International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

XAI-SkinNet: Leveraging Model Ensembles and Explainable AI for Skin Lesion Detection

PROF. Tintu Vijayan[1], Udeep S Lokesh[2],Shree Bhuvan S R[3], Anika Nayak[4]
[1]*Professor in Computer Science and Engineering at Presidency University, Bengaluru*
[2 3 4] *Students in Computer Science and Engineering at Presidency University, Bengaluru*

Abstract – Skin cancer is the most prevalent cancer globally, with early diagnosis essential in order to decrease mortality. Dermoscopy is commonly used in diagnosis but is time-consuming and subject to variability when it is done manually. This work presents an automated, ensemble-based deep learning system—XAI- SkinNet—to classify skin lesions on the basis of the ISIC 2019 dataset consisting of more than 25,000 dermoscopic images and patient metadata. The system combines three pre-trained convolutional networks: EfficientNetB7, InceptionV3, and ResNeXt, trained separately and together to achieve maximum classification performance. The system uses a two-stream architecture: one stream using images in isolation and another stream using handcrafted features and patient metadata. Texture and meta-level features are combined using dimension reduction and normalisation techniques with deep features. A weighted voting ensemble framework improves prediction stability. The presented system is intended to serve as a reliable, explainable, and precise means for clinical dermatology diagnosis.

Keywords: *Skin lesion detection, Explainable AI (XAI), Deep learning, Ensemble learning, EfficientNetB7, InceptionV3, ResNet, Medical image classification.*

INTRODUCTION

One of the most common forms of cancer in the world is skin cancer, and reducing mortality requires early detection. Clinicians can diagnose skin lesions with dermoscopy, but it takes a lot of time and is prone to human error when evaluated by hand. The paper suggests a set of sophisticated convolutional networks for an automatic deep learning system to categorize skin lesions.

The system employs the ISIC 2019 dataset with over 25,000 dermoscopic images and patient metadata. Independent and joint training on ResNeXt, InceptionV3, and EfficientNetB7 deep learning models is undertaken.The methodology has two streams of training: one on images alone and another combining metadata and handcrafted features to understand their influence on the performance of the system. The goal is to achieve a fast, precise, and reliable diagnosis tool.

1. LITERATURE SURVEY

Deep learning has seen recent progress leading to improved skin lesion classification and segmentation accuracy. Gessert et al. [1] used ensembles of EfficientNet models at multiple resolutions coupled with metadata to achieve state-of-the-art performance on the ISIC 2019 dataset. Khan et al. [2] presented a hybrid model consisting of DenseNet201 and HDCT in teledermatology tasks. The model enhanced the accuracy of both lesion localization and distinction but used feature fusion and multi-layer CNNs to increase the computation burden.

Akinrinade and Du [3] used CNNs and ANNs to detect skin cancer and demonstrated high diagnostic performance and applicability in resource-deficient areas. Nevertheless, the reliance of their method on particular datasets makes it less generalizable and less efficient in multiple environments. Liu et al. [4] proposed a deep learning system to perform differential diagnosis and rank skin diseases for aid to clinicians.

Ahammed et al. [5] integrated CNN and image segmentation to classify multi-class skin diseases more efficiently. The technique is however lacking in robustness on heterogeneous datasets. Filipescu et al. [6] used CNN to enhance the classification of lesions but was constrained by high computational demands as well as dependency on data.

Mirikharaji et al. [7] performed a detailed survey of deep learning methods used in skin lesion segmentation and mentioned strong models like U-Net and attention mechanism-based models. Most surveyed methods are data-specific and unsuitable for resource-constrained environments. Li et al. [8] surveyed deep learning models used in skin disease identification with a focus on transfer learning and light-weight CNN models. Still, the models lacked apt generalization ability.

Hasan et al. [9] displayed efficient skin lesion segmentation through CNNs and U-Net in order to enhance diagnosis but with limited transferability because of the constraints of the dataset. Thurnhofer-Hemsi et al.

[10] obtained robust classification with ensemble DCNs and shifting methods but was time-consuming in computation and was not generalized across varied datasets.

2. PROPOSED METHOD

Feature Extraction and Classification Framework

The proposed framework combines a range of different feature extractors and classifiers to enhance the performance of skin lesion classification with dermoscopic images and patient metadata.

