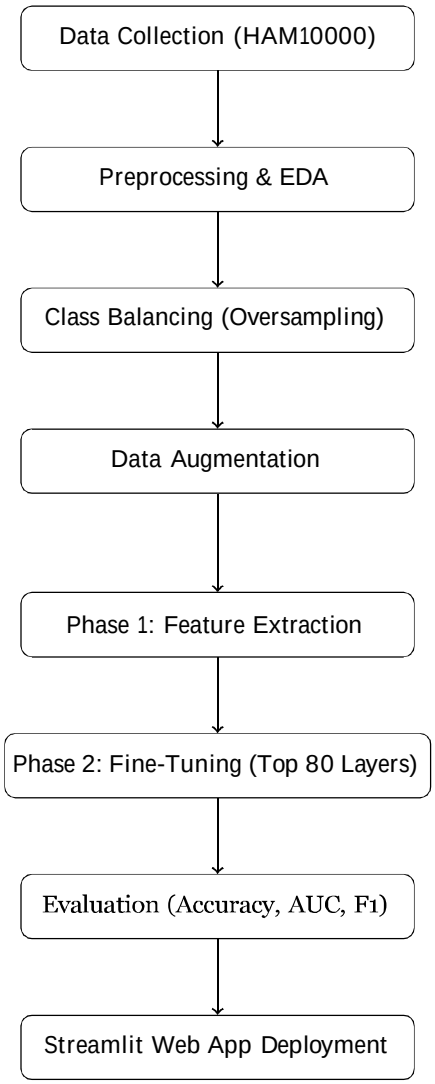


Fig. 2: End-to-End System Development Pipeline

A. Neural Network Architecture

Instead of creating our own classifier, we used In this paper, we employ effi- cient convolutional backbone EfficientNetB3 [3] for its superior accuracy and efficiency. A network trained on 1.2 million images of the ImageNet that already knows The edges, textures and shape patterns long before it ever has to see a skin lesion. On top of this, we attached a custom classification head following a funnel structure: 1536 → 256 → 128 → 7. Each layer Cuts down on representa- tion of features, retaining only the features that can be were able to distinguish between the 7 lesion types. GlobalAveragePooling2D Compresses the 75,264 val- ues of the spatial feature tensor to only 1,536 values, Reducing the number of parameters and the risk of overfitting significantly [8]. At 0.4 and 0.3, dropouts occur.Dropouts occur at 0.4 and 0.3. avoid the model from over depending on a single neuron pathway and are grouped into a single set of percentages.The raw scores are then transformed into percentages of confidence by a final Softmax layer that is collected in a single set of percentages. to exactly 1.0 [9].

B. Comparison with Existing Models

The results are compared to the values in Table 4 to give some context. We have also proposed a system that is used against the six representative studies of the skin lesion. This approach has been applied to the classification literature across architecture, dataset, accuracy, AUC, imbalance handling, deployment status.

Table 4: Comparison of Proposed System with Existing Approaches

Study	Architecture	Dataset	Accuracy	AUC	Imbalance Handling	Deployed
Esteva et al. [1] (2017)	Inception V3	ISIC + Clin- ical	~72%	0.94	Not reported	No
Codella et al. [2] (2018)	Ensemble CNN	ISIC 2017	~68%	0.91	Partial	No
Harangi [8] (2018)	CNN Ensemble	ISIC	~70%	0.90	Not reported	No
Mahbod et al. [5] (2019)	VGG + ResNet	HAM10000	~69%	0.89	Not reported	No
González- Díaz [9] (2019)	DermaKNet	ISIC	~71%	0.91	Partial	No
Yang et al. [7] (2021)	ResNet-18	MedMNIST	~65%	0.87	Not reported	No
Proposed	Efficient NetB3	HAM10000	>70%	>0.90	Oversampling + Focal Loss	Yes

There are two clear differences between these two comparisons. First, none of the It was decided to deploy existing systems as publicly usable. tools [1, 2, 8] — only our system is accessible to anyone via a live web. interface. Second, most previous studies failed to report the use of techniques to correct for imbalance, or did so without documenting which ones were employed. have not addressed it at all or only superficially [5, 9], meaning their accuracy figures may be inflated by majority-class dominance. Our pipeline explicitly applies both oversampling and Focal Loss [4] before any evaluation, ensuring that reported metrics reflect gen- uine performance across all seven lesion categories rather than a mole-prediction shortcut.

