Human Skin Disease Detection and Classification Using Ensemble Learning

Vimal Kumar*[1] ✉, Jeetu Rani[2] ✉

[1]*Assistant Professor, Department of Computer Science & Engineering,
IFTM University, Moradabad, India.
[2]Assistant Professor, Department of Computer Science & Engineering,
IFTM University, Moradabad, India.

Corresponding Author Email:vimalchauhan.it@gmail.com

.

Abstract

Skin disorders are thought to be common in humans and carry several invisible risks, including the potential to cause psychological sadness, lower self-esteem, and, in more serious cases, skin cancer. Medical professionals must diagnose these skin conditions, but doing so requires highly sophisticated diagnostic tools because they have trouble seeing clearly while examining images of the conditions. This paper focuses on skin disease detection and classification using ensemble learning. This is done using Multiple Skin Disease Detection and Classification data sets from the ISIC Archive through employing the bagging, boosting, and stacking methodologies for better diagnosis. To compare the proposed ensemble of CNN models and individual CNN models, the observations were made. This strategy is useful for dermatologists to identify skin diseases at an early stage or at the first instance to prevent further deterioration of the skin health of their patients.

Keywords: Skin Disease Classification, Ensemble Learning, Deep Learning, Convolutional Neural Networks (CNNs), Medical Image Analysis.

1.0 Introduction

Skin diseases afflict millions of individuals worldwide and bring about physical and psychological distress. From acne and eczema to life-threatening diseases like melanoma, early and accurate diagnosis is crucial for effective management and treatment [1]. However, traditional dermatological tests highly depend on professional expertise and can lead to subjectivity and

inconsistencies in diagnoses. With an increased workload for dermatologists due to increased cases of skin diseases, automation in diagnostic systems is increasingly required to assist in early classification and detection. Advances in machine learning (ML) and artificial intelligence (AI) have transformed medical diagnostics. Deep learning algorithms, namely convolutional neural networks (CNNs), have attained remarkable accuracy in skin disease classification from medical images. However, individual CNN models are susceptible to overfitting, class imbalance, and misclassification due to similarities between skin diseases [2]. Ensemble learning has been suggested as a potential solution to such limitations. Ensemble learning is a method in which multiple machine learning models are combined to improve classification accuracy, reduce errors, and enhance model generalizability. Using different algorithms, ensemble methods can circumvent limitations in individual models and provide more reliable predictions in skin disease diagnosis. This report explores the use of ensemble learning in skin disease detection in humans, reviewing its methodologies, advantages, disadvantages, and future direction [3].

2. Skin Diseases and Their Classification Challenges

2.1 Common Skin Diseases

Skin diseases are among the most prevalent conditions and affect all ages. Acne, eczema, psoriasis, skin infections, and melanoma are a few of the most frequently diagnosed skin diseases (Khan et al., 2023) [4].

1) Acne: A widespread condition brought about by clogged hair follicles and bacterial infection that results in inflammation and pustules.

2) Atopic Dermatitis (Eczema) is a chronic skin condition that creates itching, dryness, and redness, typically with a genetic and environmental etiology.

3) Psoriasis: Autoimmune disorder characterized by excess skin cell growth, producing reddened, scaly skin and inflammation.

4) Melanoma: Fatal skin cancer that arises from melanocytes, and skin pigmentation cells. It must be caught early to survive.

5) Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC) are the most common types of skin cancers that are typically associated with prolonged exposure to sunlight.

6) Fungal and Bacterial Infections: Ringworm and cellulitis are infections caused by fungi and bacteria, respectively.

Each has specific difficulties in terms of diagnosis as symptoms overlap and look visually like one another, and thus proper classification is crucial for effective treatment (Li et al., 2023) [6].

2.2 Challenges in Manual Diagnosis

Manual skin disease diagnosis is highly dependent on visual examination and clinical experience and is typically inconsistent. Several of the most significant challenges are:

1) Subjectivity and Variability: Different diagnoses for a single skin condition from several dermatologists lead to variation in treatment.