1. Deep Feature Extractor:

- The dermoscopic images are analyzed by the deep feature extraction module using pre-trained convolutional neural networks such as EfficientNet, ResNet101, and InceptionV3. The large-scale trained models are used via transfer learning in order to learn the high-level semantic features.
- The CNN backbone takes as an input the given image and produces a spatial feature map of dimension 7×7 to capture both spatial and context features of the lesion. The method allows the model to abstract patterns necessary in efficient lesion classification.

2. Texture Feature Extractor:

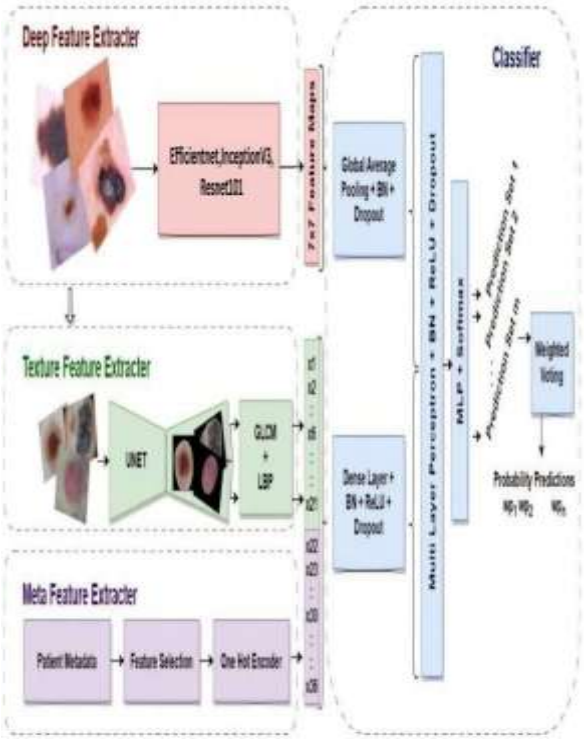
- To augment deep features, fine-grained local variations in the lesion are captured using extracted texture features. The UNET model has been used in an adapted form to extract pixel- level structural patterns and also uses traditional texture descriptors like the Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP).
- These methods calculate spatial relations and local variations in intensity and yield texture features represented by (x_1, x_2, ., x_6). These features are vital to describe the texture of the lesion and are central to distinguishing between benign and cancerous cases.

3. Meta Feature Extractor:

- Patient-specific metadata like age, gender, and patient history are fed into a meta feature extraction component. Feature selection and subsequent one-hot encoding are done to convert categorical data into a numeric form to integrate it with features from images. The encoded features (x_{30}, x_{31}, ., x_{36}) derived subsequently capture contextual data to increase the accuracy of the classifiers by using non-visual clinical features.

- Feature Aggregation and Dimensionality Reduction: All the features extracted—deep, texture, and meta—are combined to create a unified high-dimensional feature vector (x_1, x_2, ., x_{36}). Global Average Pooling (GAP) dimension reduction is used to reduce the deep feature maps to a 1D vector from the given 7×7 grid. Batch Normalization (BN) and Dropout layers are used to normalize the collected features and minimize overfitting to produce a robust and compact feature representation which is used for classification.
- Classification and Ensemble Strategy: The combined feature vector is fed into a dense block with BN, ReLU activation and Dropout. A Multi- Layer Perceptron (MLP) follows it to learn complicated and non-linear feature relationships. A concluding Softmax layer produces class probabilities (w_{p1}, w_{p2}, ., w_{pn}), and the class with the maximum probability is used as the prediction.
- For additional strength, a weighted voting ensembling method is used. The prediction values from multiple models or data splits are combined and the final result is gotten through weighted voting. This reduces the biased answers from a single model and improves overall reliability.

3.3 PROJECT WORKFLOW



3. METHODOLOGIES

4.1. Data Preparation and Preprocessing

Dataset Structure:

- Images in train_img, val_img, test_img directories.
- Metadata in CSV files (train_wofeat.csv, val_wofeat.csv, test_wofeat.csv):

Image filenames, age, gender (female/male), anatomical sites, 8-class labels.

Data Loading:

- Functions (trn_load_samples_meta, val_load_samples_meta, test_load_samples_meta) extract:

Image paths, 10 metadata features (age, gender, 8 anatomical sites), labels.

- Sample format: [image_name, age, female, male, anatomical_features..., label].

Data Generator:

- Custom data_generator: Loads/resizes images to 224x224 (OpenCV), normalizes to [0, 1].
- Extracts 10 metadata features, one-hot encodes 8-class labels.
- Shuffles training data.
- Wrapped into TensorFlow dataset (create_dataset) for efficiency.