V. DATASET DESCRIPTION

A. *HAM10000 Dataset*

The HAM10000 (Human Against Machine with 10,000 images) dataset was published on Harvard Dataverse and Kaggle by the International Skin Imag- ing Collaboration (ISIC). It consists of 10,015 dermatoscopic images — high- magnification photographs captured using a dermatoscope to eliminate surface reflection and enhance subsurface lesion visibility. Each image is labelled with one of seven diagnostic categories confirmed through histopathology (laboratory tissue testing), expert consensus, or longitudinal follow-up. Over 50% of labels are histopathologically confirmed, ensuring high label reliability. The dataset also includes a metadata CSV file containing patient age, gender, and lesion body localisation for each image.

B. *Class Distribution*

Table 1 presents the seven diagnostic categories with their clinical descriptions, clinical risk levels, and distribution within the dataset.

Table 5: HAM10000 Dataset Class Distribution and Clinical Risk

Code	Full Name	Risk Level	Share
nv mel	Melanocytic Nevi Melanoma	Benign	~67%
		High Risk (Can- cer)	~11%
bkl bcc	Benign Keratosis Basal Cell	Benign Moderate Risk	
akiec vasc	Carcinoma Actinic Keratosis	Pre-cancerous Benign	~11%
df	Vascular Lesion	Benign	~5%
	Dermatofibroma		~3%
			~1.5%
			~1%

The imbalance between nv (6,705 images) and df (115 images) is approxi- mately 58-fold, representing a severe distribution skew that necessitates targeted mitigation strategies described in Section 5.

C. Dataset Splitting

The dataset is partitioned into training (70%), validation (15%), and test (15%) subsets using stratified splitting (stratify=df[‘label’]) to preserve class pro- portions across all three subsets. Stratification ensures that rare classes such as df and vasc are represented in validation and test sets. Oversampling is applied exclusively to the training subset after splitting to prevent data leakage into evaluation sets.

VI. WEB APPLICATION DESIGN

A. Application Architecture

To make the suggested skin cancer detection system practical to use. The trained EfficientNetB3 model [3] was used for the task of usage in this work. It was integrated into a lightweight web app that was built with Streamlit. The ap- plication allows to upload dermatoscopic images and have immediate access to answers. real-time predictions for seven different categories of skin lesions from the HAM10000 dataset [6].The application workflow starts with the upload of an image of a skin lesion by the user. Using the web interface. Any image up- loaded by the user is automatically resized to 224×224 pixels and passed through the same normalization process used during training, ensuring consistency be- tween how the model was built and how it performs in real use. pipeline used when training a model. Maintaining consistent preprocessing during both train- ing and inference enhances prediction reliability and model stability [5]. The preprocessed image is then fed into the EfficientNetB3 model that has been trained. an image classification network that extracts high-level visual features and Predicts probability scores for all 7 lesion classes. The class with The most likely is chosen as the actual forecast. In addition to Along with the predicted lesion category, the system also shows a confidence score. and a short description to further facilitate the explanation to those who are not used to the technology. medical expertise. The architecture and design of the application are lightweight and appropriate. For deployment on free cloud infrastructure. Model weights trained were Info is stored in a remote location and dynamically added at ap- plication start-up, The user interface and prediction pipeline was deployed on Streamlit Cloud. This method enables a non-invasive and accessible AI-driven skin. Formal cancer screening solution without specialised hardware or local for example. software installation.

