

# Skin Insight Fusion: Advancing Precision Diagnosis with Integrated Deep Learning Models and YOLOv8 Integration

Masuk Ur Rahman Shodoy,Raitah Binte Khusru, Amitabha Chakrabarty, Md.Fahim-Ul-Islam,Ibne Hassan

*Department Computer Science and Engineering*

*Brac University*

Dhaka, Bangladesh

masuk.ur.rahman.shodoy@g.bracu.ac.bd, raitah.binte.khusru@g.bracu.ac.bd, amitabha@bracu.ac.bd, fahim.islam@bracu.ac.bd

**Abstract**—The rapid and precise identification and treatment of skin issues is critical in the field of global public health. Deep Learning (DL) has emerged as a critical method for improving medical picture analysis, particularly in the classification of skin disorders. Our research aims to improve DL models by introducing an ensemble model that combines the prediction powers of three different DL models. Furthermore, we broaden our analysis by including the cutting-edge You Only Look Once Version 8 (YOLOv8) to give an in-depth comparison of detection and classification models in the context of skin-related datasets. The ensemble methodology we propose uses feature-level fusion to create a prediction model that easily merges disparate data sources for long-term consistency. Our experimental results show that the suggested ensemble models outperform their counterparts, with an amazing accuracy of 94%. Furthermore, we use a variety of evaluation measures to test the performance of our ensemble models in the classification of skin diseases. This study advances DL applications in the vital domain of dermatological health by stressing the potential for improved diagnostic accuracy and treatment outcomes.

**Index Terms**—CNNs, Xception, Psoriasis, NASNetMobile, Melanoma, Inception V3, Skin disease, ResNet-50, Eczema, EfficientNet, Acne, DenseNet201, MobileNet, Solar lentigo, Dermatofibroma, Scabies.

## I. INTRODUCTION

Skin disorders claim the lives of people all over the world. Best therapy and management require early and accurate diagnosis. Deep learning transformed medical image classification [1]. This study proposes an ensemble model comprising two models and three cutting-edge CNNs: ResNet-50, EfficientNet, and Inception V3 [2]. The study used MobileNet, NASNetMobile, DenseNet201, Xception. To simplify and improve skin disease classification. Many people worldwide have severe skin concerns. Early skin diagnosis is key to low-risk treatment. CNN-based deep learning aids skin disease diagnosis and medical picture processing. Automated skin condition classification of dermatological images. Get a dermatological picture dataset with skin disease annotations. Enhanced data [3]. Datasets should standardise image size, quality, and class imbalance. Look at CNN models like Xception NasNetMobile, MobileNet, and DenseNet201 for skin disease categorization [4]. Improve strengths with Inception V3, ResNet-50, and

EfficientNet. CNNs and ensembles utilise transfer learning and hyperparameter tuning. Validate models on several datasets to generalise. The project seeks a reliable skin condition identification tool. CNN models Xception, NasNetMobile, MobileNet, and DenseNet201 will be employed. Inception V3, ResNet-50, and EfficientNet will build a "ensemble" neural network. Deep learning diagnoses skin well. This study will construct a dermatological diagnosis computer system from clinical images and patient data. Labelling skin diseases with MobileNet V2 and LSTM. Good data collecting and analysis expand fields. The lengthy study examined deep learning skin disease categorization. The study covered skin lesion imaging basics and technologies. Accurate dataset-based research requires processing, categorization, and evaluation. This study creates a deep learning clinical skin diagnostic framework [5]. Deep learning improves skin disease detection. This industry has many growing challenges. Deep learning, better dataset properties, publicly acceptable skin disease photo datasets, and public dataset racial and regional biases are needed. Our main contributions are :

- We propose on the development process of an ensemble model for our dataset with seven classes, emphasizing individual model fine-tuning, overfitting prevention, and model diversity encouragement via weighted averaging strategies, resulting in a robust ensemble model tailored to the dataset's characteristics.
- Our ensemble model's distinct architecture design outperforms other DL approaches based on ROC curves, classification reports, accuracy curves, and confusion matrices derived from extensive testing.
- Furthermore, we compare object recognition and classification models using the YOLOv8 on our skin disease image dataset to improve result comprehension and analysis.

