Transfer Learning Based Weighted Deep Learning Ensemble Model for Medical Image Classification

[1]Giddaluru Lalitha and [2]Riyazuddin Y MD
[1,2]Department of CSE, School of Technology, GITAM (Deemed to be University), Hyderabad, Telangana, India.
[1]lgiddalu@gitam.edu, [2]rymd@gitam.edu

Correspondence should be addressed to Giddaluru Lalitha : lgiddalu@gitam.edu

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404063
Received 10 August 2023; Revised from 12 December 2023; Accepted 08 June 2024
Available online 05 July 2024.

Abstract – Malignant melanoma is a well-known and deadly form of cancer that originates from epidermal melanocytes in humans. Early detection of such diseases, including various forms of cancer, is necessary for speeding up diagnosis and enhancing patient outcomes. A novel transfer learning-based ensemble-deep learning model was presented for diagnosing diseases at a preliminary stage. Data augmentation was used to increase the dataset, and integration of Inception-v3, DenseNet-121, and ResNet-50 techniques, along with an ensemble method, was employed to overcome the scarcity of labeled datasets and increase the accuracy as well as make the model more robust. The proposed system was trained and tested employing the International Skin Imaging Collaboration (ISIC) dataset. The suggested ensemble model gained the best performance, producing 98% accuracy, 98% area under the curve, 98% precision, and 98% F1 score. The proposed model outperformed the existing state-of-the-art models in disease classification. Furthermore, the proposed model will be beneficial for medical diagnosis and reduce the incidence of various diseases.

Keywords – Ensemble Model, Skin Disease, Classification, Deep Learning, Transfer Learning.

I. INTRODUCTION

Tumors form when normal cells have genetic abnormalities that cause them to grow uncontrollably, leading to cancer. Tumors can be classified as either malignant (cancerous) or benign (non-cancerous) [1]. Malignant tumors can damage tissue nearby and spread to other areas of the body. The skin is the largest organ of the human body, covering an area of about 20 square feet, and is essential for sensing pressure, temperature, and touch. Significant environmental changes have led to an increase in the incidence of skin diseases in humans. Skin cancer, including both melanoma and non-melanoma, is increasingly recognized worldwide, particularly in Caucasian populations [2]. It is the most common type of cancer worldwide, with more than 3.5 million cases diagnosed annually, including breast, lung, and colon cancers [3]. Melanoma claims a life every 57 seconds. As with all cancers, early screening and detection offer the best chance for a full recovery. Early detection of skin cancer shows a promising ten-year survival rate of 94%. However, this rate drops significantly as the cancer progresses to later stages. In the case of melanoma, the ten-year survival rate drops to only 15% when diagnosed at an advanced stage. Melanoma is challenging to diagnose early and can be complex and tough to understand, even for experts. Therefore, simplifying the diagnostic process can be advantageous for professionals [4]. The broad classification of skin lesions in dermatology complicates automatic skin cancer diagnosis. However, several automated methods have been developed by researchers for the detection and diagnosis of skin cancer [5]. There were two primary issues with skin image categorization systems: insufficient data and image quality. To categorize skin images, traditional methods required extensive preprocessing, segmentation, and feature extraction. Nevertheless, today, machine learning (ML) can perform categorization tasks without manual feature extraction because of AI developments that are comparable to those seen in popular tech use. Recently, there has been a trend of using ML approaches to diagnose cancer. ML algorithms have increased the accuracy of cancer prediction by 15% to 20% within the last 20 years. Deep learning (DL) is a fast-developing topic with a wide range of possible applications and one of the best ML techniques for classifying and identifying images [6].

This paper introduces an innovative Transfer learning with an Ensemble -DL model with the purpose of early skin disease recognition and different state-of-the-art algorithms are compared for this reason. The developed model is undergoing extensive training to address the drawback of a poorly imbalanced dataset, data scarcity, and image quality in the existing skin image categorization systems and how it could affect the accuracy of the model. Even though maximum papers achieved outstanding accuracy, there are some overfitting issues, and the model could be performing nicely for

particular classes. So, we are therefore attempting to fill those gaps. The research focuses on improving the early detection of skin diseases, mainly cancer, given its rising global prevalence. This motivation to offer healthcare professionals enhanced resources for timely diagnosis, potentially saving lives. The research is driven by the urgent need to improve diagnostic accuracy, particularly for skin cancer, where early detection significantly improves patient outcomes.

