TWO-STEP HIERARCHICAL CLASSIFICATION OF SKIN LESION USING CONVOLUTIONAL NEURAL NETWORK AND RANDOM FOREST

Presentation by:

Taofik Ahmed Suleiman
Daniel Tweneboah Anyimadu
Andrew Dwi Permana
Hsham Abdalgny Abdalwhab Ngim

LIST OF CONTENTS

INTRODUCTION

- Skin cancer is an excessive growth characterized by the uncontrolled proliferation of abnormal skin cells.
- Significant advances regarding its treatment modalities have been made.
- However, skin cancer remains a global health concern due to its increasing incidence and associated morbidity and mortality rates.
- AI technology such as machine learning (ML) and deep learning (DL) techniques have been employed in the diagnosis, prediction, and treatment optimization of diseases, including skin lesions.
- **In this project, we used CNN and RFC algorithms to tackle the challenge of classifying three distinct types of skin lesions: benign, melanoma, and seborrheic keratosis.**
- **Our approach is based on a 2-step hierarchical binary classification because it offers a unique solution to address the class imbalance and the inherent variability in skin lesions.**

RELATED WORK

Several studies have demonstrated the effectiveness of hierarchical classification in predicting skin lesions, highlighting its potential in this field.

01

A. Esteva et al.

Utilized a hierarchical approach with a CNN, showing improved performance compared to flat classification models.

02

P. Tschandl et al.

Used a two-step hierarchical classification for skin cancer diagnosis to improve accuracy by first differentiating benign and malignant lesions before classifying subcategories.

03

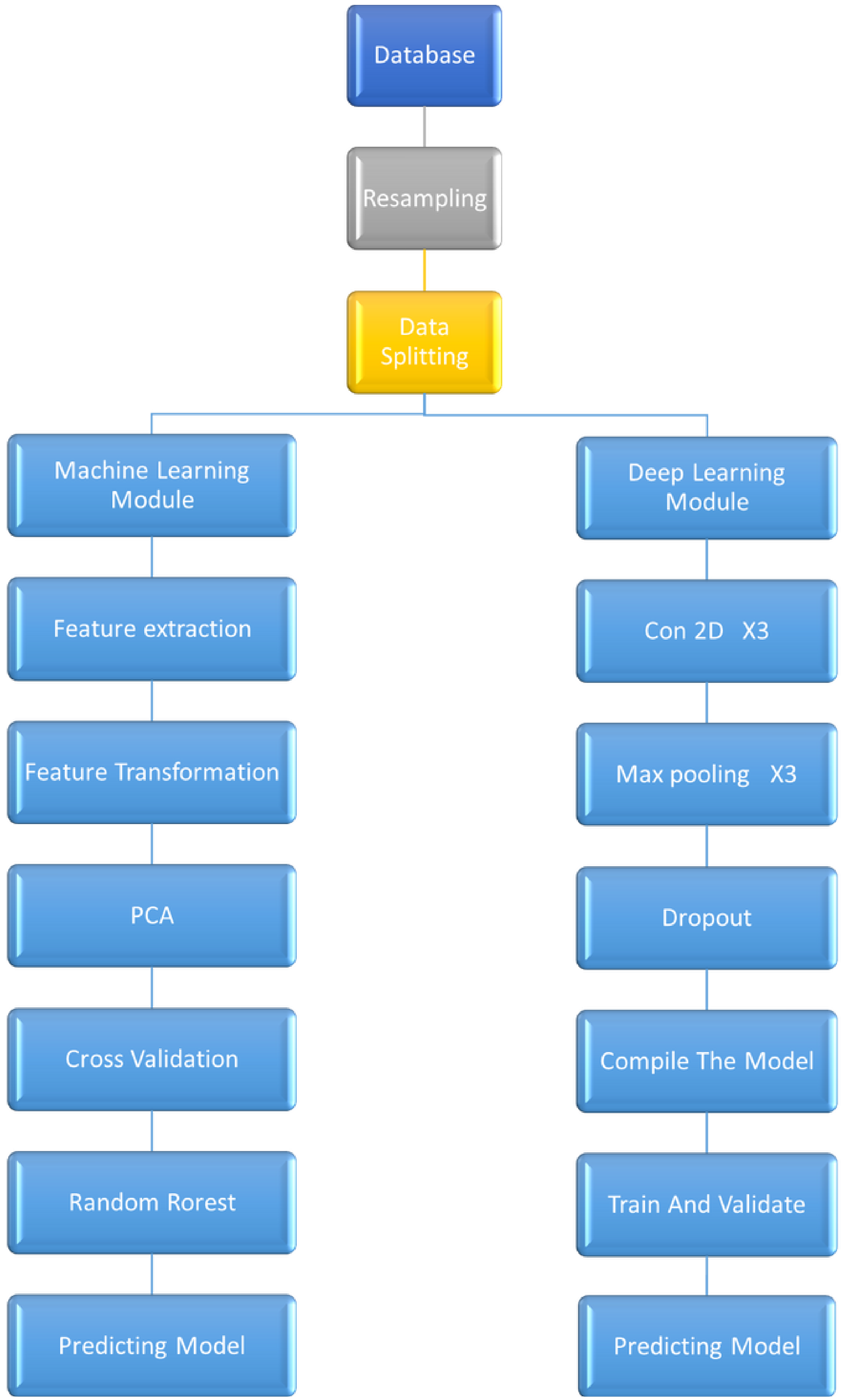
F. Afza et al.