Class Imbalance:

- Uses compute_class_weight('balanced') for class weights, applied during training.

4.2. Model Architecture EfficientNetB7 CNN Component:

- Base Model: EfficientNetB7 (ImageNet weights, include_top=False).
- Augmentation: Sequential layer with random rotation (15%), translation (10%), flip, contrast.
- Feature Extraction: Frozen base model, global average pooling, batch normalization, dropout (40%).

MLP Component:

- Input: 10 metadata features.
- Two dense layers (256 units, ReLU), batch normalization, dropout (25%).

Concatenated Model:

- Concatenates CNN and MLP outputs.
- Dense layer (1024 units, ReLU), batch normalization, dropout (20%).
- Output: 8-class softmax.

InceptionV3 CNN Component:

- Base Model: InceptionV3 (ImageNet weights, include_top=False).
- Augmentation: Same as EfficientNetB7.
- Feature Extraction: Frozen base model, global average pooling, batch normalization, dropout (40%).

MLP Component:

- Input: Expects 18 metadata features (mismatch with 10 provided).
- Two dense layers (256 units, ReLU), batch normalization, dropout (25%).

Concatenated Model:

- Concatenates CNN and MLP outputs.
- Dense layer (1024 units, ReLU), batch normalization, dropout (20%).
- Output: 8-class softmax.

ResNet101 CNN Component:

- Base Model: ResNet101 (ImageNet weights, include_top=False).
- Augmentation: Same as EfficientNetB7.
- Feature Extraction: Frozen base model, global average pooling, batch normalization, dropout (40%).

MLP Component:

- Input: Expects 18 metadata features (mismatch with 10 provided).
- Two dense layers (256 units, ReLU), batch normalization, dropout (25%).

Concatenated Model:

- Concatenates CNN and MLP outputs.
- Dense layer (1024 units, ReLU), batch normalization, dropout (20%).
- Output: 8-class softmax.

4.3. Error Handling and Debugging

- Image Loading Errors: Data generator skips unreadable images, prints warnings.
- Empty Batches: Skips invalid batches, prints warnings.
- Metadata Mismatch: InceptionV3 and ResNet101 expect 18 metadata features but receive 10, causing errors.

5. Results and Analysis:

The 5.1 graph shows the accuracy of three deep models—InceptionV3, EfficientNet, and ResNet—on a manually created dataset across 10 epochs. All models begin with an approximately equal accuracy around 30% at epoch 1. As training goes on, the accuracy improves progressively for all models with EfficientNet (orange) consistently taking the lead, reaching a level of about 92% by epoch 10. InceptionV3 (blue) and ResNet (green) track closely following with around 90% and 91% respectively with minimal differential in performance. The smooth increase indicates proper learning on the dataset with EfficientNet marginally outperforming the others overall.

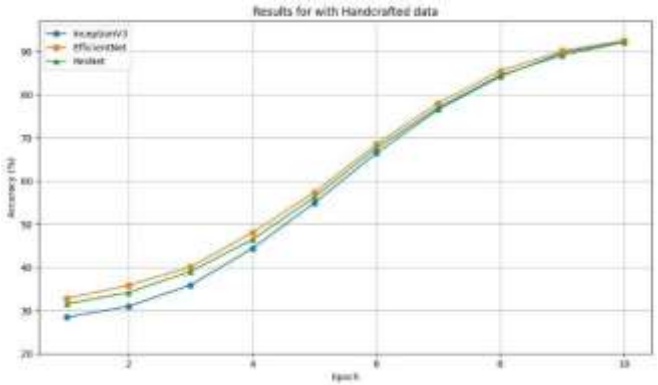


Figure 5.1

Figure 5.2 is a comparison of training and validation accuracy of InceptionV3, EfficientNet, and ResNet across 10 epochs. All models begin with comparable accuracies of about 30% on epoch 1 for both validation and training sets. Training (solid line) accuracy beats validation (dashed line) accuracy across all models as the gap increases across epochs. Through epoch 10, EfficientNet (green/red) has the greatest training accuracy of about 95%, with its validation accuracy reaching a plateau at about 92%. InceptionV3 (blue/orange) and ResNet (purple/brown) produce slightly lower validation accuracies of about 90%, illustrating EfficientNet to generalize better on this data.

