Early Detection and Classification of Melanoma Skin Cancer Using Machine Learning Techniques

Introduction:

Skin cancer is one of the most common cancers worldwide. Among its types of melanomas is the most dangerous. It starts with the pigment-producing cells called melanocytes. If not found early it spreads fast to other parts of the body and becomes life-threatening [1]. Melanoma may appear anywhere on the skin but is often shown in areas exposed to sunlight such as the face neck and arms [2]. Though it makes up a small part of all skin cancer cases it causes many deaths [3]. The main problem is that malignant and benign lesions look very similar. Even trained doctors face trouble in telling them apart [4]. Many cases stay undetected until later stages when treatment is less effective and survival drops [5]. This creates the need for early detection and accurate classification which is a major challenge in skin cancer diagnosis.

To address this issue researchers turned to technology. Machine learning is widely used to improve diagnosis [2]. These models study large sets of medical images and spot hidden patterns. For example researchers at Stanford built a deep learning model. It matched the accuracy of expert dermatologists in classifying skin cancer [6]. Another effort the ISIC challenge gave a big open-source dataset for melanoma. It helped boost model performance [7]. In one study pretrained CNN models like ResNet-50 and VGG-16 reached over 90 percent accuracy [8]. These works aim to improve early diagnosis through smart systems though they depend heavily on quality data and correct model selection which remains a challenge.

There are many useful ways to apply this technology. Machine learning tools are now part of mobile apps. Users take a photo to screen skin lesions [9]. In hospitals these tools help doctors make fast and correct decisions. This is useful in places with few dermatologists [10]. It also lowers the workload of medical staff. Low-risk cases are filtered out. Only serious ones are flagged for review [11]. The goal is to make skin cancer detection simple and more available to all. One challenge is to ensure that these tools work well in real-world situations and stay easy to use.

This research helps improve skin cancer diagnosis using deep learning. It adds value to the growing use of artificial intelligence in healthcare. With better models and large medical datasets skin cancer detection becomes more accurate and faster. The aim is to support early treatment and save lives. But more work is needed to test and improve these models in real clinical settings where conditions vary.

2. Method Design:

2.1 Dataset

For this study, we utilized the HAM10000 (Human Against Machine with 10,000 training images) dataset (Tschandl et al., 2018), which contains 10,000 dermatoscopic images across seven different categories of skin lesions:

- Melanoma (mel)
- Melanocytic nevi (nv)
- Basal cell carcinoma (bcc)
- Actinic keratoses and intraepithelial carcinoma (akiec)
- Benign keratosis (bkl)
- Dermatofibroma (df)
- Vascular lesions (vasc)

The dataset includes images from different populations, captured with various dermatoscopic systems. Each image is associated with metadata including patient demographics, lesion location, and diagnostic confirmation method (histopathology, follow-up examination, expert consensus, or confirmation by in-vivo confocal microscopy).
For our binary classification task (melanoma vs. non-melanoma), we reorganized the dataset accordingly, resulting in 1,113 melanoma images and 8,887 non-melanoma images. This significant class imbalance was addressed in our methodology.

2.2 Data Preprocessing:

Several preprocessing steps were implemented to enhance image quality and standardize the input data:

1. Image Resizing: All images were resized to 224×224 pixels to maintain consistency and compatibility with pre-trained deep learning architectures.

2. Color Normalization: We applied color constancy algorithms to minimize variations in lighting conditions and camera settings. Specifically, we implemented the Shades of Gray algorithm (Finlayson et al., 2004) which performs well for dermatological images.

3. Hair Removal: Dermoscopic images often contain hair that can interfere with lesion analysis. We implemented a Dullrazor algorithm (Lee et al., 1997) followed by inpainting to remove hair artifacts.

4. Data Augmentation: To address class imbalance and improve model generalization, we applied the following augmentation techniques to the training set:
    - Random horizontal and vertical flips
    - Random rotations (±30 degrees)

- Random brightness and contrast adjustments (±10%)
- Random zoom (±15%)
- Elastic deformations

2.3 Feature Engineering

We implemented and compared two distinct approaches to feature extraction:

2.3.1 Handcrafted Features:

For traditional ML algorithms, we extracted the following feature sets:

- Color Features: Color histograms in RGB, HSV, and LAB color spaces; color moments (mean, standard deviation, skewness)
- Texture Features: Gray Level Co-occurrence Matrix (GLCM) features including contrast, correlation, energy, and homogeneity; Local Binary Patterns (LBP)
- Shape Features: Compactness, asymmetry, border irregularity, and fractal dimension
- ABCD Rule Features: Quantitative representations of the clinical ABCD criteria

These features were concatenated to form a 137-dimensional feature vector for each image.