Proposed a hierarchical framework using two-dimensional superpixels and deep learning. They segmented lesions, boosted dermoscopy contrast, and used deep learning transfer learning.

MATERIAL AND METHODOLOGY

This section presents the materials used in the project, the source of the skin lesion image dataset, as well as the methodologies used to achieve the two-step hierarchical classification.

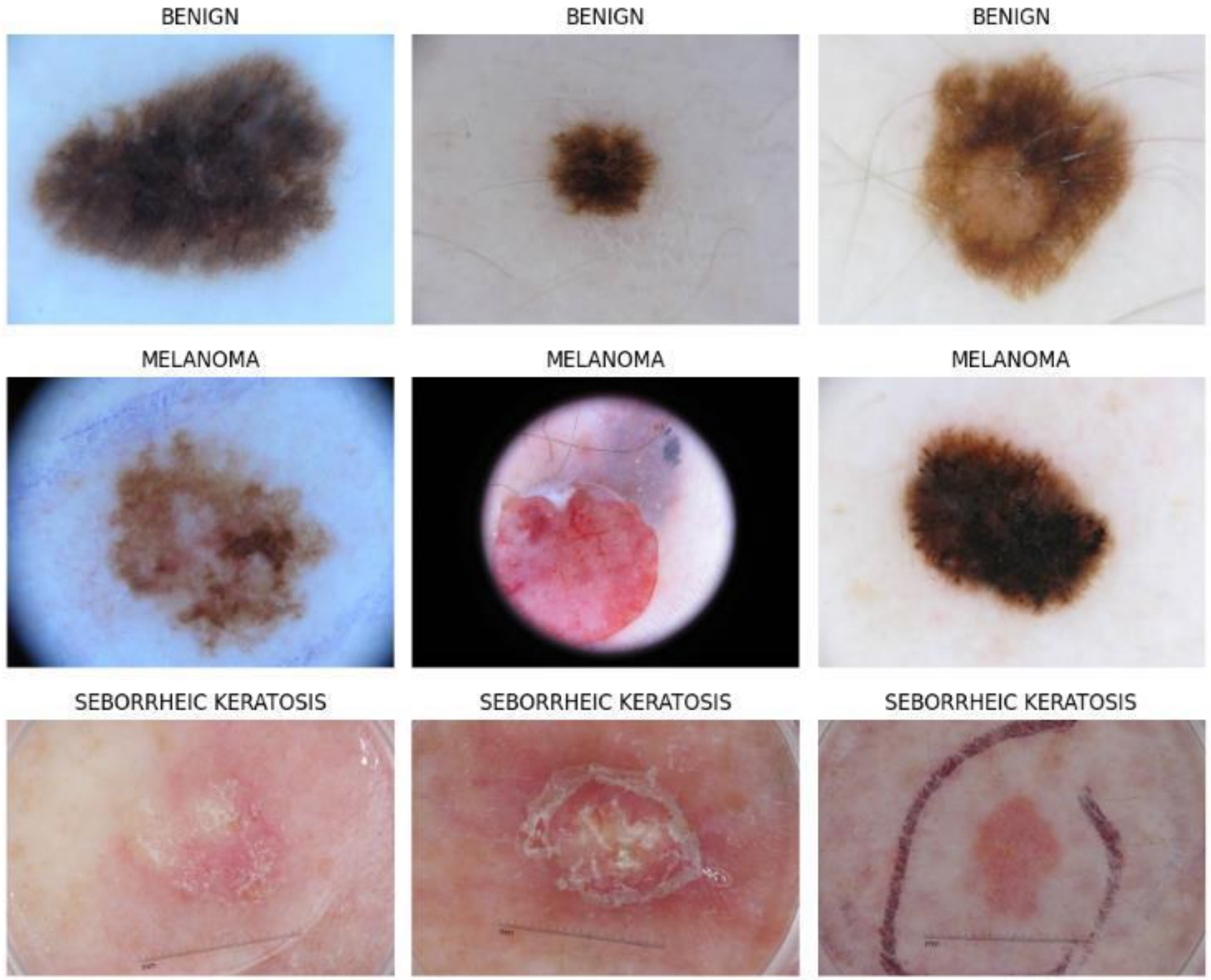
FLOW DIAGRAM OF THE ALGORITHM USED



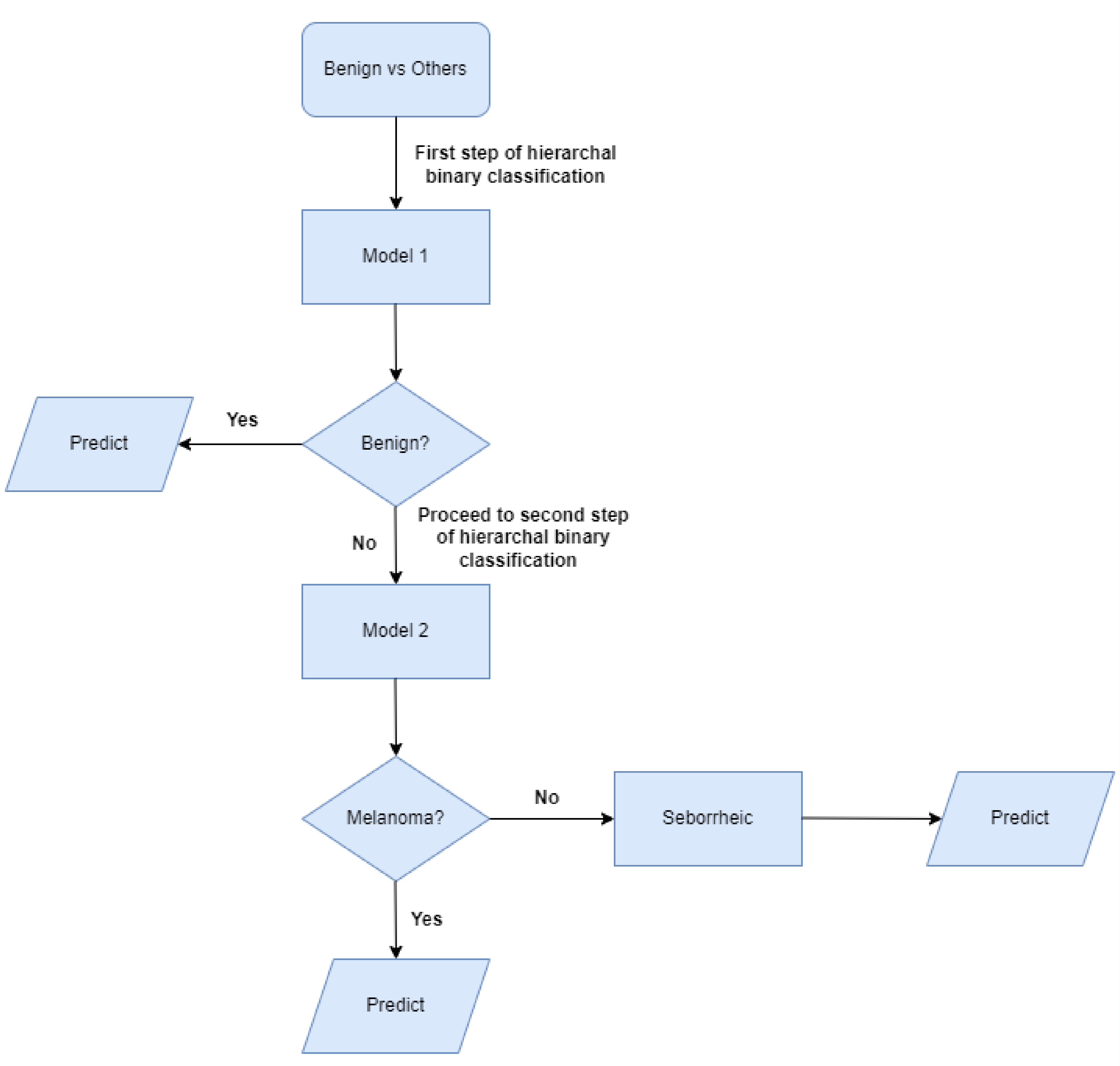
MATERIAL

Image Dataset

The images were taken from the International Skin Imaging Collaboration (ISIC) 2017 dataset, which consists of 2000 lesion images with three tumor types: melanoma (374 images), seborrheic keratosis (254 images), and benign (1372 images).