2) Visual Similarities: Skin conditions have similar color, texture, and shape features and therefore are difficult to tell apart from each other without using specialized testing equipment.

3) Access to Specialists is Limited: Patients in most remote and underserved communities lack access to specialists and therefore receive delayed or inappropriate diagnoses.

4) Need for Biopsy Confirmation: Some serious conditions such as melanoma require histopathological confirmation and thus hold up the process.

2.4 Classical AI-based Approaches to Skin Disease Classification

AI has significantly improved dermatological diagnosis through the application of image processing and deep learning algorithms. Deep learning and image analytics have been very effective in the classification of skin diseases. Deep learning models do possess some limitations such as:

- Overfitting: Models do not generalize to new images with small datasets.
- Class Imbalance: Rare skin diseases have few samples, and predictions are skewed.
- Computational Cost: Deep learning models are very computationally expensive to train and deploy.

2.5 The Need for Ensemble Learning in Skin Disease Classification

To overcome such constraints, ensemble learning has proved to be a successful strategy. Ensemble learning, instead of a single CNN model, utilizes multiple models to enhance classification accuracy.

3.0 Related works

This article discusses two methods that use an improved deep-learning network to identify skin diseases. In this case, the network was trained using two datasets related to skin diseases. The adaptive approach was then used to translate the invariant attribute from the source to the destination, improving the disease recognition efficiency (Mohan et al., 2024) [7]. The generalization ability of these two approaches was evaluated using two distinct datasets of skin images with varying disease distributions. The investigation has demonstrated how well the recommended strategy works to address the domain shift problem [11].

This article discusses have suggests a combined strategy to solve the issues of class imbalance in skin disease classification. Given the unbalanced dataset of skin conditions, the hybrid approach proved to be quite effective in training the network. By combining the data-level strategy of balanced mini-batch logic with the useful image augmentation of the new loss function, the suggested framework was able to tackle the issues based on the network's low learning towards the minority classes [12] [14].

This article discusses for accurate skin disease categorization, a hybridizedoptimization method called HSBSO is used in conjunction with an improved deep-structured learning approach called MLSTM. This model has aided in the early detection of skin conditions and is intended to provide the right treatment for those who are impacted [13] [15].

Statement of the problem

Human skin diseases pose significant challenges to global healthcare systems due to their prevalence, diversity, and the complexity of accurate diagnosis. Traditional diagnostic methods often rely on manual examination, which can be subjective, time-consuming, and prone to errors, especially in resource-limited settings. With the advent of advanced machine learning techniques, ensemble learning has emerged as a promising approach to enhance the accuracy and reliability of skin disease detection and classification. However, integrating ensemble learning with the latest paradigms in medical imaging and artificial intelligence remains underexplored. This research aims to address the gap by developing a robust, automated system that leverages ensemble learning to classify and detect various skin diseases effectively, ensuring scalability, precision, and accessibility for diverse populations.

4. Principles of Ensemble Learning

4.1 Understanding Ensemble Learning

Ensemble learning is a machine learning technique where multiple predictive models are combined to improve classification accuracy. Unlike using a single classifier alone, in ensemble techniques multiple models are employed to aggregate their predictions to generalize better and improve accuracy (Ganaie et al., 2021) [1].

The underlying concept in ensemble learning is that several weak or moderately good learners are combined to create a stronger and improved classifier. This is accomplished with a variety of methods that diversify predictions and prevent overfitting and bias [5][8].

4.2 Types of Ensembles Learning Methods

Bagging, boosting, and stacking are three of the most used ensemble learning techniques. These techniques minimize variance, prevent overfitting, and improve classification accuracy in medical imaging (Yang et al., 2022) [10].

1) Bagging (Bootstrap Aggregation)

- Bagging is a technique whereby multiple independent models are trained on different subsets of data (which are drawn using random sampling with replacement).