The primary contribution of the study is as follows:

- We have designed an Ensemble DL model for the classification of skin cancer.
- We used data augmentation techniques for improved dataset size, enhanced generalization, robustness to variability, balance classes, and Reducing Annotation.
- In our research, we have used transfer learning as a base model, and we have fine-tuned the transfer learning model according to our research.
- We have conducted several experiments, including different combinations of base learners, different optimizers, and a set of learning rates.

II. LITERATURE REVIEW

Several scholars have delved into several studies to increase efficacy in the field of skin cancer classification over the years. Here, we present some notable and recent studies on skin cancer research.

Early detection of skin diseases can save lives. To address this, Nigar et al. [7] suggested a new DL model for analyzing six familiar skin conditions: actinic keratosis, benign keratosis, melanoma, basal cell carcinoma, insect bites, and skin acne. They compared different algorithms employing standard datasets and discovered that convolutional neural networks (CNN) achieved the highest accuracy at 97%, with precision, recall, and F1-score all at 91%. In another study, Riaz et al. [8] proposed a joint learning system combining CNN and Local Binary Patterns (LBP) and linked features from both architectures and testing on a skin cancer dataset. Results demonstrated that the fused architecture acquired 98.60% accuracy and 97.32% validation accuracy, showing robustness. Lakshmi et al. [9] developed a Hybrid AI Model (HAIM) for skin cancer classification by using different representation techniques for feature extraction and an efficient Exponentially Weighted and Heaped Multi-Layer Perceptron (EWHMLP) for classification. The HAIM included Curvelet (CurT), Contourlet (ConT), and Shearlet (SheT) transforms for feature extraction. While MLP was appropriate for classification, weight learning was tough, and optimization was often required to converge, leading to instability. To address this, EWHMLP was designed, and results indicated that the combination of these transforms performed 98.33% accuracy in multi-class classification on the PH2 database. In another study, Ogundokun et al. [10] introduced a hybrid-deep CNN architecture for skin cancer identification, combining two main heuristics: Xception and MobileNetV2 models. They presented data augmentation to balance the dataset and used transfer learning to manage challenges related to the lack of labeled datasets. Xception, alongside MobileNetV2, delivered the best performance and performed 97.56% accuracy, 97.00% area under the curve, 100% sensitivity, 93.33% precision, 96.55% F1 score, and 0.0370 false positive rate on the evaluated dataset.

Natha et al. [11] incorporated Contourlet Transform (CT) and Local Binary Pattern (LBP) techniques to identify skin cancer image features accurately. To handle computational costs and overfitting, they employed Particle Swarm Optimization (PSO) to select key features and used a Support Vector Machine (SVM), Random Forest (RF), and Neural Networks (NN) for classification. SVM had the quickest training time (0.0458 seconds), followed by NN (0.08730 seconds) and RF (0.1622 seconds). SVM and NN are appropriate for real-time applications, and their model performed the highest accuracy of 86.9% compared to state-of-the-art models. After reviewing the entire literature, we discover particular Skin cancer-related limitations.

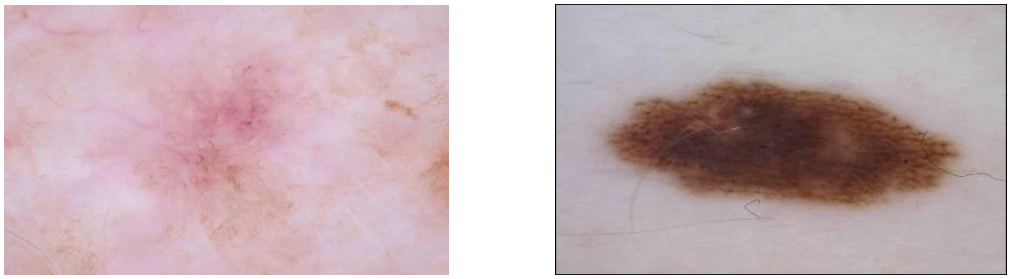
- The drawbacks of the dataset include the primarily fair-skinned representation, which may distort results; also, the preponderance of small lesions in the current datasets presents difficulties for training because of data scarcity, imbalance, and limited diversity [12], [13], [14].
- Despite progress, there is still a need to address the scaling of datasets through extensive augmentation and analysis of enriched imaging data, such as hyperspectral images [15].
- Time-consuming classifier training hampers efficiency because boosting techniques have a slow convergence time, and high-dimensional feature sets are complicated [16].
- While algorithms show unique strengths, the persistent challenge of optimization and generalization gives a notable gap in algorithmic applicability [16].
- The strength and robustness of dermatological studies need to be improved by the low generalizability of study populations concerning age, skin type, and risk profiles [17].
- It is essential for comprehensive data augmentation techniques, such as synthetic image generation through GANs, to emphasize a gap in improving model resilience and adaptability to different scenarios [18].
- Extracting features from skin lesions faces notable accuracy constraints, posing a crucial gap in the accuracy and reliability of diagnostic systems [5].
- Despite improvements, frameworks still require work on high complexity in terms of time and memory usage, defining a marked gap in computational efficiency and scalability [19]