METHODOLOGY



Two-step Hierarchical Binary Classification

This technique offers robustness compared to traditional multiclass classification methods.
In the first step, we classified Benign as the majority class in the dataset, followed by the classification of the remaining classes which are melanoma and seborrheic keratosis.

Resampling and Data Splitting

We combined the class melanoma and seborrheic keratosis and labeled them as "others", moreover we upsampled to 1000 samples, while the benign class was downsampled to 1000 samples, resulting in 2000 samples in total.

We splitted the data using the class-wise splitting method, where the first 70% of images were assigned as a training set, while the remaining 30% were allocated to the testing set.

MACHINE LEARNING MODULE

- **Features Extraction and Transformation**

Convolutional filters were utilized to extract several features from the input data. The convolutional layers of the network effectively captured spatial patterns and hierarchical representations within the images. To ensure fair comparison and prevent features with larger magnitudes from dominating the classification process, we applied the standardized scaler method.

- **Principal Component Analysis**

We utilized PCA to address the challenge of high-dimensional feature spaces and mitigate potential overfitting.

- **Five-Fold Cross-validation**

Cross-validation ensured the resilience and generalization of our machine learning model. During training, we used 5-fold cross-validation.

- **Random Forest Classifier**

We set our number of estimators, or decision trees to 50 for each step of the hierarchical classification. This value was determined based on experimentation and fine-tuning to achieve a good balance between model complexity and performance.

DEEP LEARNING MODULE

- A short convolutional neural network (CNN) architecture was implemented in this module.

- A sequential model was built utilizing:
- **Convolutional layers:** For feature extraction.
- **Activation function:** To enable the model to learn complex patterns.
- **Max pooling:** To reduce spatial dimensions.
- **Dropout layers:** To mitigate overfitting.

- The resulting outcome was flattened and fed to a dense layer for classification.

- Adam optimizer was utilized in training the neural network.