2.3.2 Deep Features for deep learning approaches, we utilized:

- Transfer Learning: Feature extraction using pre-trained CNNs (VGG16, ResNet50, EfficientNetB3) on ImageNet
- End-to-End Learning: Direct training of CNN architectures with various depths and configurations
- Hybrid Approach: Combination of deep features with handcrafted features

2.4 Model Selection:

We implemented and compared the following models:

2.4.1 Traditional Machine Learning Models:

- Support Vector Machine (SVM) with RBF kernel
- Random Forest (RF) with 100 estimators
- XGBoost
- k-Nearest Neighbors (k-NN)

These models were trained using the handcrafted feature vectors.

2.4.2 Deep Learning Models:

- Custom CNN: A network with 4 convolutional blocks followed by fully connected layers
- Transfer Learning: Fine-tuned pre-trained networks:
    - VGG16
    - ResNet50
    - EfficientNetB3 (Final Chosen Model)

For transfer learning models, we froze the early layers (capturing generic features) and fine-tuned later layers for melanoma-specific features.

2.4.3 Ensemble Methods:

We also implemented ensemble techniques to improve classification performance:

- Soft voting of multiple CNN models
- Stacking ensemble combining traditional ML and deep learning predictions

2.5 Training Methodology:

The dataset was split into training (70%), validation (15%), and test (15%) sets with stratification to maintain class distribution. To address class imbalance, we applied the following techniques:

- Weighted loss function
- Class-weighted sampling
- SMOTE (Synthetic Minority Over-sampling Technique) for traditional ML models

Models were trained using the following configurations:

- Traditional ML: 5-fold cross-validation for hyperparameter tuning
- Deep Learning: Trained for 50 epochs with early stopping based on validation loss
- Learning rate scheduling with initial rate of 1e-4 and reduction on plateau
- Batch size of 32
- Adam optimizer

2.6 Evaluation Metrics:

Given the class imbalance and clinical context, we utilized multiple evaluation metrics:

- Accuracy
- Sensitivity (Recall)
- Specificity
- Precision

- F1-score
- Area Under the ROC Curve (AUC)
- Balanced Accuracy

Additionally, we generated confusion matrices and precision-recall curves to visualize model performance.

References:

[1] American Cancer Society. (2023). *Skin Cancer Facts.* Retrieved from **https://www.cancer.org/cancer/skin-cancer.html**

[2] World Health Organization. (n.d.). *Skin cancer.* Retrieved from **https://www.who.int/news-room/fact-sheets/detail/skin-cancers**

[3] Cancer Research UK. (n.d.). *Melanoma skin cancer.* Retrieved from **https://www.cancerresearchuk.org/about-cancer/melanoma**

[4] Mayo Clinic. (n.d.). *Skin cancer - Symptoms and causes.* Retrieved from **https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605**

[5] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. **https://doi.org/10.1038/nature21056**

[6] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 180161. Retrieved from **https://www.nature.com/articles/sdata2018161**

[7] Stanford University. (2017). Dermatologist-level classification of skin cancer. Retrieved from **https://ai.stanford.edu/blog/dermatology/**

[8] ISIC Archive. (2020). *International Skin Imaging Collaboration: Melanoma Project.* Retrieved from **https://www.isic-archive.com**

[9] Codella, N. C., Nguyen, Q. B., Pankanti, S., et al. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, 61(4/5), 5:1–5:15. **https://doi.org/10.1147/JRD.2017.2704659**

[10] World Health Organization. (n.d.). *Global report on health equity for marginalized populations.* Retrieved from **https://www.who.int/publications/i/item/9789240055470**

[11] SkinVision. (n.d.). *Skin cancer screening using a smartphone app. Journal of Dermatology*. Retrieved from **https://www.skinvision.com**

[12] McKinsey Global Institute. (2020). *Artificial Intelligence in Healthcare: Opportunities and Risks.* Retrieved from **https://www.mckinsey.com/industries/healthcare**