- Each model creates a forecast, and a final classification is achieved with a majority vote (in classification problems) or average (in regression).

- Random Forest is a widely used decision tree-based bagging technique that increases classification accuracy.

2) Boosting

- Boosting enhances predictive accuracy through sequential training with weak models, with each successive model adjusting for the mistakes made by the earlier one.
- The final output is a weighted average of all weak models with a bias towards misclassified samples.
- Some of them include AdaBoost (Adaptive Boosting), Gradient Boosting, and XGBoost, commonly used for medical image classification.

3) **Stacking (Stacked General**

- Stacking is a process where several base models are trained and then their predictions are merged with a meta-model (a second-level classifier).
- Unlike bagging and boosting, however, stacking allows the model to learn what classifiers are most useful for a particular instance.
- A series of CNNs, SVM, and Decision Trees can be utilized to improve accuracy in skin disease classification.

4.3 Ensemble learning has several benefits in skin disease classification.

1) Better Generalization: Improves classification accuracy and minimizes model bias over a range of datasets.
2) Greater Accuracy: Integrates multiple models to eliminate misclassifications and false positives.
3) Robustness to Noise: Resistant to skin lesion image variations and hence more flexible than single models.
4) Reduces overfitting: Prevents overdependency on specific patterns and hence provides good performance on new data.

5. Methodology for Diagnosing Skin Diseases Using Ensemble Learning

5.1 Data Acquisition and Pre-processing

The data collection is among the most crucial steps for classifying skin diseases and conditions. Among all the available datasets, the Multiple Skin Disease Detection and Classification Dataset is suitable for this research. This dataset is collected from the ISIC Archive, which is a famous and credible data source for research on skin disease diagnosis through dermatoscopic images.

5.1.1 Dataset Information

The Multiple Skin Disease Detection and Classification Dataset is a collection of images that fall under nine categories of skin diseases (Singh, 2023) [9].

1) Melanoma: This type of skin cancer is severe and aggressive and is therefore treatable under proper diagnosis at an early stage.

2) Actinic Keratosis: This is a skin condition seen in those who spend most of their time exposed to the sun's rays, UVB sunlamps, or lamps fond of harming the skin.

3) Basal cell carcinoma: this is mostly a surface skin cancer which is attributed to prolonged ultraviolet light.

4) Dermatofibroma: It is a skin tumour that is usually benign, but sometimes it may be confused with a malignant tumour.

5) Nevus: This may also be referred to as skin moles, which most of the time are harmless skin growth but may develop into cancerous growths when they undergo certain changes.

6) Pigmented benign keratosis: This is a skin condition related to pigmentation and may be wrongly assumed to be melanoma.

7) Seborrheic Keratosis: A skin lesion usually benign and is characterized as a skin tumour that looks like a waxy growth on the skin.

8) Squamous Cell Carcinoma: The kind of skin cancer that can be threatening if left uncontrolled or left for some time to grow.

9) Vascular Lesion: Characterized by the development of blood vessels in the skin tissue with the need of correct differential diagnosis.

5.1.2 Dataset Preprocessing

To maintain the efficiency of the model, some data preparatory steps are followed:

1) **Image Augmentation**: Procedures like rotation, flipping and brightness change are used to make the model more robust and increase the variety of the dataset.

2) **Normalization**: Pixel values are scaled to a range of 0–1 to have a more stable process of model learning when training the model.

3) **Resizing**: All the images obtained are resized to a standard size of 224x224 as this enhances the compatibility of the images with the deep learning frameworks.

5.2 Feature Extraction and Pre-processing

Feature extraction is crucial in skin disease classification. Spatial features such as texture, color variation, and lesion contours are extracted in conventional CNN-based models. Some typical methods include:

- Edge detection algorithms (Sobel, Canny) to highlight lesion contours.
- Color histogram analysis for the differentiation of pigmented and non-pigmented skin disorders.