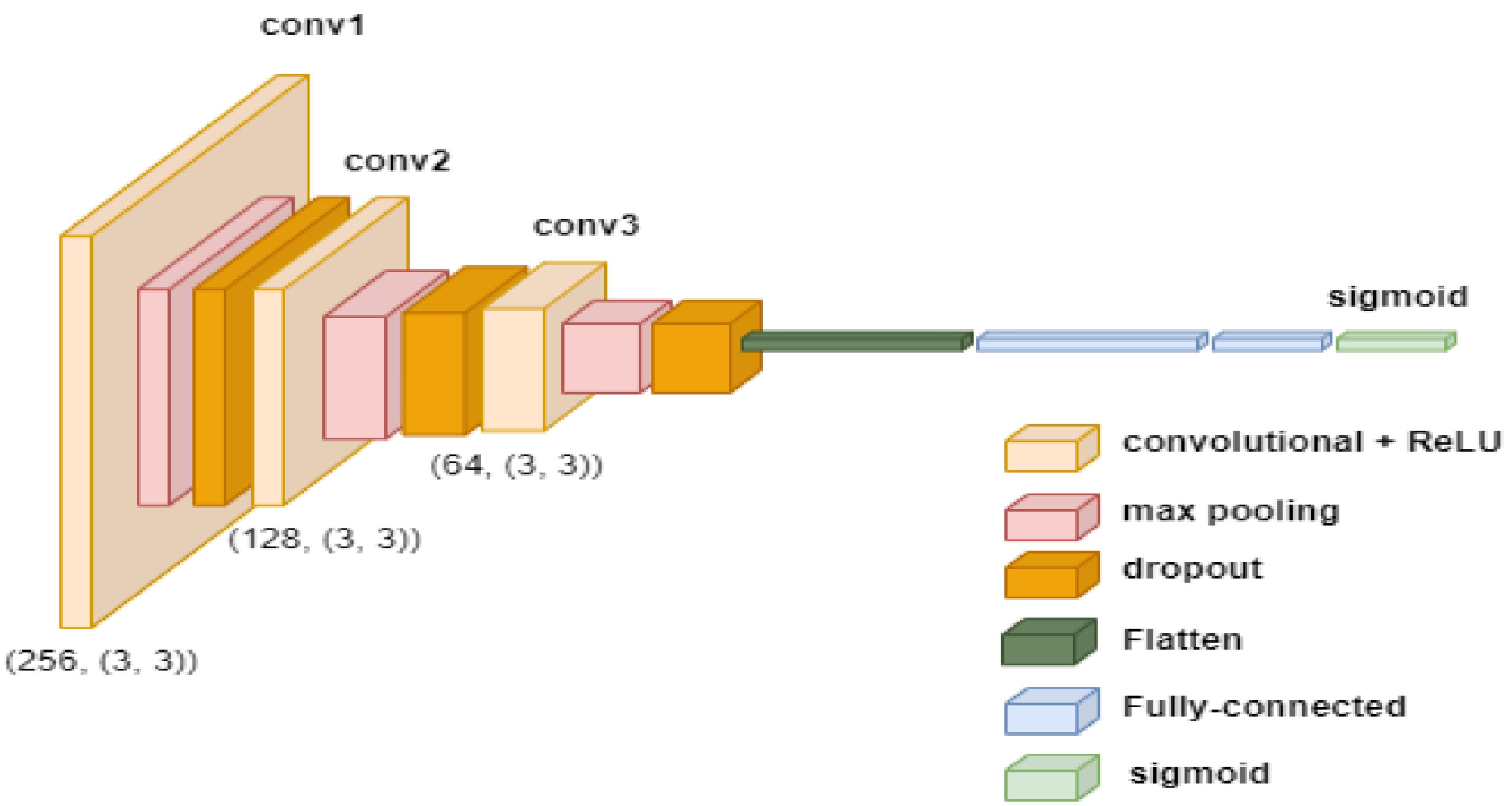


Fig : A DL network architecture configured using 256, 128, and 64 convolutional layers respectively with a 3 x 3 kernel/filter size.

RESULTS

In this section, we present and discuss the results obtained from each method, highlighting their respective strengths.

Cross-Validation Results for the Machine Learning Module

The Random Forest classifier consistently outperformed the other classifiers in each fold, achieving an average accuracy of 81.93%. SVM, KNN, and logistic regression yielded lower average accuracies of 67.02%, 69.21%, and 66.50% respectively.

Result of the cross validation for the first stage

Evaluation Metrics	Random Forest	SVM	KNN	Logistic regression
Accuracy	**0.8267**	0.6533	0.7050	0.6383
Precision	**0.8245**	0.6556	0.7078	0.6402
Recall	**0.8300**	0.6533	0.7050	0.6383
F1-Score	**0.8272**	0.6521	0.7040	0.6371
AUC	**0.8267**	0.6533	0.7050	0.6383
BA	**0.8283**	0.6533	0.7050	0.6383

Result of the cross validation for the second stage

Cross Validation	Random Forest	SVM	KNN	Logistic regression
1-fold	**0.8857**	0.7643	0.6714	0.7786
2-fold	**0.9143**	0.7929	0.7429	0.7786
3-fold	**0.8929**	0.7571	0.6071	0.7643
4-fold	**0.8857**	0.7929	0.6143	0.7929
5-fold	**0.9281**	0.7266	0.6547	0.7770

RESULTS

In this section, we present and discuss the results obtained from each method, highlighting their respective strengths.

Performance Evaluation of the Machine Learning Module

To evaluate the performance, we are using a confusion matrix, where we calculated accuracy, precission, recall, F1-score, AUC, and balanced accuracy.