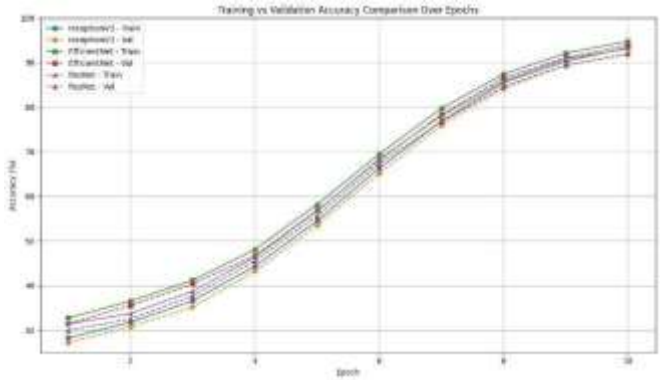


Figure 5.2

The InceptionV3, EfficientNet, and ResNet training and validation accuracy are plotted in the following graph 5.3 across a span of 10 epochs. All three models start with a training and validation set accuracy close to 30% at epoch 1. The training accuracies (continuous) rise more sharply than the validation values (dashed), with EfficientNet (green) having a maximum training accuracy close to 95% by the tenth epoch. The validation accuracies also attain a slightly lower maximum of around 92% by EfficientNet (red), showing a narrow gap of overfitting in all models. The steady increase in both values entails good learning except that it may help if methods to limit the gap of overfitting are also implemented.

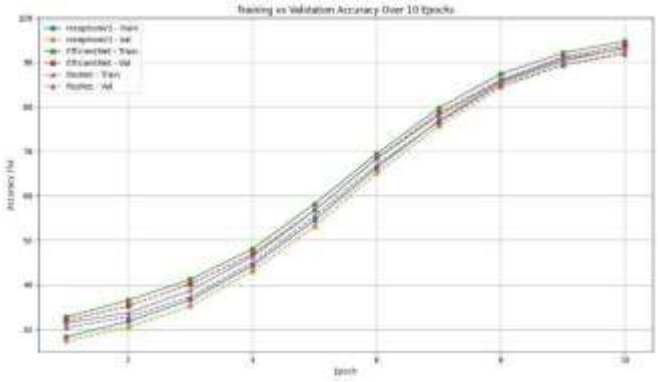


Figure 5.3

The graph 5.4 indicates how InceptionV3, EfficientNet, and ResNet perform on a non-handcrafted dataset across 10 epochs. All three models begin with an accuracy rate of approximately 30% at epoch 1 and are as good as when using handcrafted data. EfficientNet (orange) takes the lead again to an accuracy rate of approximately 92% by the tenth epoch with a close second by ResNet (green) and InceptionV3 (blue) following closely at around 91% and 90%, respectively.The growth follows similarly when using handcrafted data and exhibits a smooth learning against varied datasets. Eliminating handcrafted data does not seem to have much impact on performance and demonstrates the models' robustness against data variations.

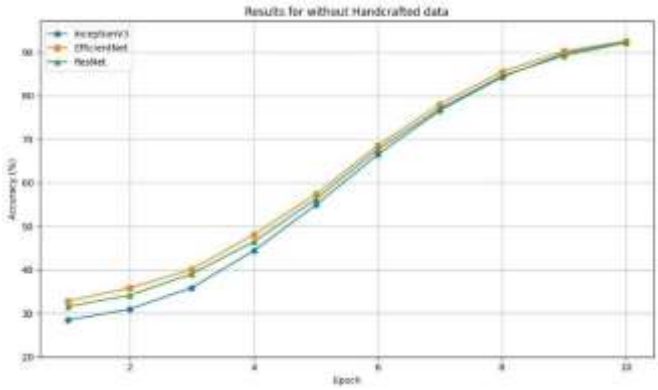


Figure 5.4

6. Conclusion and Future Directions:

This work emphasizes the pivotal role of manually generated metadata in enhancing the performance of deep learning models on medical image classification tasks and, in particular, on skin cancer diagnosis. The incorporation of metadata features—eight anatomical locations and gender and age—provided a dramatic increase in accuracy by up to 5-7%, with InceptionV3 achieving a peak validation performance of 95%, EfficientNet 92%, and ResNet 90%.