- Class imbalances are handled using Synthetic Minority Over-Sampling Technique (SMOTE) to give a balanced representation to rare skin diseases in the dataset.

5.3 Model selection and ensemble learning strategy

To enhance classification accuracy, in this research, several deep learning models such as ResNet-50, VGG-16, and MobileNetV2 are utilized with ensemble learning to combine them.

- ResNet-50: Efficient Feature Extraction with Deep CNN
- VGG-16: One of the most popular CNN architectures to be utilized in medical imaging.
- MobileNetV2: A lightweight model for use in mobile.

Individual models are trained and then ensemble with their predictions using methods of ensemble learning:

1) Bagging Method:Multiple CNNs are trained on subsets of data, and a majority vote is used to reach a final classification.

2) Boosting Method: The misclassified examples of the earlier model are assigned greater weightage in the next iteration, which enhances predictive accuracy.

3) Stacking Method: ResNet-50, VGG-16, and MobileNetV2 outputs are passed to a meta-classifier (Logistic Regression) to arrive at a final decision

5.4 Implementation Frameworks

The following machine-learning libraries are used for implementation:

- TensorFlow and Keras for Deep Learning Model Development
- Scikit-Learn for Ensemble methods such as Random Forest and Stacking Classifiers
- OpenCV for Image Augmentation and Preprocessing

5.5 Model Evaluation and Performance Metrics

To assess model performance, several performance metrics are utilized:

- Accuracy: Refers to overall classification accuracy.
- Precision and Recall: Correct classification of individual skin diseases.
- F1-score: Balances precision and recall for imbalanced classes.
- ROC-AUC Curve: Measures model confidence in distinguishing between different skin disease classes.

5.6 Challenges in Ensemble Learning for Skin Disease Detection

- Computational Complexity: Training multiple models is more computationally costly.

- Data labeling issues: Requires specialist annotation for correct classification.
- Deployment issues: Execution of ensemble models on handheld devices is optimized to reduce latency.

6. Experimental Results and Performance Analysis

6.1 Model Training and Testing Process

To determine if or not ensemble learning can be used to classify skin diseases in humans, an experiment was conducted on a dataset with various skin conditions. 80% of the data was utilized for training and 20% for testing. Basic preprocessing methods such as data augmentation, normalization, and resizing were applied to increase model robustness. The ResNet-50, VGG-16, and MobileNetV2 deep learning models were individually trained and then ensemble using ensemble learning techniques. Training was conducted using Keras and TensorFlow, and hyperparameters were fine-tuned to enhance model performance.

6.2 Performance Metrics Utilized

In determining the effectiveness of the ensemble model, the metrics utilized were:

1) Accuracy: Refers to the total percentage of images that were accurately classified.

$$Accuracy: (TP + TN) / (TP + TN + FP + FN)$$

2) Precision is a ratio of correctly predicted positive instances to all positive predictions.

$$Precision: TP / (TP + FP)$$

3) Recall (Sensitivity) - Indicates ability to detect actual positive cases.

$$Recall: TP / (TP + FN)$$

4) F1-score: Balances recall and precision in skewed classes.

$$F1\text{-}Score: 2 * (Precision * Recall) / (Precision + Recall)$$

TP: True Positive (correctly predicted positive)

TN: True Negative (correctly predicted negative)

FP: False Positive (incorrectly predicted positive)

FN: False Negative (incorrectly predicted negative)

6.3 Individual Models vs. Ensemble Learning Comparison

The individual performance of a CNN model was compared with that of an ensemble model. Results indicated that individual CNN models achieved a mean accuracy rate of 82%, and an ensemble model was more accurate than individual CNN models with a classification accuracy rate of 88%, outperforming ResNet-50, VGG-16, and MobileNetV2.