Performance evaluation for the first stage

Evaluation Metrics	Random Forest	SVM	KNN	Logistic regression
Accuracy	**0.8267**	0.6533	0.7050	0.6383
Precision	**0.8245**	0.6556	0.7078	0.6402
Recall	**0.8300**	0.6533	0.7050	0.6383
F1-Score	**0.8272**	0.6521	0.7040	0.6371
AUC	**0.8267**	0.6533	0.7050	0.6383
BA	**0.8283**	0.6533	0.7050	0.6383

Performance evaluation for the second stage

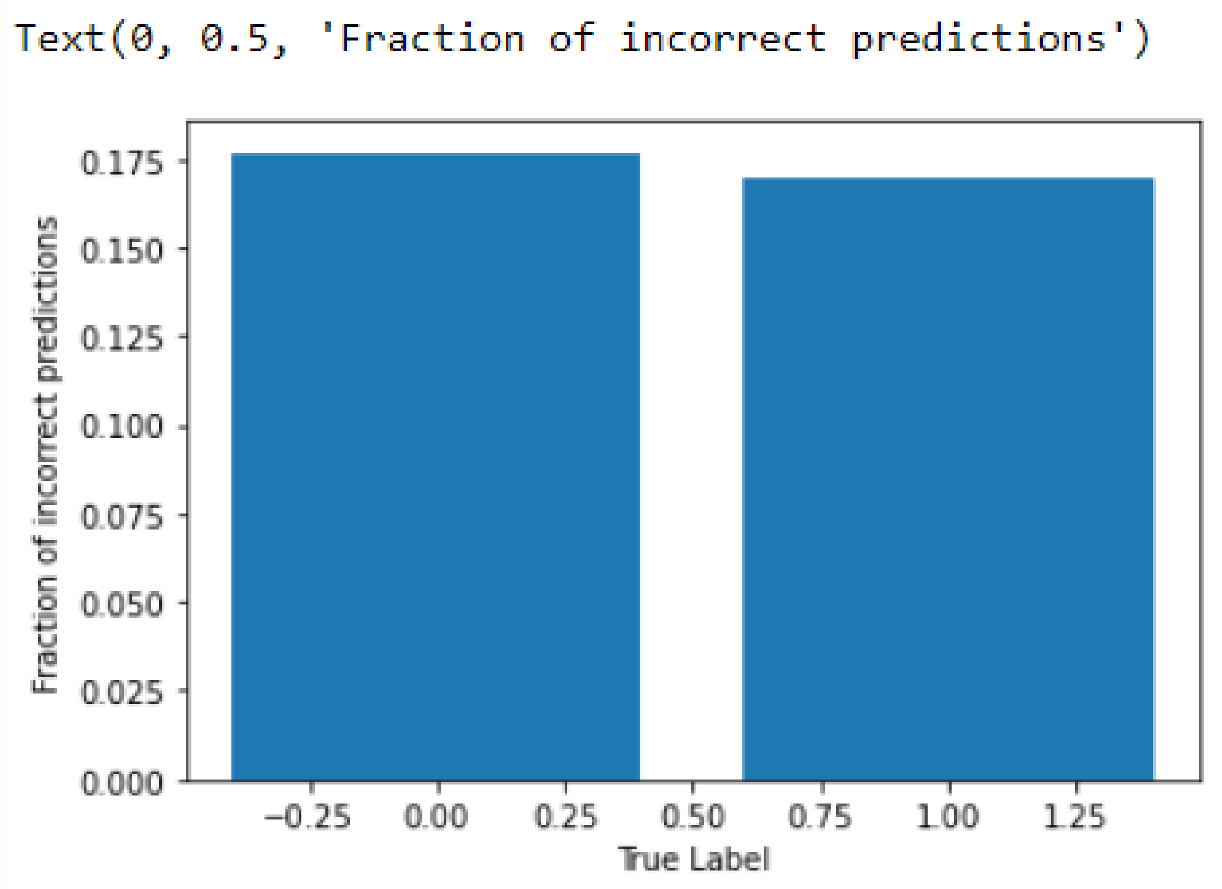
Evaluation Metrics	Random Forest	SVM	KNN	Logistic regression
Accuracy	**0.9037**	0.8073	0.7674	0.8206
Precision	**0.8667**	0.8068	0.7667	0.8197
Recall	**0.8888**	0.8073	0.7674	0.8206
F1-Score	**0.8776**	0.8070	0.7671	0.8200
AUC	**0.9010**	0.7957	0.7538	0.8081
BA	**0.8963**	0.8073	0.7674	0.8206