## II. LITERATURE REVIEW

The goal of this section is to discuss recent research on skin diseases using DL-based methods.

This study uses CNNs and DL on dermatology pictures. This study investigates six CNN models for three Middle

Eastern dermatological diseases. Dermatology and deep learning reduce other difficulties by focusing on skin cancer. This study used data from DermNet, the University of Iowa Dermatology Department, and other credible sources. Misrepresented database depth. A study finds dermatological databases inadequate. Patients must stay private. Additionally, patient privacy is ignored [13]. Automated hybrid deep learning predicts skin diseases. BBO and DCNN automatically classify skin problems. This research aims to improve skin disease forecast accuracy using the Biogeography-Based Optimisation Algorithm (BBOA) and Deep Convolutional Neural Network (DCNN). It reduces computational and human resources, improves diagnosis, and saves time. The model's contextual data emphasis ensures accurate prognoses, especially with earlier timestamp data [12].

FC7, softmax, and cross-entropy classify skewed multiple skin lesions (MSLC) in fourteen deep learning networks without the last fully connected layer using transfer learning. Dialogue improves. Study used 10,15 dermoscopic images. Resized, improved photographs. Iterations, batch size, and learning rate assess networks. Use 14 deep learning networks to classify skin lesions. Performance indicators for network analysis include accuracy, recall, precision, and F1. Expand and standardise transfer learning. Network analysis compares various factors. Skin lesions are classified by DNN borders or data [11].

This study accurately detects skin diseases utilising optimal region growth segmentation and autoencoder classification. Grey Wolf Optimisation (GWO) and optimal region growth segment sick regions best. Both WL and GLCM examine segmented lesion textures. Latent representations and autoencoders minimise features. CNN and neural networks discover pathogenic lesions using autoencoder latent representation. The study showed the suggested method detects skin abnormalities better than deep categorization. Grey Wolf segmental lesions improved. Quantity and variety of classification model training data effect autoencoder-based framework performance [10].

SpaSA deep transfer learning improves skin cancer detection and categorization, with five unique U-Net models and eight pre-trained CNN models optimizing research hyperparameters. The method employs U-Net DenseNet201 and utilizes public databases for education and assessment. Understudied are performance metrics and method impacts. The technique's efficacy is uncertain without field best practises or technical advancements. The work eliminates deep transfer learning and SpaSA optimizer skin cancer diagnosis disadvantages [9].

Researchers trained deep neural networks for skin cancer diagnosis using dermoscopic images to reduce human interpretation errors. They aimed for early disease detection and quality service, using transfer learning and fine-tuning for faster model construction. Progressive learning reduces EfficientNet overfit. Activities go beyond research, this study accurately detects skin problems using DNNs and disease taxonomy [8]. ISIC Archive and DermNet train deep neural

networks. DermNet and ISIC Archive skin illnesses were detected using strong DNNs. Disease taxonomy improves modeling. We found advanced DermNet disease classification. The proposed approach has 98% AUC and 80% accuracy.

The 622-subclass DermNet dataset had 67% accuracy and 98% AUC. The ISIC Archive found all seven diseases 99-93%. Deep learning may classify skin issues like humans, according to research. This approach offers great potential for real-time skin condition detection. DNNs and CAD users classify skin conditions well. This revolutionary dermatological disorder classification leverages disease taxonomy for accuracy and AUC. Clinical or dermoscopy photographs can assist clinicians diagnose skin problems using deep learning. This study found unhomogeneous classification unhelpful. The method may alter skin disease detection. Different proprietary and public dataset train-test splits and class numbers cause this. Public database photos can be watermarked, loud, or small. Clinical interferences can alter fine-grained deep learning AI classifier item categorization data [7].