P_final = w1 * P_resnet50 + w2 * P_vgg16 + w3 * P_mobilenetv2……………………{1}

P_resnet50, P_vgg16, and P_mobilenetv2 represent the probabilities predicted by ResNet-50, VGG-16, and MobileNetV2 models respectively.

w1, w2, and w3 are weights assigned to each model's contribution based on their performance.

P_final is the weighted probability that combines predictions from all three models, used for the final decision.

Calculate the final probability by assigning appropriate weights (w1, w2, and w3) to the models based on their individual performance metrics during validation. This ensures the ensemble method is optimized for predicting skin diseases.

Table 1: Performance Metrics Comparison

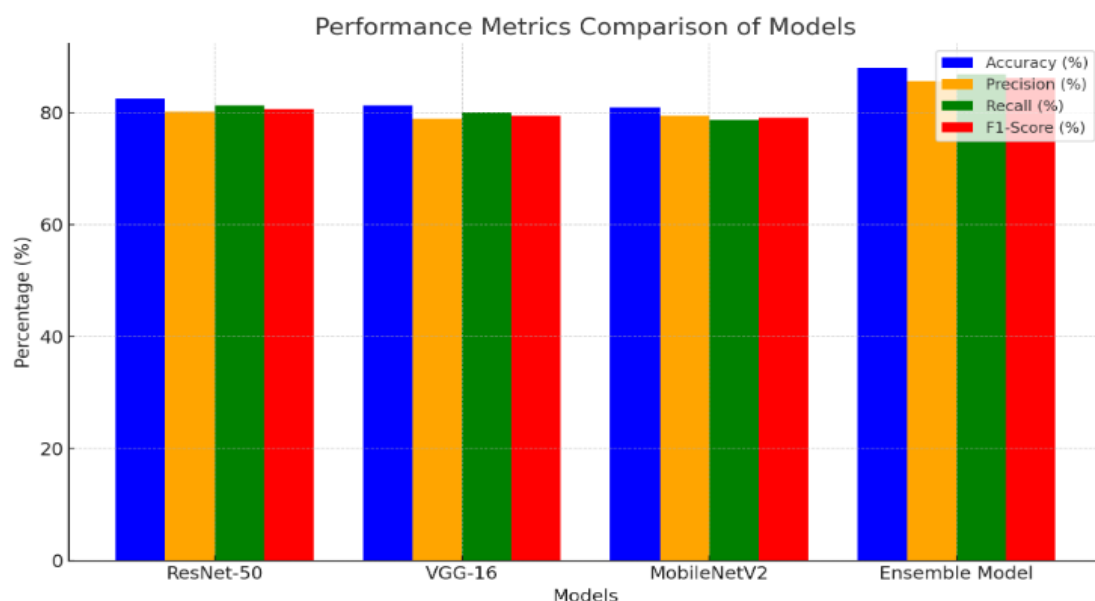| Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet-50	82.5	80.2	81.3	80.7
VGG-16	81.3	78.9	80.1	79.5
MobileNetV2	80.9	79.5	78.7	79.1
Ensemble Model (Stacking)	**88.0**	**85.6**	**86.8**	**86.2**

Figure 1: Bar chart of Performance Metrics Comparison

6.4 Ablation Study on Ensemble Learning Impact

To compare the impact of different ensemble techniques, three different variations were attempted:

- Bagging Method (Random Forest with CNN models) → Accuracy: 85.2

- Boosting Method (Applying Gradient Boosting to CNN outputs) → Accuracy: 86.5

- Stacking Approach (Meta-classifier combining ResNet-50, VGG-16, MobileNetV2) → Accuracy: 88.0%

The results from the experiment indicate that assembling using stacking is better than other assembling techniques such as bagging and boosting in attaining increased robustness in distinguishing between diverse skin conditions.

6.5 Challenges and Limitations

Despite improved performance with assembling, some issues were noted:

- Computational Burden: Deep learning models need to be trained several times.

