Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 245 (2024) 1129–1138

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Computer Science and Computational Intelligence 2024 (ICCSCI 2024)

A comparative study of deep learning algorithms for image-based classification of hyperpigmented skin disease

I Gusti Bagus Ramadha Saverian Ranuh[a*], Marvellino Christian Sanjoto[a], Alfi Yusrotis Zakiyyah[b], Meiliana[a]

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia*
[b]*Mathematics Department, School of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia*

Abstract

There are growing numbers of significant skin disorders, including skin pigmentation. It states that skin color is determined by the amount of melanin produced by the body. The two main categories of skin pigmentation are hyperpigmentation, in which pigment seems to overflow, and hypopigmentation, in which pigment appears to decrease. However, many skin conditions share characteristics, making it difficult for dermatologists to correctly early diagnose their patients. Consequently, the accurate detection of skin disorders and the diagnosis of dermatoscopy pictures can be greatly aided by machine learning and deep learning approaches. The most effective deep learning technique for picture identification was investigated to diagnose hyperpigmented skin diseases. YOLO, DenseNet201, GoogLeNet, InceptionResNetV2, and MobileNet were among the pretrained models used to classify four most common hyperpigmented skin disorders. Using assessment metrics like accuracy and AUC (Area Under Curve), it was determined which deep learning method would work best for creating a clinical diagnostic system. The study analyzed the accuracy rates of five pretrained models, including GoogleNet, MobileNet, DenseNet201, InceptionResNetV2, and YOLO, after 50 iterations. Using metrics such as accuracy and Area Under the Curve (AUC), the study evaluated the models' performance on a small dataset split into 80% for training and 20% for testing. Training accuracy rates were 93.8%, 100%, 100%, 98.77%, and 97.43%, respectively, while test accuracy rates were 87.18%, 79.49%, 87.18%, 89.74%, and 97.56%. DenseNet201 showed strong performance, particularly for cafe-au-lait spots (CS), melasma (ML), and nevi (MN), but struggled with congenital nevus (CN). Despite DenseNet201's strong traditional CNN capabilities and generalization, YOLO emerged as the top model due to its stable accuracy and AUC values, as confirmed by confusion matrices. While these models show promise as diagnostic tools for dermatologists, further research, including expanding the dataset and exploring hybrid models, is needed to enhance their clinical accuracy and effectiveness.

* Corresponding author.
E-mail address: i.ranuh001@binus.ac.id

1877-0509 © 2024 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 9th International Conference on Computer Science and Computational Intelligence 2024
10.1016/j.procs.2024.10.342

Keywords: skin disorders; pigmentation; diagnosis; machine learning; deep learning; Convolutional Neural Network; Area Under Curves; Confusion Matrix

1. Introduction

Skin is the biggest organ in humans and the body's outermost layer. While the skin shields the body from UV radiation, sunlight exposure also causes the skin to create vitamin D, which is vital for human health [1]. Stated that there are several difficulties in identifying skin illness, such as the difference on the skin tone among individuals and the presence of hair and debris [2] In addition to this, many characteristics of skin illness are identical, especially when they are first discovered. When melanoma cells proliferate throughout the body, treating the disease becomes difficult [1]. For this reason, early diagnosis is essential to save lives. Dermoscopy pictures show that there are many kinds of skin diseases. Melanocytic and nonmelanocytic skin diseases can be distinguished as the two primary categories. Two forms of melanocytic diseases include nevi and melanoma [2]. Which is the focus on this paper.

Various picture and medical record types, including dermatoscopy and medical records, have been classified using artificial intelligence approaches. Recent developments in deep learning have made it easier to create artificial intelligence programs that can diagnose skin conditions from pictures [3]. With pre-processing techniques, skin illness identification, and separation from healthy skin to concentrate on the affected region, deep learning is expected to have a capacity to complete evaluations quickly and accurately [2]. It is expected to have an impact on the functions that image specialists play in biological diagnostics.

Through this research, we aim to address the critical challenge of early and accurate skin disease diagnosis, contributing to improved patient outcomes and advancing the field of medical image analysis. Our contribution seeks to provide a comparative analysis on the effectiveness of certain models with respect to these issues. Researchers seek to identify the most successful method for categorizing hyperpigmented skin conditions by evaluating the algorithms' prior performance. Our study evaluates five deep learning models, selected based on their demonstrated high performance and accuracy across a range of image classification and object detection tasks. These models are recognized for their reliability and effectiveness in medical image analysis, which establish their selection for this research.