Performance excellence is measured by sensitivity, specificity, and AUC. Deep learning and machine learning increase dermoscopy melanoma diagnosis. Breakthroughs change industry standards. ROC curve, average accuracy, and specificity rose. On identical test photos, it beats dermatologists. Report lists two downsides. Performance comparisons suffer from predefined dataset partitions and no-n-fold assessment tools. Skin lesion characteristics influence dermatologists' diagnoses [6].

### III. DATA SET

We collected our dataset from the online source kaggle 'skin disease image' dataset where it consists of seven different skin diseases. Our dataset includes acne, melanoma, solar lentigo, dermatofibroma, eczema, psoriasis, and scabies [14]. We grew our limited data collection using photo augmentation methods like rotation, shearing, and scaling. Clahe converted the photographs to black and white from the augmented data. Data augmentation efficiently improves machine learning models' robustness and real-world accuracy. Model improvement requires data augmentation, especially when training data is scarce. CLAHE and histogram equalization are utilized in our study. In areas with high intensity variability, CLAHE Histogram Equalisation avoids over-enhancement. It tiles the image and applies Histogram Equalisation separately. It restricts tile augmentation by clipping. This balances noise augmentation and prevents uniform zone amplification.

### IV. METHODOLOGY

#### A. Workflow

The initial stage is gathering skin disease data. Image division follows data preprocessing. Additionally, we calculate tile Histograms. Additionally, we calculate CDF. Adjusting tile intensity, sewing, and clipping following. Additionally, we split the data into test and train. Data split 80/20. Models are trained on 80% and tested on 20%. Afterwards, all models are

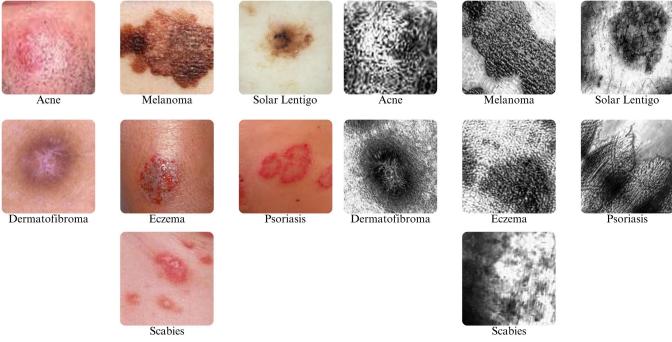


Fig. 1. Data Set Before and After CLAHE

performed, and the results are measured on different evaluation metrics. The whole process is displayed in Figure 2.

#### B. DenseNet201

Here, CNN-201 divides thicker slabs into transitional layers by using deep convolutional refers to DenseNet layers model for thick layer networking in DenseNet model. Transitional layers minimize feature mappings and spatial dimensions, resulting in global average pooling and a Softmax-activated layer that is fully linked [15].

#### C. Xception

Google researchers developed Xception, a deep convolutional network with 71 layers, using residual connections to separate channel and spatial correlations [16]. The architecture prevents gradient disappearance and smooths propagation, outperforming ResNet and Inception V3.

#### D. Ensemble Model-X1 architecture

For ensemble learning, three deep learning models were used: ResNet50, EfficientNetB0, and InceptionV3. Each model was fine-tuned using a dataset with seven classes. Ten layers were frozen during the fine-tuning process, up to the GlobalAveragePooling2D layer. The top layer was removed from each model, and a Dense layer was added. Figure 3 depicts the architectural model. Ensemble-averaging models average learned models, calculating weighted, total, and basic averages. Weighted sum ensembles are the best performance booster, as poor execution may worsen.