III. PROPOSED METHODOLOGY

The research was conducted in several steps. Each of the steps is described as follows:

Data Collection and Augmentation

In our research, we have used the International Skin Imaging Collaboration (ISIC) dataset, and this set consists of 2357 images of malignant and benign oncological diseases. All images were sorted according to the classification taken with ISIC, and all subsets were divided into the same number of images, except for melanomas and moles, whose images are slightly dominant. The data set contains the following diseases: actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, vascular lesion. After that, we converted the dataset into two classes: Benign and Melanoma. We assumed that the use of subpar devices by unskilled users would be the most frequent issue related to the practical implementation of the suggested strategy in isolated rural locations in terms of image quality degradation. Because the camera was used in real-world situations, motion blur, overexposure, and low resolution of the user device camera are predicted to affect the photographs. Resolution downsampling: To meet the neural network's minimum input size requirement, the original images were down-sampled to 32 ~ 32 (64 ~ 64, 96 ~ 96, and 192 ~ 192) pixels and up-sampled to 224 ~ 224 pixels. As there is a limited number of samples in the dataset, we increase the dataset using data augmentation techniques. We have performed scaling, rotation, zoom, and height width shifting. We used a range for performing these techniques as an example. Rotation range ± 15 degrees means that the image rotated from -15 degrees to +15 degrees from its original position. The sample of the dataset is shown in **Fig 1**.



a) Melanoma b) Benign

Fig 1. Sample of the Dataset.

Proposed Model

In the research, firstly, we employed transfer learning models, namely Inceptionv3, ResNet50, and DenseNet121. In the transfer learning models, we have customized the models according to our research. Since the final fully connected layers were utilized to categorize ImageNet's 1,000 classes, they have been eliminated. A modification is made by freezing all of the pre-trained layers' weights to tackle the new classification challenge. The model is extended by the addition of two fully connected layers. The fully connected back layer's output provides the classification for every class. After customizing the transfer learning models, we used them as base learners of the ensemble model. In our research, we have proposed a novel ensemble model. Once the pre-trained models were deployed one at a time, we went up to the transfer learning ensemble. Using weighted average approaches, we ensembled to consider all potential combinations of classifier models. Since the fundamental results of the base models might affect the ensemble outcomes, the weighted average ensembles have particular implications on the ensembled outcome. Consequently, the model with the highest performance level supersedes the others. Assume for the moment that we have 'C' test instances and 'S' pre-trained models. An instance, each model categorizes Ci into one of the pre-established groups from the n class. Hence, a model Sj provides a softmax probability distribution vector (prb[]) of size n class for each instance Ci. As a result, the output is:

Model 1:
prb11, prb12, prb13, ..., prb1n **:::** prb21, prb22, prb23, ..., prb2n **::::** ...prbC1, prbC2, prbC3, ..., prbCn

Model 2:
prb11, prb12, prb13, ..., prb1n **:::** prb21, prb22, prb23, ..., prb2n**:::** ...prbC1, prbC2, prbC3, ..., prbCn

Model 3:
prb11, prb12, prb13, ..., prb1n **:::** prb21, prb22, prb23, ..., prb2n**:::**...prbC1, prbC2, prbC3, ..., prbCn

Here, Models 1 to 3 represent the three pre-trained models, and prb_ij means the softmax probability that the instance i belongs to class j.C means the number of test instances.n define the number of classes.