RESULTS

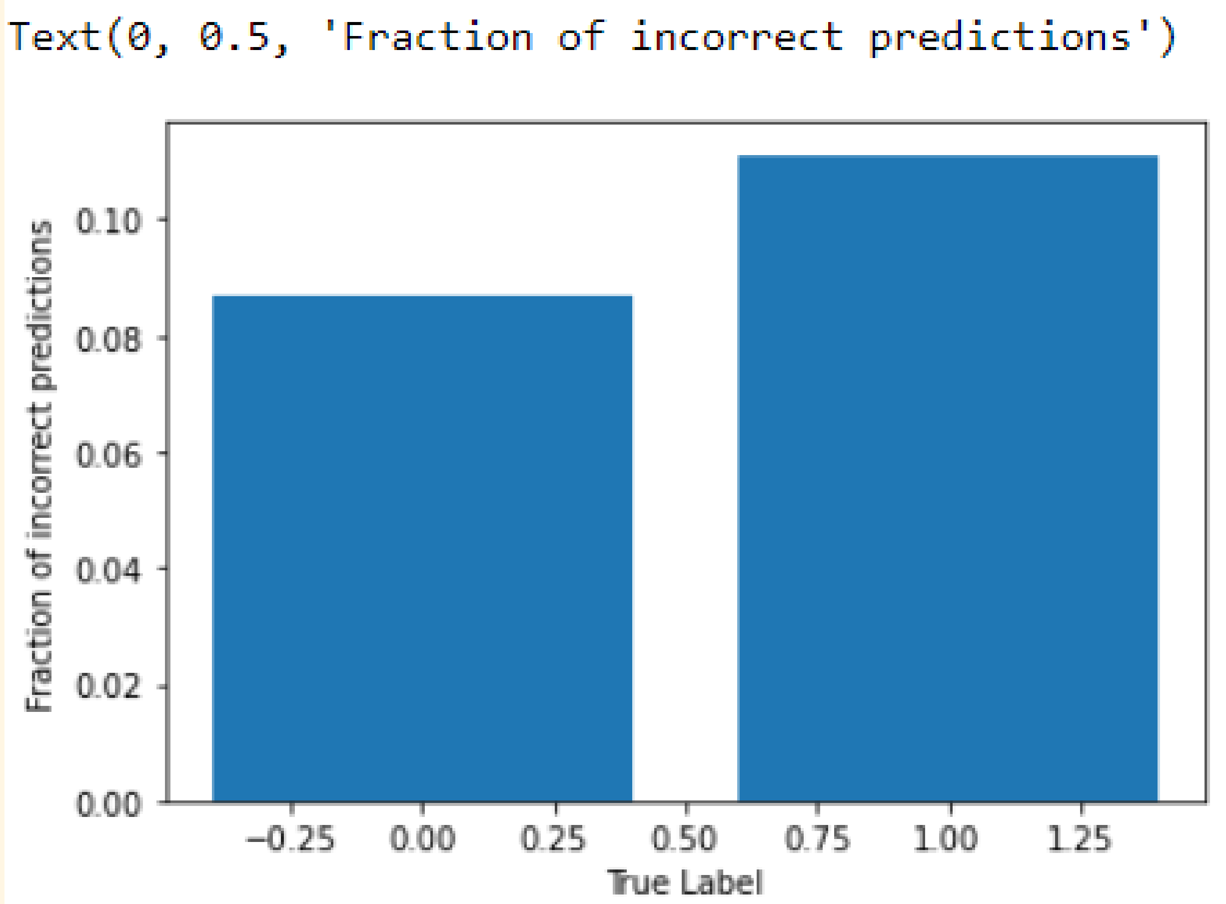
In this section, we present and discuss the results obtained from each method, highlighting their respective strengths.

Plot of fractional incorrect misclassifications

Fraction of incorrect predictions for step 1



Fraction of incorrect predictions for step 2



RESULTS

In this section, we present and discuss the results obtained from each method, highlighting their respective strengths.

Comparison Performance Evaluation Deep Learning (CNN) and Random Forest

Comparison results achieved by CNN and Random Forest in first stage

Evaluation Metrics	Random Forest	CNN
Accuracy	**0.8267**	0.7500
Precision	**0.8245**	0.7483
Recall	**0.8300**	0.7533
F1-Score	**0.8272**	0.7508
AUC	**0.8267**	0.7500
BA	**0.8283**	0.7517

Comparison results achieved by CNN and Random Forest in second stage

Evaluation Metrics	Random Forest	CNN
Accuracy	**0.9037**	0.8904
Precision	**0.8667**	0.8333
Recall	0.8888	**0.8974**
F1-Score	**0.8776**	0.8642
AUC	**0.9010**	0.8917
BA	**0.8963**	0.8939