In the model development process outlined, each individual model undergoes fine-tuning prior to ensemble creation. The dataset, containing seven distinct classes, guides the architecture adjustments. For every solo model, a “GlobalAveragePooling2D” layer freezes ten preceding layers, the top layer is excluded, and a dense layer is appended. This setup serves as the foundation for following ensemble approaches. To achieve this, the top layers are frozen, and global average pooling is applied. The resulting ensemble model, denoted as  $E$ , combines the predictions of individual models ( $M_i$ ) using weighted averaging. Each  $M_i$  represents a distinct model, and

$W_i$  denotes the weight assigned to the corresponding model based on its quality and contribution.

$$E = W_1 \cdot M_1 + W_2 \cdot M_2 + \dots + W_n \cdot M_n \quad (1)$$

The prevention of overfitting and the encouragement of model diversity are critical considerations. This is accomplished by designing each model with its own architecture and applying transfer learning. The weighted sum ensembles are preferred over basic averages to avoid equal contribution, with the weights assigned based on the quality estimates ( $Q_i$ ) of the individual models. The multi-model quality estimation weighting ( $Q_{\text{ensemble}}$ ) further refines the ensemble’s predictive power by incorporating quality estimates into the final decision-making process.

$$\text{Weighted Sum}(E) = \frac{W_1 \cdot Q_1 + W_2 \cdot Q_2 + \dots + W_n \cdot Q_n}{W_1 + W_2 + \dots + W_n} \quad (2)$$

In summary, the architecture and ensemble procedures outlined in this process aim to improve model performance by combining varied, fine-tuned individual models with weighted averaging strategies that prefer higher-performing models. This method ensures a resilient and effective ensemble model that is tuned to the specifics of the provided dataset.

#### E. YoloV8

One of the newest and most advanced yolo models that is highly respected for object identification or picture segmentation applications is the YOLOv8 model. This model, created by Ultralytics, includes 25.9 million parameters and is pre-trained using the COCO dataset. The yolov8 version that we use is called yolov8m. In addition to image classification’s advantage in terms of resource efficiency, performance is another crucial consideration. A model or method cannot be deemed superior to another without demonstrating both strong performance and efficiency. Thus, we apply the yolov8 model to our dataset to conduct object recognition and track the outcomes it yields.

## V. RESULTS AND DISCUSSION

#### A. Results

1) *Accuracy:* DenseNet20 train accuracy starts at 0.67. Train accuracy is 0.95 at 15 epochs, nearly perfect. Training accuracy is 0.99 after 40 epochs. Validation accuracy evolves from 0.5. Epoch 35: 0.94. Validation accuracy is 0.87 after 40 epochs. Extreme Inception (Xception) diagnosed skin issues 86% accurately after jogging. Set 0.69 train accuracy. Epoch 15 train accuracy is 0.95, nearly perfect. After 40 epochs, training accuracy approaches 0.97. Validation accuracy varies from 0.67. Epoch 31 peaked at 0.96. Validation accuracy averages 0.86 after 40 epochs. Ensemble Model-X1 initial train accuracy is 0.97. Train accuracy is 1.00 at 5 epoch. Training accuracy is 0.99 after 40 epochs. Validation accuracy improves from 0.70. Epoch 38 maxes at 0.96. Validation accuracy is 0.94 after 40 epochs. The Figure 4 displays all models accuracy curves.