A weight based on the models' softmax probabilities is used in the weighted average ensemble technique. The related precision score (pre), recall score (rec), fl-score (fi), and AUC score (AUC) are produced by generating and comparing the

predictions of the i[th] model (') with the true labels (y). This, it is assumed, creates an array A = {pre, rec, f, AUC"). The hyperbolic tangent function is then used to calculate the weight (w) assigned to each classifier, as seen below.

$$w(i) = \sum_{x \in A(i)}^{i} \tanh(x) \quad (1)$$

$$\text{Ensemble prob} = \frac{\sum w^{(i)} * p_j^{(i)}}{\sum w^i} \quad (2)$$

Table 1. Hyperparameters Of Pre-Trained Models (Base Learners)

Hyper Parameter	Value
Optimizer	Adam, Nadam
Loss	Cross Entropy
Learning Rate	0.0001, 0.00001
Epochs	50

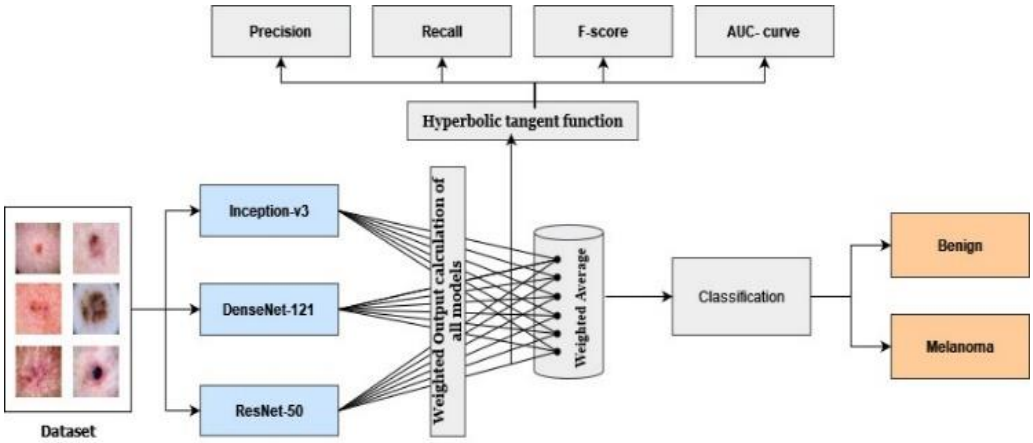


Fig 2. Proposed Ensemble Model Schema.

The equation calculates these weights (1) and multiplies them by the decision scores of corresponding pre-trained models (DenseNet-121, ResNet-50, Inception-v3) to compute the weighted average probability ensemble. Finally, the proposed model is predicted the class is computed by Equation 3.

$$\text{Prediction} = \text{argmax}(\text{Ensemble\_prob}) \quad (3)$$

The graphical representation of the proposed model is shown in **Fig 2**, and the hyperparameters used in the pre-trained model are shown in **Table 1**.

IV. EXPERIMENTAL ANALYSIS & RESULTS

For conducting the research, we have divided our dataset into three parts training, testing, and validation .10% data was used for validation, and 80%data was used for training. The rest of the amount is used for testing purposes to see how our proposed model can perform on unseen data. We have also used Nadam and Adam optimizers and 0.0001 and 0.00001 learning rate in the proposed model where Adam with 0.00001 performed best for the proposed ensemble model**.** After developing the model, we evaluated each one individually, operating accuracy, recall, F-score, and precision calculations. We also displayed the training, validation, loss, and accuracy curves for the suggested model. In the end, we calculated the confusion matrix to analyze the error of the model occurred.

Table 2. Performance Of the Testing Accuracy, Training Accuracy, And Validation Accuracy for The Proposed Model

Model	Testing Accuracy	Training Accuracy	Validation Accuracy
ResNet50	86%	76%	76%
InceptionV3	93%	97%	83%
DenseNet121	97%	97%	87%
Ensemble	98%	98%	89%

For several models, we evaluated the testing, training, and validation accuracy. Each model showed specific characteristics and performance patterns, which were indicative of variations in architecture, optimization methods, and

potential challenges like overfitting. **Table 2** presents the testing accuracy, training accuracy, and validation accuracy of individual models. Let's discuss individual models and then compare them to specify which one presented the best result.

The ResNet50 model architecture, which was optimized with Adam Optimizer, demonstrated a qualified test accuracy of 86%. However, its validity and training accuracy, both 76%, were relatively poor. This indicated a potential problem of underfitting, where the model fitted the training data, leading to poor generalization.