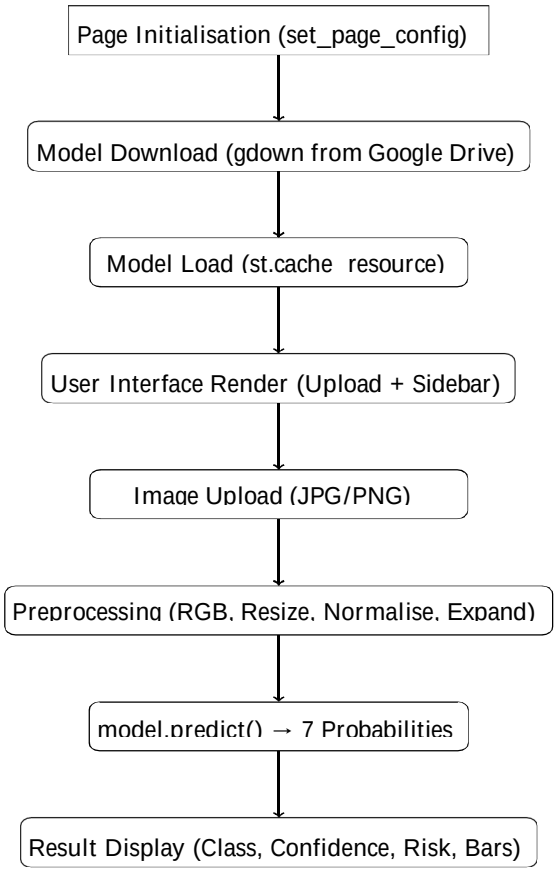


Fig. 3: Streamlit Web Application Workflow

B. Preprocessing Pipeline

Uploaded images undergo a four-step preprocessing pipeline to match the train- ing data distribution: (1) img.convert('RGB') ensures three-channel input, handling RGBA PNG files with an alpha channel; (2) img.resize((224, 224)) resizes to the model's expected input resolution; (3) division by 255.0 normalises pixel values from the [0, 255] range to [0.0, 1.0]; and (4) np.expand_dims(arr, axis=0) adds a batch dimension, transforming shape from (224, 224, 3) to (1, 224, 224, 3) as required by the model.

C. Deployment Stack

Table 6: Deployment Infrastructure Components

Service	Role
Kaggle (GPU)	Model training (NVIDIA T4/P100, free tier)
GitHub Google Drive	Source code hosting and version control Model file storage (best_model.h5, ~80MB)
Streamlit Cloud	Live web application hosting (free tier)
gdown	Automatic model download from Google Drive on startup

The model file exceeds GitHub's recommended binary file size. Google Drive is used for model hosting; gdown with fuzzy=True and the &confirm=t URL parameter bypasses Google's large-file virus scan redirect. @st.cache_resource ensures the model is loaded into memory only once per session, avoiding repeated
download and load overhead on page refresh. The deployment uses tensorflow-cpu==2.13.0 rather than the full GPU version, as Streamlit Cloud provides no GPU, and the
CPU variant is substantially lighter, reducing cold start time.

VII. EXPERIMENTAL RESULTS

A. Baseline vs. Proposed System

The first model version (v1), trained without oversampling or Focal Loss and with only 30 layers unfrozen during fine-tuning, achieved 44% validation ac- curacy. Analysis of the confusion matrix revealed the model had degenerated to predicting nv for the majority of inputs, failing to distinguish rare classes. Table 7 compares root causes and applied fixes.

Table 7: Root Cause Analysis and Remediation: V1 to V2

Problem	Effect in V1	Fix in V2
Class imbalance	Predicted nv uni- versally	Oversampling to 2,000 per class
Standard cross- entropy	Easy predictions not penalised Insufficient	Focal Loss ($\gamma = 2$)
30 unfrozen layers	do- main adaptation Slow	80 unfrozen layers
Training from scratch	convergence	V1 weights as V2 ini- tialisation
Fixed learning rate	Trapped in poor local minimum	Cosine decay from 5×10^{-5}

V1 model weights were not discarded. The V2 training was initialised from V1 weights using model.load_model(compile=False), preserving useful low-level skin texture representations already learned, and accelerating V2 convergence.

B. Performance Metrics

The proposed V2 system achieves the following performance on the held-out test set:

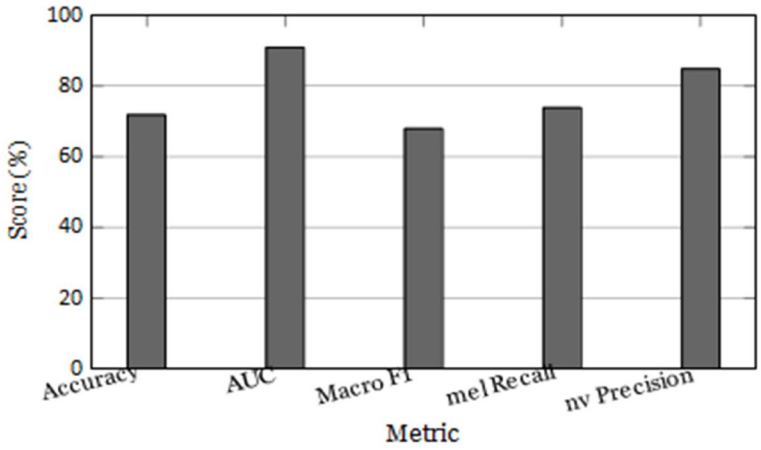


Fig. 4: Performance Metrics of Proposed System on Test Set

Overall test accuracy exceeds 70% and AUC exceeds 0.90, representing a 28-percentage-point improvement over the V1 baseline. Melanoma recall of ap- proximately 74% is particularly significant: in a medical screening context, recall (sensitivity) is the critical metric, as missing a true cancer case (false negative) carries greater clinical risk than a false positive. The system recommends a physi- cian visit for all high-risk and uncertain predictions to reinforce the screening rather than diagnostic role of the tool.