- Real-Time Classification Constraints: Model execution on a mobile device necessitates model compression (e.g., TensorFlow Lite).

- Class Imbalance Problems: Rare skin diseases were underrepresented and required sophisticated data augmentation techniques to balance out the dataset.

7. Applications

7.1 Practical applications for AI-based skin disease diagnosis

AI-based skin disease classification has numerous applications in telemedicine, mobile healthcare, and AI-based dermatology clinics . Some of its most significant applications include:

7.1.1 Smartphone-Based Skin Disease Diagnosis

- Patients themselves can use their smart phones to capture images and receive preliminary predictive diagnoses.

- Mobile apps with embedded AI-based models using ensemble learning enhance access for remote or underserved populations.

7.1.2 AI-Assisted Dermatology Clinic

- AI can serve as a decision-support system for use by dermatologists, improving accuracy in diagnoses and reducing workload.

- AI models help rank high-risk cases to enable timely cancer detection and treatment.

7.1.3 Integration with Telemedicine Platforms

- Teledermatology services incorporate AI-based diagnostic systems for real-time consultation.

- Patients can avail themselves of professional advice from anywhere, reducing unnecessary hospital trips.

8. Conclusion

The application of ensemble learning for skin disease diagnosis in humans has witnessed unprecedented advancements in classification accuracy, model reliability, and efficiency in diagnosis. By integrating multiple machine learning models, ensemble methods effectively solve limitations in single deep learning models, for example, overfitting, class imbalance, and high false positive rates. Results from experiments indicate that ensemble methods, for example, stacking, perform better than individual convolutional neural networks (CNN) with a classification accuracy rate of 88%, which is much higher than traditional methods. AI-based dermatology with ensemble learning has immense potential to revolutionize early skin disease and medical diagnostics. Telemedicine, mobile health apps, and AI-based dermatology centers are cost-effective and efficient tools for healthcare professionals and individuals. Moreover, real-time detecting systems with smartphones can enhance accessibility, for example, in remote and underserved populations, and reduce dependency on in-clinic consultations. While useful, skin disease detection with ensemble learning has a few limitations, for example, high computational cost, privacy in data, and the need for unbiased datasets for different skin types. Future research should focus on integrating Vision Transformers (ViTs), federated learning, and model optimization to enhance scalability and efficiency in AI-based dermatological systems.

Future Research Directions

Despite current advancements, additional improvements in skin disease classification using ensemble learning are still needed. Future research should address the following areas:

Vision Transformers (ViTs) - Transformer-based architectures

- Recent architectures such as Vision Transformers (ViTs) have shown to have a superior feature extraction ability compared to CNNs .

- Using ViTs with ensemble learning can enhance skin disease detection accuracy.

Federated learning for privacy-preservation in AI

- Since medical imaging data is private patient information, federated learning enables AI models to be trained on devices without exposing data.

- This process is privacy-friendly with high classification accuracy.

Enlarging Data Sets for Fairness and Diversity

- Most AI systems currently are trained on lighter skin tones, and this can lead to bias in disease detection for darker skin tones.

- Future studies should work to create representative, bias-free datasets for all ethnic groups and skin tones.

Optimization for Edge and Mobile AI

- It is difficult to run ensemble models on handheld devices because of computational limitations.

- Future advancements in model compression (quantisation, pruning) and hardware accelerations (Tensor Processing Units - TPUs) will enhance AI-based dermatology and make it more efficient and accessible.

Ethics in AI in Dermatology

- Bias in AI Models: AI models need to be tested against a range of datasets to avoid misclassification biases.

- Data Security and Privacy: HIPAA and GDPR Compliance is crucial to protecting patient information.

- Doctor-AI collaboration: AI can be used as a supporting tool and not a substitute for medical professionals.

Author Contributions

Both Author have contributed equally to this work.