In contrast, the Inception architecture, which was optimized by Nadam optimizer, achieved the highest testing accuracy at 93%. However, a big difference between its 97% training and 83% validation accuracy was observed. This gap indicated overfitting, wherein the model overly memorizes patterns in the training data, restricting its ability to generalize effectively. Nonetheless, its superior testing accuracy meant a powerful capability to discriminate between unseen samples, with likely issues of over-reliance on training data.

The DenseNet-121 architecture, along with the Adam optimizer, exhibited a high level of testing accuracy at 97%. Additionally, it preserved a relatively small gap between its training and validation accuracies, with 98% and 87%, respectively. This suggested that the model is better at generalizing to new data, with less overfitting approximated to other models. Although there is a small difference between training and validation performance, the model's power to manage unseen data indicated that it has effectively learned and extracted features from the training set.

After evaluating three particular models, the ensemble model was designed by incorporating their predictions. It reached the highest testing accuracy of 98%, corresponding to the training accuracy of 98% and a negligible validation accuracy of 89%. This implied that the ensemble model has outstanding generalization when compared to individual models, indicating that it can satisfactorily predict results with new and unseen data. The ensemble model showed how collective intelligence may improve overall predictive performance by combining the strengths of many architectures and minimizing the shortcomings of each one separately.

Every model provided specific perspectives on the relationship between architecture, optimization, and generalization. Although transfer learning models such as ResNet50, InceptionV3, and DenseNet-121 faced overfitting or poor generalization issues, merging various perspectives through ensemble approaches underscored the ability to perform better predictive accuracy and robustness. When we applied the ensemble model, the overfitting was reduced, and the model can perform best on unseen data.

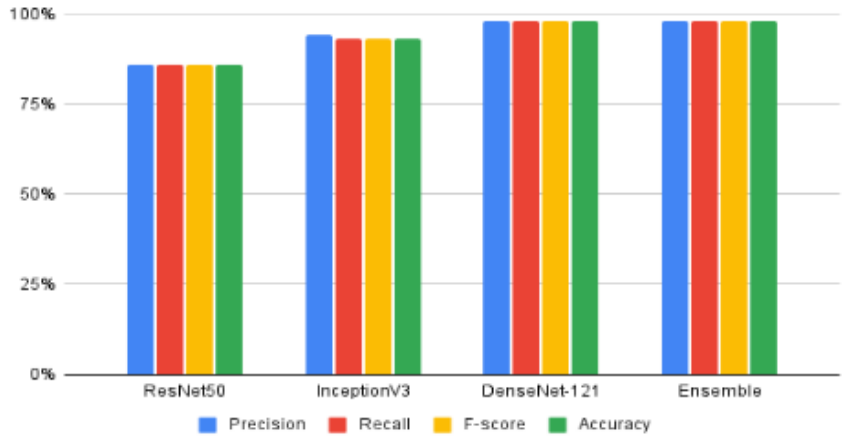


Fig 3. The Evaluation Metrics Of The Models.

In **Fig. 3**, the ensemble model displayed exceptional performance across all measures, with precision, recall, F-score, and accuracy, all at 98%. The high precision of the ensemble model, 98%, indicated its capability to identify positive instances while minimizing false positives accurately.

Moreover, the ensemble model's recall score of 98% showed its ability to effectively capture an extensive majority of positive instances from the dataset. This high recall provided that the model can detect most of the relevant information, which is mainly important in applications where missing positive instances could be detrimental. The F-score of 98% highlighted the ensemble model's balanced performance in terms of both precision and recall. The ensemble model appeared as the best performer among the evaluated models, excelling in precision, recall, F-score, and accuracy. It's important to train and validate metrics when assessing how well a model performs, specifically for tasks like detecting skin cancer. Training and validation accuracy, as well as training and validation loss, are important metrics which are shown in **Fig 4**. Loss notifies us how closely the model's forecasts match the actual data labels. Training loss indicates how well the model matches the data it was trained on, while validation loss measures its performance on new, unseen data, meaning how well it generalizes. The proposed ensemble model training testing and validation loss accordingly 0.0336, 0.0.807, and 0.5134, respectively.