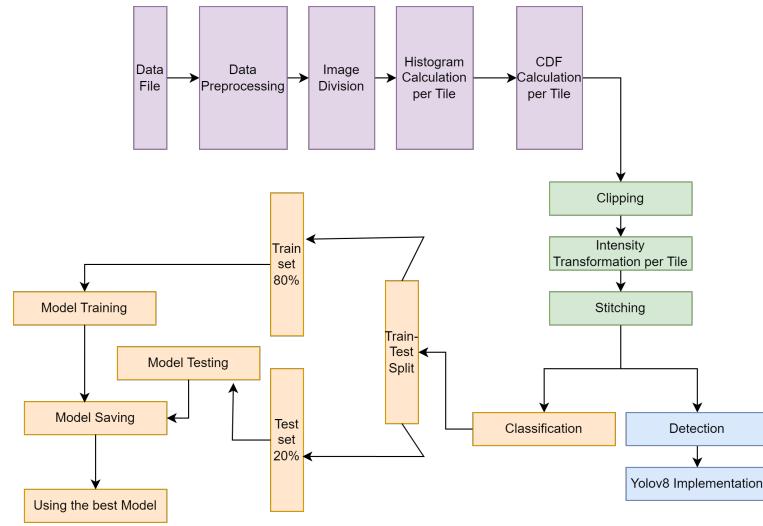


Fig. 2. Work Flow of the research

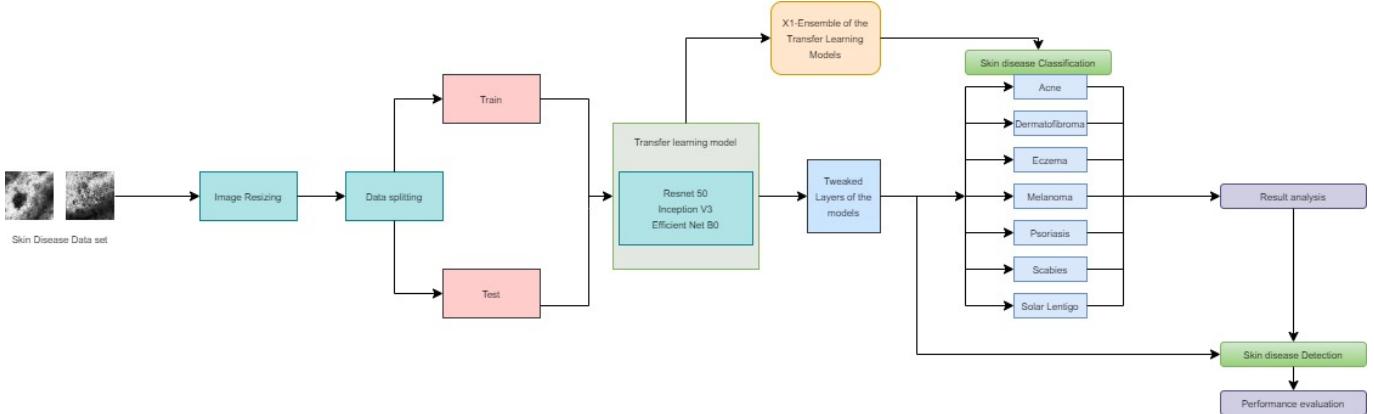


Fig. 3. Ensemble Model-X1 Architecture

2) *Confusion Matrix*: The Figure 5 shows extensive confusion matrices for several skin disease classification models, including Densenet201, individual models for acne, dermatofibroma, eczema, melanoma, psoriasis, scabies, and solar lentigo, and Ensemble Model-X1. The matrix shows the number of correct and incorrect identifications for each illness category, illustrating the strengths and drawbacks of each model in detecting skin disorders. Overall, the models vary in their accuracy and misidentification across different skin disease groups.

3) *ROC Curve*: AUC values from 0.97 show strong ROC curve discrimination for acne, dermatofibroma, eczema, melanoma, psoriasis, scabies, and solar lentigo. With AUC values near 1.00, this graph shows that the classification model accurately identifies skin conditions from non-conditions. Psoriasis and dermatofibroma have lower AUCs but differ. Dermatology benefits from the model's skin diagnosis. Xception's ROC curves exhibit strong AUC values for acne, Dermatofibroma, eczema, melanoma, psoriasis, scabies, and solar lent. Our ensemble X1's ROC curves for acne, dermatofibroma,

eczema, melanoma, psoriasis, scabies, and solar lentigo have 0.99 AUCs showing better performance. Multiple curves suggest model skin disease detection. With AUCs of 1.00, acne, eczema, melanoma, psoriasis, scabies, and solar lentigo differentiate perfectly, while the others have 0.99 as shown in Figure 6.