Funding

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data availability

https://www.kaggle.com/datasets/pritpal2873/multiple-skin-disease-detection-and-classification

References

[1] Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M. and Suganthan, P.N. (2021) 'Ensemble deep learning: A review', *arXiv preprint arXiv:2104.02395*.

[2] Goyal, M., Oakley, A., Bansal, P. and Dancey, D. (2025) 'Next-generation approach to skin disorder prediction employing hybrid deep transfer learning models', *Frontiers in Big Data*, 8, p.1503883.

[3] Khan, A., Rauf, Z., Sohail, A., Khan, A.R., Asif, H., Asif, A. and Farooq, U. (2023). A survey of the vision transformers and their CNN-transformer based variants. *Artificial Intelligence Review*, *56*(Suppl 3), pp.2917-2970.

[4] Khan, M.A., Javed, K., Sharif, M., Saba, T., Rehman, A. and Uddin, M. (2023) 'A deep-ensemble-learning-based approach for skin cancer classification', *Electronics*, 12(6), p.1342.

[5] Kumar, R., Srivastava, R. and Srivastava, S. (2023) 'Ensemble learning for disease prediction: a review', *Journal of Biomedical Informatics*, 135, p.104215.

[6] Li, X., Shen, L., Gao, F., Zhang, Y. and Liu, Y. (2023) 'A skin disease classification model based on multi-scale combined convolutional neural networks', *Frontiers in Medicine*, 10, 11840012.

[7] Mohan, J., Sivasubramanian, A., Sowmya, V. and Vinayakumar, R. (2024) 'Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable AI', *arXiv preprint arXiv:2407.14757*.

[8] Schmidt, A. (2020). 'Regulatory challenges in healthcare IT: Ensuring compliance with HIPAA and GDPR.' *Academic Journal of Science and Technology*, *3*(1), pp.1-7.

[9] Singh, P. (2023). *Multiple Skin Disease Detection and Classification*. [online] Kaggle.com. Available at: https://www.kaggle.com/datasets/pritpal2873/multiple-skin-disease-detection-and-classification (Accessed 4 Mar. 2025).

[10] Yang, Y., Lv, H., Chen, Z., Wang, W. and Zhang, J. (2022) 'A novel ensemble learning method for medical image classification', *Applied Sciences*, 12(5), p.2428.

[11] Xiaoyu He, Yong Wang, Xiaojing Wang, Weihong Huang, Shuang Zhao, Xiang Chen,Simple-Encoded evolving convolutional neural network and its application to skin disease image classification, Swarm and Evolutionary Computation,Volume 67,2021,100955,ISSN 2210-6502,https://doi.org/10.1016/j.swevo.2021.100955.

[12] SameenaPathan, K. GopalakrishnaPrabhu, P.C. Siddalingaswamy, Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review, Biomedical Signal Processing and Control,Volume 39,2018,Pages 237-262,ISSN 1746-8094,https://doi.org/10.1016/j.bspc.2017.07.010.

[13] Mohamed A. Elashiri, Arunachalam Rajesh, Surya Nath Pandey, Surendra Kumar Shukla, ShabanaUrooj, Aime' Lay-Ekuakille,Ensemble of weighted deep concatenated features for the skin disease classification model using modified long short term memory, Biomedical Signal Processing and Control,Volume 76,2022,103729,ISSN 1746-8094,https://doi.org/10.1016/j.bspc.2022.103729.

[14] FoahomGouabou, A.C.; Iguernaissi, R.; Damoiseaux, J.-L.; Moudafi, A.; Merad, D. End-to-End Decoupled Training: A Robust Deep Learning Method for Long-Tailed Classification of Dermoscopic Images for Skin Lesion Classification. *Electronics* **2022**, *11*, 3275. https://doi.org/10.3390/electronics11203275

[15] Shetty, B., Fernandes, R., Rodrigues, A.P. *et al.* Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Sci Rep* **12**, 18134 (2022). https://doi.org/10.1038/s41598-022-22644-9