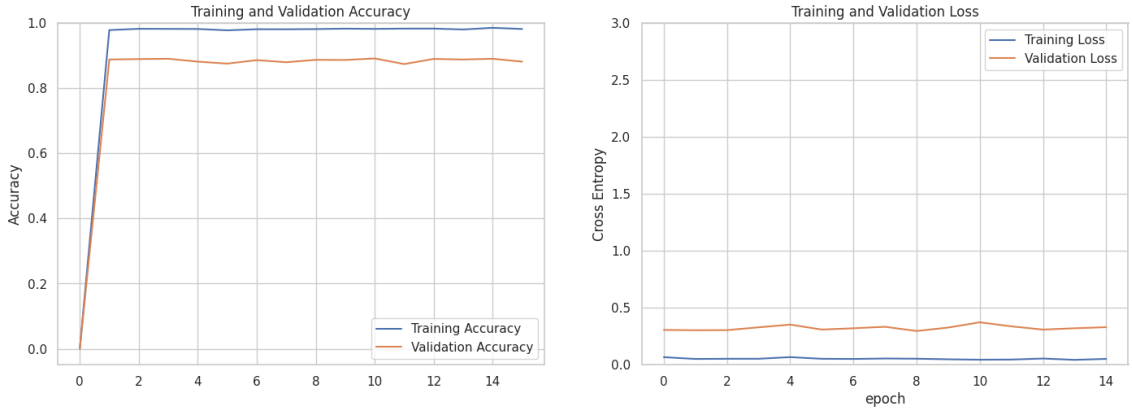


Fig 4. Training And Validation Loss Of The Proposed Model.

Furthermore, the model is graphically evaluated. We visualize the ROC curve and confusion matrix. The confusion matrix is shown in **Fig 5** and in **Fig 6**, we displayed the ROC curve of our proposed model. The figure helped us to analyze the error. The figure shows that the proposed model did not make any error in classifying the class melanoma; on the other hand, among 250 benign instances, the model incorrectly predicted 10 classes as melanoma. Furthermore, the model can accurately identify all melanoma instances. Additionally, we have calculated the Area Under the(AUC) value under the ROC Curve. The ROC curve demonstrated how well the model performed at different threshold values in differentiating between the true positive rate (Sensitivity) and the false positive rate (1-Specificity).

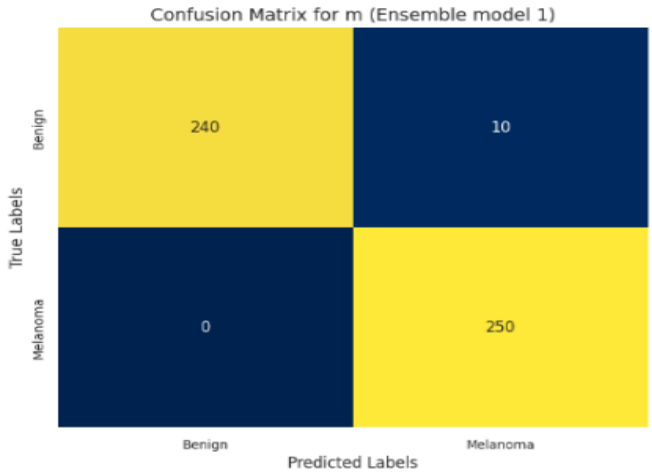


Fig 5. Confusion Matrix of The Proposed Ensemble Model.

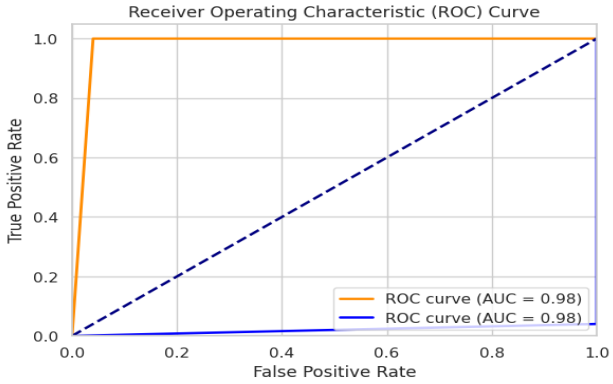


Fig 6. ROC Curve of The Proposed Model.

When the AUC is 0.98, it indicates that the model has a near-perfect capability to distinguish between the two classes. Put differently, the false positive rate is extremely low, but the genuine positive rate is very high.