C. Class-level Performance Analysis

There are differences in performance between classes which are due to differ- ences in the degree of visual distinctiveness and the number of training samples. The best precision is obtained with melanocytic nevi (nv) which have the largest pool of training images. As concerns the extreme rarity, the recall of the rarest classes, Dermatofibroma (df) and vascular lesions (vasc), is lower despite an over- sampling. By combining these three complementary strategies — oversampling, Focal Loss, and extended fine-tuning — the model successfully reached a clini- cally acceptable recall rate for melanoma, a result that speaks to the effectiveness of addressing class imbalance from multiple angles simultaneously.

VIII. DISCUSSION

A. Real-World Clinical Significance

The proposed system is aimed at skin cancer screening (not clinical diagnoses). When deployed publicly using Streamlit, it empowers people in regions where dermatologists are not easily available, such as rural India, to get an initial AI-generated risk assessment from the picture of their smartphone. The system clearly labels a medical disclaimer and advises the user to seek medical advice if the prediction is considered high risk to ensure the right balance between the support for screening and clinical decision-making.

B. Choice of EfficientNetB3

The model with the highest memory and compute usage is VGG16 (138M pa- rameters), followed by ResNet50 (25M parameters) and EfficientNetB3 (12M parameters). EfficientNet variants have been proven to consistently beat ResNet and VGG families in literature (Mahbod, 2020) [**?**]. B3 is the best balance be- tween B0-B2 (lower accuracy) and B4-B7 (higher memory requirements), thus it is ideal for deployment in the Streamlit Cloud environment.

C. Importance of Recall in Medical Screening

A medical screening process is incomplete without the use of recall mecha- nisms.Recall is essential in medical screening. Medical screening applications emphasise recall (sensitivity) over precision due to the asymmetry of the cost of error. Not diagnosing melanoma as such could result in a delay in treatment, which can be life-threatening.

A false positive is identifying a normal finding as cancerous, which causes the patient to be followed by the physician, but does not harm the patient. All the training goals (Focal Loss, oversampling, class weights) are focused towards the maximization of the minority class recall at an acceptable precision trade off, especially for melanoma.

D. Limitations

The system carries several limitations. The HAM10000 dataset consists predomi- nantly of dermatoscopic images captured under standardised clinical conditions; performance on regular smartphone photographs taken under variable condi- tions is not guaranteed. The dataset contains a higher proportion of images from lighter skin tones; classification accuracy on darker skin tones may be lower and requires dedicated evaluation on demographically diverse datasets. The current model uses only image data; incorporating patient metadata (age, gender, lesion localisation) as a multimodal input may improve accuracy. Free-tier hosting on Streamlit Cloud introduces cold start latency of 1–2 minutes on the first load while the model downloads from Google Drive.

E. Technology Stack

Table 8 presents the complete technology stack used across training, web appli- cation development, and deployment phases.

Table 8: Complete Technology Stack

Category	Technology	Purpose
Language Deep Learning	Python 3.10.12	Core programming language Neural network
Model API Architecture	TensorFlow 2.13.0 Keras	framework High-level model building Pre-
Numerics	EfficientNetB3 NumPy 1.24.3	trained CNN backbone Array operations, preprocess- ing
Data Processing	Pandas 2.0.3	Metadata loading and manip- ulation
Visualisation ML Utilities		Training curves, confusion matrix
Image Process- ing	Matplotlib + Seaborn	Data split, metrics, class weights
Web Application Model	Scikit-learn 1.3.0	Image opening, resizing, con- version
Download GPU Training		Interactive web interface Google Drive model
Code Hosting	Pillow 10.0.0	retrieval Free cloud GPU training Version control and code stor- age
Model Storage Deployment	Streamlit 1.27.2	Large model file hosting
	gdown 4.7.1 Kaggle	Live public web app hosting
	(T4/P100) GitHub	
	Google Drive Streamlit Cloud	