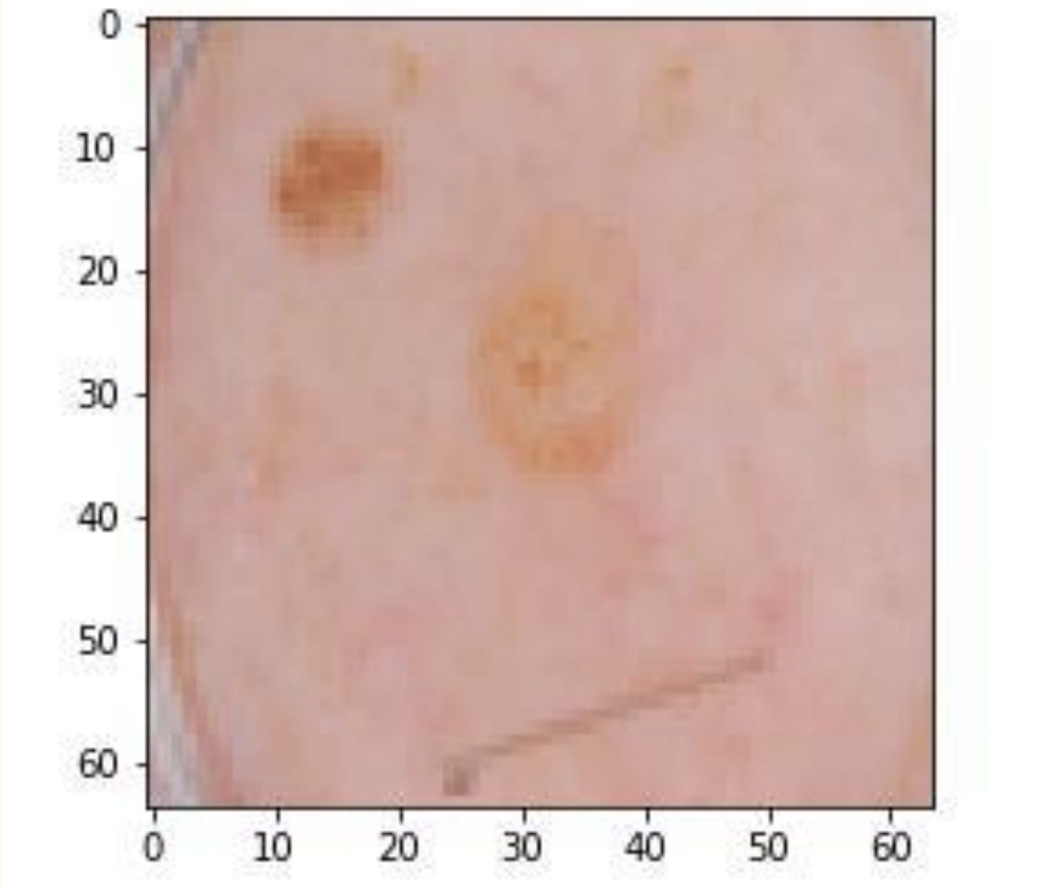
RESULTS

In this section, we present and discuss the results obtained from each method, highlighting their respective strengths.

Implementation of Predictive Model for the two-step Hierarchical Binary Classification

We proceeded to implement a predictive model that integrates the two-step hierarchical binary classification process.
A visual representation of the results obtained from the predictive model.

```
1/1 [==============================] - 0s 180ms/step
1/1 [==============================] - 0s 79ms/step
The prediction for this image is: seborrheic_keratosis
The actual label for this image is: seborrheic_keratosis
```



Balanced Multiclass Accuracy

The random forest classifier achieved a BMA of 87.51%, indicating its effectiveness in accurately classifying the different skin lesion classes. On the other hand, CNN achieved a BMA of 84.33%

CONCLUSION

- In conclusion, this project aimed to develop a robust and accurate classification system for skin lesion diagnosis.
- A 2-step hierarchical binary classification using both machine learning and deep learning approaches was implemented.
- The results obtained from the evaluation of the machine learning module showcased the potential of the random forest classifier in accurately classifying skin lesions.
- The deep learning module showed competitive performance in classifying seborrheic keratosis and melanoma, even though the random forest classifier slightly outperformed it in terms of accuracy.
- Overall, this project has shown how machine learning and deep learning techniques can help dermatologists in the early detection and diagnosis of skin lesions.
- **The Balanced Multiclass Accuracy (BMA) metric further solidifies the reliability of our models, with the random forest classifier achieving a BMA of 87.51% and the CNN achieving a BMA of 84.33%.**