In practical terms, an AUC of 0.98 suggested that the model is remarkably accurate in its predictions. For instance, if the model is being utilized for skin cancer classification, it means that the model can accurately identify the extensive majority of cancerous lesions while also maintaining misclassifications of non-cancerous lesions to a minimum.

An AUC of 0.98 in the ROC curve represents a model with exceptional performance, presenting reliable and accurate predictions with minimal errors in classification. The minimal error is also observed in the confusion matrix.

Discussion And Comparison of The Proposed Method with Prior Research

Table 3. Compare the Proposed Model With Current State-Of-Art

Reference	Dataset	Model	Performance
[20]	ISIC	InSiNet	Accuracy -94.59%
[21]	ISIC	Transfer learning	AUC-96.81
Proposed Model	ISIC	Ensemble transfer learning	Accuracy -98%

We have compared our outcomes with recent research work that is presented in Table 3. Reis et al. [20], used a CNN-based deep learning method named InSiNet. They used several datasets in their research, and ISIC is one of them. They got 95.59% accuracy in their research on the ISIC 2018 dataset. However, our proposed model outperformed their model. Our proposed ensemble deep learning achieved 98% accuracy. The other's evaluations were also higher in our research compared to them. In another work [21], the author used the ISIC dataset. For classifying the skin cancer, they used a transfer learning model, namely EfficientNet.To evaluate their model, they have calculated the AUC value in their research. Their average AUC value was 96.81%, whereas the AUC value also evaluates our proposed model, and we got 100% AUC for each class. Therefore, the proposed model is surprised by their model.

V. CONCLUSION

Contemporary advancements in skin disease recognition, mainly in the realm of early skin cancer detection, have shown promising results with the novel deep-learning models. This work advances the rapidly expanding field by introducing a novel Transfer learning model with an Ensemble-Deep learning model designed for the classification of skin cancer. After a thorough analysis of the literature, it is clear that even though many models have demonstrated impressive accuracy rates, challenges, including imbalanced datasets, a lack of data, and poor image quality, continue to impede ideal performance. Our model aims to address these challenges by using balanced datasets, applying data augmentation techniques, and focusing on enhancing overall model robustness. The findings of this study provide worthy insights into the development of more influential skin disease recognition models, and our Transfer learning with an Ensemble-deep learning model proposes a promising solution to the challenges faced in the field, aiming to enhance early detection rates and eventually protect lives. Future research should highlight comprehensive data augmentation techniques, analyze the use of hyperspectral images for richer data analysis, and focus on enhancing model resilience and adaptability to diverse scenarios. Additionally, efforts should directed toward enhancing the generalizability of models across various demographics and skin types. The identification of skin diseases can progress by resolving these issues, offering healthcare professionals improved tools for proper and timely diagnosis, thus positively impacting patient care and outcomes.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

Reference

[1]. J. Amin et al., “Integrated design of deep features fusion for localization and classification of skin cancer,” Pattern Recognition Letters, vol. 131, pp. 63–70, Mar. 2020, doi: 10.1016/j.patrec.2019.11.042.

[2]. D. Bisla, A. Choromanska, R. S. Berman, J. A. Stein, and D. Polsky, “Towards Automated Melanoma Detection With Deep Learning: Data Purification and Augmentation,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2019, doi: 10.1109/cvprw.2019.00330.

[3]. H. M. Balaha and A. E.-S. Hassan, “Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm,” Neural Computing and Applications, vol. 35, no. 1, pp. 815–853, Sep. 2022, doi: 10.1007/s00521-022-07762-9.

[4]. V. L. Cohen et al., “Staging uveal melanoma with whole-body positron-emission tomography/computed tomography and abdominal ultrasound: Low incidence of metastatic disease, high incidence of second primary cancers,” Middle East African Journal of Ophthalmology, vol. 25, no. 2, p. 91, 2018, doi: 10.4103/meajo.meajo_96_18.