4) *Yolov8 Experimentation*:: Firstly, as we are analyzing the completion of detection, it is more suitable to detect proper objects enclosed by a frame, like skin-marked places, for each of the classe datasets. So, we select the images of the seven categories that are acne, Dermatofibroma, eczema, melanoma, psoriasis, scabies, and solar lentigo. All these Categories are divided into training, validation, and testing. The YOLOV8 model executes detection on the training images and testing images by using bounding boxes that recognize the desired labeling. The Figure 7 shows the detection outcomes achieved from four testing images. From the figure the Melanoma, Eczema, and Dermatofibroma show almost 99% correct detection but psoriasis shows only 73% correct detection; moreover, there are more marks of psoriasis that the model is unable to

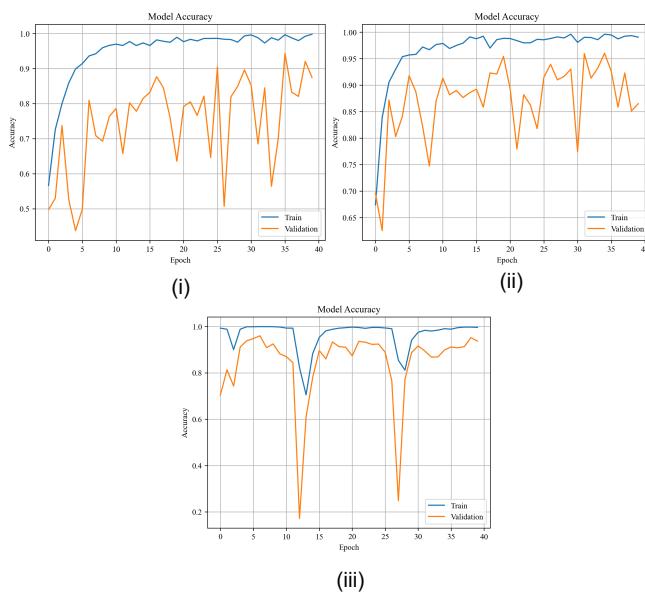


Fig. 4. (i)Densenet201 Accuracy curve (ii) Xception Accuracy curve (iii) Ensemble Model-X1 Accuracy curve

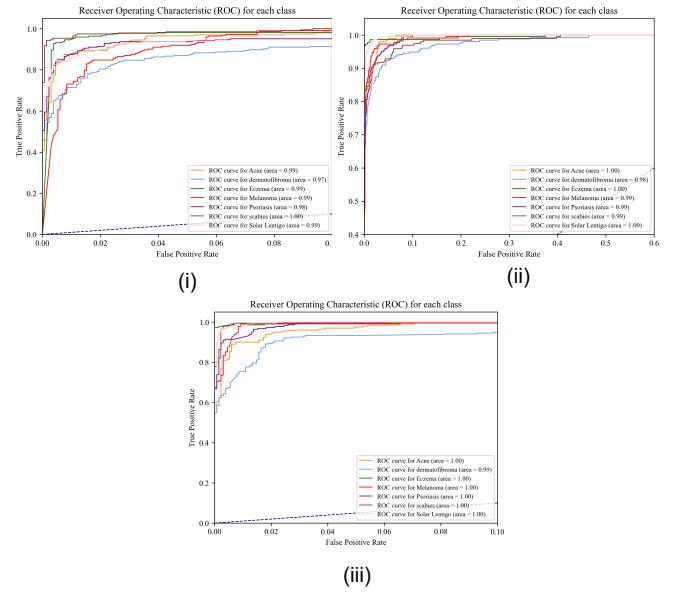


Fig. 6. (i)Densenet201 ROC curve (ii) Xception ROC curve (iii) Ensemble Model-X1 ROC curve

detect. Not only that, for some testing images of Acne and solar Lentigo it is completely unable to detect any other specific marks. From this study, it is pointed out that the model needs additional training data and higher parameters to come up with better performance.