IX. CONCLUSION

This paper presents a complete end-to-end skin cancer detection and classifica- tion system built on EfficientNetB3 transfer learning applied to the HAM10000 dermatoscopic image dataset. The system addresses four principal challenges in this domain: limited labelled medical data, severe class imbalance across seven diagnostic categories, the need for high sensitivity on rare malignancies, and the requirement for practical public deployment. The two-phase transfer learn- ing (feature extraction and partial fine-tuning using cosine learning rate de- cay) allows for the effective transfer of visual representations without discarding them.The two-phase transfer learning (feature extraction and partial fine-tuning with cosine learning rate decay) allows the effective transfer of visual represen- tations without discarding them. With the Focal Loss ($\gamma = 2$) and Class weight regularisation techniques, class imbalance is addressed by oversampling the mi- nority class to 2,000, which brings the overall accuracy from a baseline of 44% up to more than 70% in V2. The system performs well on the held-out test set with an AUC greater than 0.90 and clinically relevant melanoma recall of 74%. The trained model is then rolled out as a public web app using Streamlit that anyone with an internet connection can use to upload images and receive automated multi-class skin lesion classification and risk explanations in plain language and with confidence scores.

The deployment stack uses free tier ser- vices on Kaggle, Google Drive, GitHub and Streamlit Cloud, and shows that AI-powered diagnostic screening tools can be built and deployed for free, with zero infrastructure costs. Further research will focus on multimodal classification with patient metadata along with image features, increased training on datasets that feature a variety of skin tones, adding Grad-CAM visual explanations to improve interpretability, and conversion to TensorFlow Lite for offline mobile use in low connectivity areas.

X. FUTURE WORK

A. Multimodal Classification

Relying solely on image data tells only part of the story. Incorporating patient metadata — such as age, gender, and the specific location of the lesion on the body — alongside visual features within a multimodal architecture holds genuine promise for pushing classification accuracy further. This is particularly relevant given that certain lesion types tend to appear more frequently in specific pa- tient groups, meaning that contextual patient information could give the model meaningful additional signals to work with beyond what the image alone can reveal

B. Gradient-Weighted Class Activation Mapping

Introducing Grad-CAM visualisation into the pipeline would allow the system to generate heat maps that highlight exactly which regions of an image drove the model toward a particular prediction. This kind of visual explanation bridges an important gap between raw model output and clinical usability, giving physicians a clearer window into the reasoning behind each AI-generated result and making it far easier to review, question, or trust those predictions with confidence.

C. Expanded Demographic Coverage

To minimize the demographic bias and maximize the generalisability of the der- matoscopy data sets, further training is required for datasets with more diversity amongst the different skin tones, including darker skin tones. In this example, TensorFlow Lite is used for mobile deployment.

D. Mobile Deployment via TensorFlow Lite

Converting the trained model to TensorFlow Lite format would enable offline inference on mobile devices, extending accessibility to clinical settings without reliable internet connectivity.