[5]. A. Dahou, A. O. Aseeri, A. Mabrouk, R. A. Ibrahim, M. A. Al-Betar, and M. A. Elaziz, "Optimal Skin Cancer Detection Model Using Transfer Learning and Dynamic-Opposite Hunger Games Search," Diagnostics, vol. 13, no. 9, p. 1579, Apr. 2023, doi: 10.3390/diagnostics13091579.
[6]. G. Frisso et al., "Functional Studies and In Silico Analyses to Evaluate Non-Coding Variants in Inherited Cardiomyopathies," International Journal of Molecular Sciences, vol. 17, no. 11, p. 1883, Nov. 2016, doi: 10.3390/ijms17111883.
[7]. T. H. Johansen et al., "Recent advances in hyperspectral imaging for melanoma detection," WIREs Computational Statistics, vol. 12, no. 1, Apr. 2019, doi: 10.1002/wics.1465.
[8]. A. Julian, B. N. Narendra, A. Chiranjeevi, and A. V. C. Reddy, "Prediction of Brain Tumor Classification by using CNN," 2023 International Conference on Computer Communication and Informatics (ICCCI), Jan. 2023, doi: 10.1109/iccci56745.2023.10128408.
[9]. M. A. Khan, M. Sharif, T. Akram, S. A. C. Bukhari, and R. S. Nayak, "Developed Newton-Raphson based deep features selection framework for skin lesion recognition," Pattern Recognition Letters, vol. 129, pp. 293–303, Jan. 2020, doi: 10.1016/j.patrec.2019.11.034.
[10]. T. Kränke, K. Tripolt-Droschl, L. Röd, R. Hofmann-Wellenhof, M. Koppitz, and M. Tripolt, "New AI-algorithms on smartphones to detect skin cancer in a clinical setting—A validation study," Plos One, Vol. 18, no. 2, p. e0280670, Feb. 2023, doi: 10.1371/journal.pone.0280670.
[11]. V. Vidya Lakshmi and J. S. Leena Jasmine, "A Hybrid Artificial Intelligence Model for Skin Cancer Diagnosis," Computer Systems Science and Engineering, vol. 37, no. 2, pp. 233–245, 2021, doi: 10.32604/csse.2021.015700.
[12]. S. O. Manoj, K. R. Abirami, A. Victor, and M. Arya, "Automatic Detection and Categorization of Skin Lesions for Early Diagnosis of Skin Cancer Using YOLO-v3 - DCNN Architecture," Image Analysis & Stereology, vol. 42, no. 2, pp. 101–117, Jul. 2023, doi: 10.5566/ias.2773.
[13]. T. Mazhar et al., "The Role of Machine Learning and Deep Learning Approaches for the Detection of Skin Cancer," Healthcare, vol. 11, no. 3, p. 415, Feb. 2023, doi: 10.3390/healthcare11030415.
[14]. A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin Lesion Classification Using Hybrid Deep Neural Networks," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, doi: 10.1109/icassp.2019.8683352.
[15]. P. Natha And R. Pothuraju, "Skin Cancer Detection Using Machine Learning Classification Models," Nov. 2023, doi: 10.20944/preprints202311.0248.v1.
[16]. N. Nigar, A. Wajid, S.Islam, and M.K.Shahzad, "Skin Cancer Classification: A Deep Learning Approach," Pakistan Journal of Science, vol. 75, no. 02, Jul. 2023, doi: 10.57041/pjs.v75i02.851.
[17]. R. O. Ogundokun et al., "Enhancing Skin Cancer Detection and Classification in Dermoscopic Images through Concatenated MobileNetV2 and Xception Models," Bioengineering, vol. 10, no. 8, p. 979, Aug. 2023, doi: 10.3390/bioengineering10080979.
[18]. L. Riaz et al., "A Comprehensive Joint Learning System to Detect Skin Cancer," IEEE Access, vol. 11, pp. 79434–79444, 2023, doi: 10.1109/access.2023.3297644.
[19]. Haldorai, B. L. R, S. Murugan, and M. Balakrishnan, "An Investigation on Different Approaches for Medical Imaging," EAI/Springer Innovations in Communication and Computing, pp. 57–75, 2024, doi: 10.1007/978-3-031-53972-5_3.
[20]. J. S M, M. P, C. Aravindan, and R. Appavu, "Classification of skin cancer from dermoscopic images using deep neural network architectures," Multimedia Tools and Applications, vol. 82, no. 10, pp. 15763–15778, Oct. 2022, doi: 10.1007/s11042-022-13847-3.
[21]. M. Zambrano-Román, J. R. Padilla-Gutiérrez, Y. Valle, J. F. Muñoz-Valle, and E. Valdés-Alvarado, "Non-Melanoma Skin Cancer: A Genetic Update and Future Perspectives," Cancers, vol. 14, no. 10, p. 2371, May 2022, doi: 10.3390/cancers14102371.