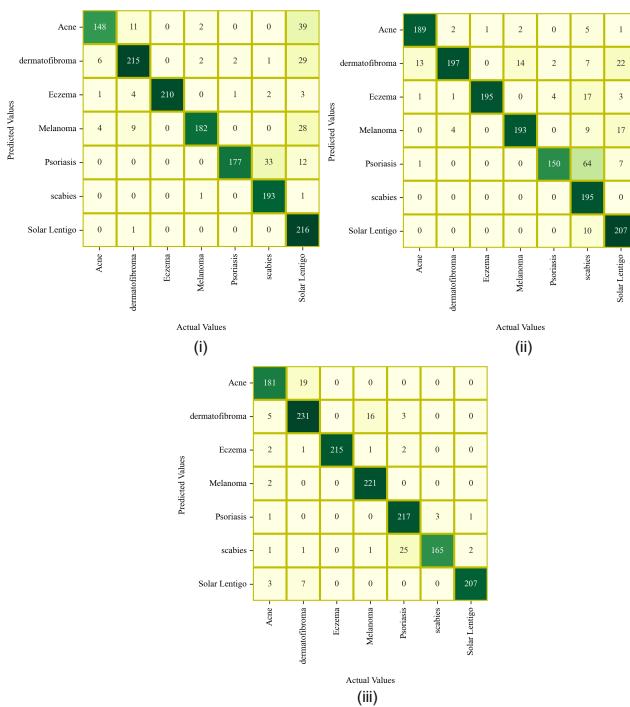


Fig. 5. (i)Densenet201 Confusion Matrix (ii) Xception Confusion Matrix (iii) Ensemble Model-X1 Confusion Matrix

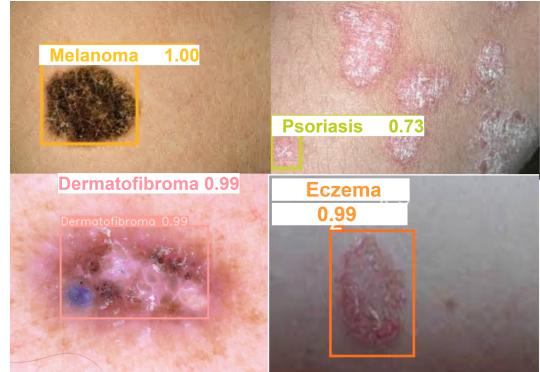


Fig. 7. results of object detection using YOLOv8 implementation

**TABLE I**  
PRELIMINARY COMPUTATIONAL RESULTS OF THE ML ALGORITHMS USED  
DURING THIS STUDIES

| Model             | Precision | Recall | F1Score | Accuracy | AUC   |
|-------------------|-----------|--------|---------|----------|-------|
| Densenet201       | 0.90      | 0.87   | 0.88    | 0.87     | 0.987 |
| Xception          | 0.89      | 0.86   | 0.87    | 0.87     | 0.993 |
| Ensemble Model-X1 | 0.94      | 0.94   | 0.94    | 0.94     | 0.94  |

#### B. Contrast of CNN classification models and yolov8 detection model

Among the implementation of our deep learning models, our custom Ensemble Model-X1 outperforms rest of the models

with almost 94% that focuses on classification, for detection, we implement the YoloV8m model that has a lower number of correct detection and huge parameters of 25.9 million that is much bigger than any other CNN classification model's parameters. Thus, it is wiser to decide on any classification model to select instead of any detection model in this scenario. Our Ensemble Model-XI performance outperforms other DL models as shown in Table I.

## VI. CONCLUSION AND FUTURE WORKS

In terms of conclusion, our research suggests an ensemble model fitted to a seven-class skin disease dataset that achieves improved performance by rigorous fine-tuning and diversity promotion. A comparison with other DL techniques demonstrates its effectiveness. In addition, we use YOLOv8 for object recognition, which improves the comprehension of our data and advances our understanding of dermatological image analysis. In the field of skin health, our contributions promise enhanced diagnosis accuracy and treatment outcomes. In the future, prospective improvements include improving the model for computational efficiency, with a focus on deployment on low-power devices such as smartphones. The creation of a model app for skin disease detection opens the door to discovering symptoms, causes, therapies, and other information. Future updates may include expanding the model to include all global skin conditions, as well as making many model tweaks to improve performance. The addition of varied dermatological imaging datasets to the training data repository is encouraged to further enhance the model's capacity. This augmentation adds diversity and quantity to the model, allowing it to detect unique skin disorders and perhaps discover rare diseases. These factors, taken together, pave the door for a more robust and adaptable use of deep learning in dermatological diagnostics.