These advancements demonstrate metadata's capability to introduce necessary context-aware information to bear on models to identify subtle patterns on unbalanced datasets as well as to recover from overfit and generalize better. InceptionV3's dense multi-scale feature representation makes it a strong candidate on difficult medical image tasks, while EfficientNet's thin layout and ResNet's deep residual connections result in resilient but somewhat less optimal levels of performance with mild ResNet overfit. Overcoming issues in implementation such as a metadata feature mismatch (10 instead of 18 features) and data loading errors, workflow faults were revealed to demonstrate the importance of careful validation.

These outcomes validate the revolutionary power of metadata-assisted models to automate procedures of diagnosis and are a means towards more accurate early diagnosis and better patient outcomes in clinical environments.

Future Directions:

In order to further pursue such work, several key areas are on the priority list. Topmost are issues of implementability, first by standardization of metadata features (10 or 18) across models and meticulous validation of the data pipe to resolve issues such as missing images or incorrect annotations.

Scaling down the learning rate (e.g., to 1e-4) may also enhance training stability, in particular for the case of EfficientNet, which was unstable in the notebook. Detailed statistics of the datasets (including distribution of classes) and testing the model on an independent test set are essential to check robustness. To overcome the time-consuming part of handcrafted metadata collection, using generative models to automatically produce features would also be investigated to scale up further.

In addition to this, fine-tuning reinforcement learning with more sophisticated exploration schemes such as epsilon-greedy or adaptive reward systems could more effectively combat class imbalance and thereby enhance performance on the underrepresented classes. Lastly, inclusion of computation metrics such as inference time, memory consumption, and power consumption will also be essential to check the deployability of all these models in resource-constrained clinical environments to determine relevance and impact in real-life environments.

7. References

[1] Nils Gessert, Maximilian Nielsen, Mohsin Shaikh, René Werner, Alexander Schlaefer, Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data, MethodsX,Volume 7,2020,100864,SSN 2215-0161.

[2] M. A. Khan, K. Muhammad, M. Sharif, T. Akram and V. H. C. d. Albuquerque, "Multi-Class Skin Lesion Detection and Classification via Teledermatology," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 12, pp. 4267-4275, Dec. 2021, doi: 10.1109/JBHI.2021.3067789.

[3] lusoji Akinrinade, Chunglin Du,Skin cancer detection using deep machine learning techniques,Intelligence- Based Medicine,Volume 11,2025,100191,ISSN 2666-5212.

[4] Liu, Y., Jain, A., Eng, C. et al. A deep learning system for differential diagnosis of skin diseases. Nat Med 26, 900–908 (2020).

[5] Mostafiz Ahammed, Md. Al Mamun, Mohammad Shorif Uddin, A machine learning approach for skin disease detection and classification using image segmentation,Healthcare Analytics,Volume 2,2022,100122,ISSN 2772-4425.

[6] Filipescu SG, Butacu AI, Tiplica GS, Nastac DI. Deep-learning approach in the study of skin lesions. Skin Res Technol. 2021 Sep.

[7] Zahra Mirikharaji, Kumar Abhishek, Alceu Bissoto, Catarina Barata, Sandra Avila, Eduardo Valle, M. Emre Celebi, Ghassan Hamarneh,A survey on deep learning for skin lesion segmentation,Medical Image Analysis,Volume 88,2023,102863,ISSN 1361-8415.

[8] L. -F. Li, X. Wang, W. -J. Hu, N. N. Xiong, Y. -X. Du and B. -S. Li, "Deep Learning in Skin Disease Image Recognition: A Review," in IEEE Access, vol. 8, pp. 208264-208280, 2020.

[9] SN Hasan, M Gezer, RA Azeez, S Gülseçen - 2019 medical technologies congress (TIPTEKNO), 2019.

[10] K. Thurnhofer-Hemsi, E. López-Rubio, E. Domínguez and D. A. Elizondo, "Skin Lesion Classification by Ensembles of Deep Convolutional Networks and Regularly Spaced Shifting," in IEEE Access, vol. 9, pp. 112193-112205, 2021.

[11] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with

deep neural networks. Nature 542, 115–118 (2017).

[12] Nawal Soliman ALKolifi ALEnezi, A Method Of Skin Disease Detection Using Image Processing And Machine Learning, Procedia Computer Science,Volume 163,2019,Pages 85-92,ISSN 1877- 0509

[13] Muhammad Romail Imran, Abdul Wahab Paracha, Hamza Anjum, Haris Anjum, Muhammad Abbas, & Muhammad Fasih. (2024). SKIN DISEASE CLASSIFICATION USING DEEP LEARNING