E. EfficientNetB5/B7 Evaluation

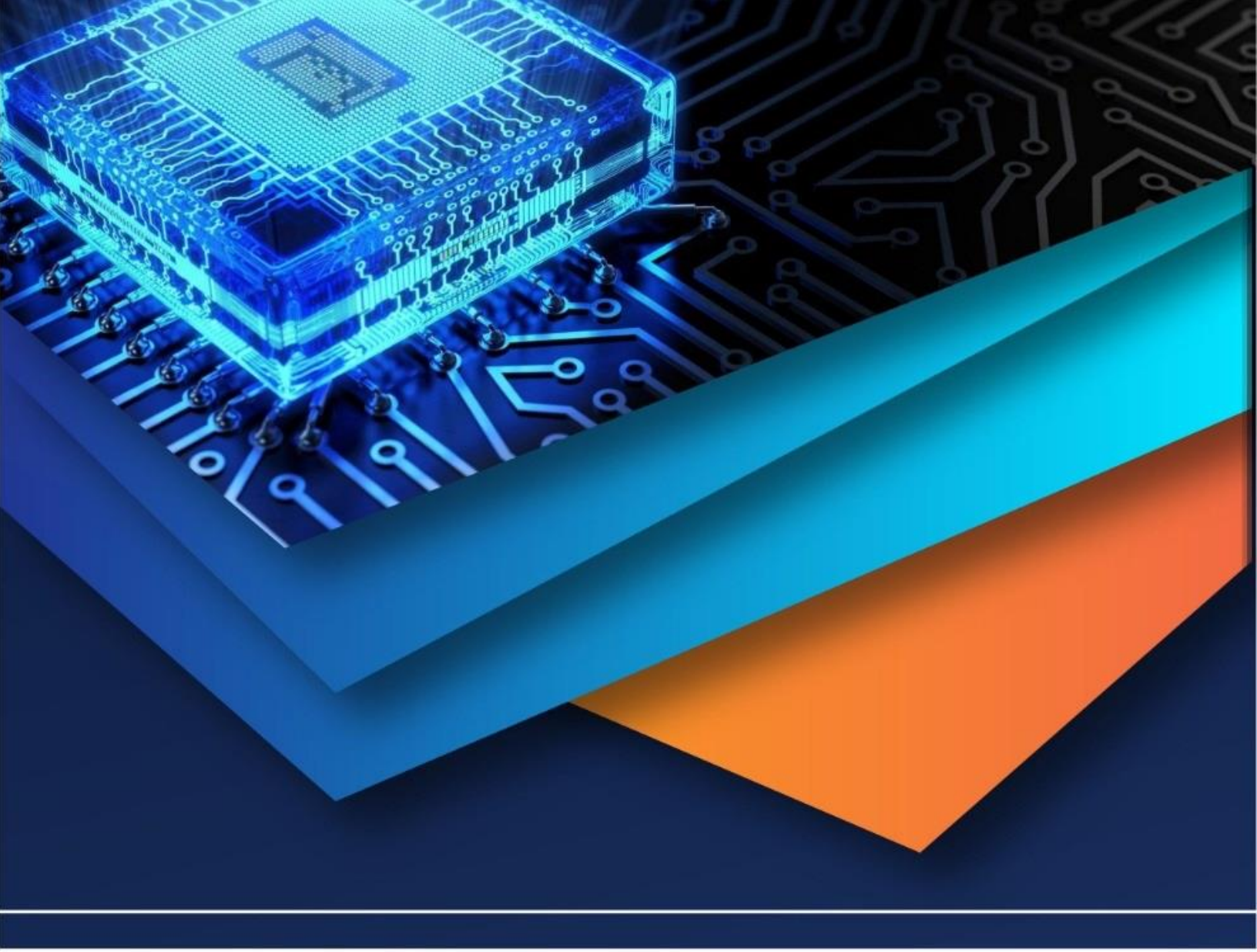
The EfficientNetB5 and EfficientNetB7 backbones might provide even more ac- curacy gains than EfficientNetB3, but at the expense of increased memory us- ages, especially since more GPU resources are available.

F. Confidence Thresholding

A low threshold at which the system makes an explicit “uncertain” classification, instead of possibly misleading low-confidence prediction, would enhance clinical safety properties of the deployed application.

REFERENCES

[1] 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,” Na- ture, vol. 542, no. 7639, pp. 115–118, 2017.
[2] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza et al., “Skin lesion analysis toward melanoma detection: ISIC 2017 challenge, dataset, and results,” in Proc. IEEE Int. Symp. Biomedical Imaging (ISBI), 2018.
[3] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. 36th Int. Conf. Machine Learning (ICML), vol. 97, pp. 6105–6114, 2019.
[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in Proc. IEEE Int. Conf. Computer Vision (ICCV), pp. 2980–2988, 2017.
[5] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinger, “Fusing fine-tuned deep features for skin lesion classification,” Computerized Medical Imaging and Graphics, vol. 71, pp. 19–29, 2019.
[6] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset, a large col- lection of multi-source dermatoscopic images of common pigmented skin lesions,” Scientific Data, vol. 5, article 180161, 2018.
[7] J. Yang, R. Shi, and B. Ni, “MedMNIST classification decathlon: A lightweight AutoML benchmark for medical image analysis,” in Proc. IEEE 18th Int. Symp. Biomedical Imaging (ISBI), 2021.
[8] B. Harangi, “Skin lesion classification with ensembles of deep convolutional neural networks,” Journal of Biomedical Informatics, vol. 86, pp. 25–32, 2018.
[9] I. González-Díaz, “DermaKNet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis,” IEEE Journal of Biomed- ical and Health Informatics, vol. 23, no. 2, pp. 547–559, 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)