## REFERENCES

- [1] Poornima, G., Sakkari, D. S. Diagnosis of skin diseases based on deep learning and machine learning approach: Technical review.
- [2] Jeyakumar, J. P., Jude, A., Priya Henry, A. G., Hemanth, J. (2022). Comparative Analysis of Melanoma Classification Using Deep Learning Techniques on Dermoscopy Images. *Electronics*, 11(18), 2918.
- [3] Salamaa, W. M., Aly, M. H. (2021). Deep learning design for benign and malignant classification of skin lesions: a new approach. *Multimedia Tools and Applications*, 80(17), 26795-26811
- [4] Arora, G., Dubey, A. K., Jaffery, Z. A., Rocha, A. (2023). A comparative study of fourteen deep learning networks for multi skin lesion classification (MSLC) on unbalanced data. *Neural Computing and Applications*, 35(11), 7989-8015.
- [5] Tabrizchi, H., Parvizpour, S., Razmara, J. (2023). An improved VGG model for skin cancer detection. *Neural Processing Letters*, 55(4), 3715-3732. [Magnetics Japan, p. 301, 1982].
- [6] Codella, N. C., Nguyen, Q. B., Pankanti, S., Gutman, D. A., Helba, B., Halpern, A. C., Smith, J. R. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, 61(4/5), 5-1.
- [7] Bajwa, M. N., Muta, K., Malik, M. I., Siddiqui, S. A., Braun, S. A., Homey, B., ... Ahmed, S. (2020). Computer-aided diagnosis of skin diseases using deep neural networks. *Applied Sciences*, 10(7), 2488.
- [8] Venugopal, V., Raj, N. I., Nath, M. K., Stephen, N. (2023). A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images. *Decision Analytics Journal*, 100278.
- [9] Balaha, H. M., Hassan, A. E. S. (2023). Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. *Neural Computing and Applications*, 35(1), 815-853.
- [10] Sun, J., Yao, K., Huang, G., Zhang, C., Leach, M., Huang, K., Yang, X. (2023). Machine Learning Methods in Skin Disease Recognition: A Systematic Review. *Processes*, 11(4), 1003.
- [11] Zhu, C. Y., Wang, Y. K., Chen, H. P., Gao, K. L., Shu, C., Wang, J. C., ... Liu, J. (2021). A deep learning based framework for diagnosing multiple skin diseases in a clinical environment. *Frontiers in medicine*, 8, 626369.
- [12] Zhang, B., Zhou, X., Luo, Y., Zhang, H., Yang, H., Ma, J., Ma, L. (2021). Opportunities and challenges: Classification of skin disease based on deep learning. *Chinese Journal of Mechanical Engineering*, 34(1), 1-14.
- [13] Muhaba, K. A., Dese, K., Aga, T. M., Zewdu, F. T., Simegn, G. L. (2022). Automatic skin disease diagnosis using deep learning from clinical image and patient information. *Skin Health and Disease*, 2(1), e81.
- [14] Sanskar, "skin disease image", Kaggle Dataset 2019.
- [15] Sanghvi, H. A., Patel, R. H., Agarwal, A., Gupta, S., Sawhney, V., Pandya, A. S. (2023). A deep learning approach for classification of COVID and pneumonia using DenseNet-201. *International Journal of Imaging Systems and Technology*, 33(1), 18-38.
- [16] Gülmek, B. (2023). A novel deep neural network model based Xception and genetic algorithm for detection of COVID-19 from X-ray images. *Annals of Operations Research*, 328(1), 617-641.