The following paragraphs outline the structure of this text and its components. Chapter 2, "Material and Related Works," provided a thorough overview of all the earlier research. Chapter 3, this part included information on the experimental setup and dataset utilized in the study. The experimental result and conclusion of the experiment are discussed in Section 4 (Results and Discussion). Finally, in the conclusion section, both the conclusion reached about the system and the work that will be done in the future to improve it were discussed.

2. Material and Related Work

2.1. Deep Learning for Image Classification

Image categorization has undergone a revolution because of deep learning, especially in the realm of medical diagnosis. Because Convolutional Neural Networks (CNNs) have been the backbone of many successful models due to their ability to automatically learn features from images without the need for manual feature extraction. Since little variations in picture patterns can reveal various disorders, this feature makes them especially well-suited for medical imaging [[2].

Once the software is developed, pattern identification of photos may be done automatically using the deep learning approach. High fidelity images may be fed into CNN (Convolutional Neural Network), and significant characteristics can be automatically extracted. As a result, this approach does not require information extraction from pictures before the learning process. Simple aspects in the pictures, like their edges, are learnt at shallow levels. While more intricate high-order characteristics are learnt at deep levels close to the output layer [[2].

The application of Deep Learning in image classification includes using algorithms to analyze the data in the image and identify the disease it has. Several approaches have been examined, such as GoogleNet Architecture, MobileNet, DenseNet201, InceptionResNetV2, and DCNN(YOLO). These algorithms were developed using a sizable dataset that iencludes skin hyperpigmentation.

2.2. Hyperpigmented Skin Disease

Many factors can cause skin pigmentation, which is a common disorder. Sun exposure, certain drugs, and heredity are the main three factors that cause skin pigmentation. Our ability to treat and prevent skin pigmentation will improve if we comprehend its underlying causes [1].

Sun exposure is the most prevalent cause of hyperpigmentation since it dramatically stimulates the synthesis of melanin [1]. The skin will appear lighter than the surrounding skin after an injury heals. Certain skin areas may develop hyperpigmentation because of other hereditary causes [2].

2.3. Related Works

Image classification of hyperpigmented skin disease has drawn a lot of interest, with researchers looking at advanced deep learning techniques. Deep learning algorithms, such as convolutional neural network (CNN), have drawn attention to image recognition, because it seems promising in dealing with image pattern in recognizing something.

Due to their visual similarities, hyperpigmentation illnesses such malignant melanoma, congenital nevi, melasma, and café-au-lait spots provide special hurdles for categorization. For this goal, many deep learning architectures have been used in previous research. To achieve high accuracy in classifying various forms of skin lesions, Tschandl et al. [3] used deep learning models; however, they also noted the necessity for bigger and more diversified datasets.

Jianpeng Zhang et al. [4] proposed a cooperative learning method for medical image classification, significantly improving accuracy but limited to certain types of skin diseases. Yuan Liu et al. [[5] developed a neural network approach to identify 26 skin conditions, providing supplementary diagnosis for physicians. While these studies have advanced the field, they frequently run into issues like the quantity of the dataset and the requirement for extensive computational resources.

While significant progress has been made in using deep learning for skin disease classification, challenges remain, particularly in the classification of hyperpigmented skin diseases. By offering a comparative examination of various models particularly for the categorization of hyperpigmented skin diseases, this study seeks to close these gaps. This research aims to determine which model is the most successful for use in clinical diagnostics by assessing the performance of different models using measures like accuracy and AUC.

3. Methods

3.1. Dataset

Images of four prevalent hyperpigmented skin conditions from Kaggle [6] and GitHub [7]. The features of the lesions caused by each skin ailment are listed below: Brown or dark brown lesions are symmetrical, unevenly dispersed, or butterfly-shaped in ML(Melasma). They also appear in the zygomatic and temporal areas. CS (cafe-au-lait spots) are distinct, light brown, evenly pigmented macules that can be solitary or many. CN (Congenital Nevus) is a skin lesion that appears at birth or develops in the first few weeks and is caused by benign proliferations of nevomelanocytes. Giant hairy nevi are another name that may be used to describe these lesions, and it indicates that excessive hair development is often present clinically. A mole's change or the appearance of a new spot are usually the first signs of MM (Malignant Melanoma). It may have varying color tones or become blotchy, enlarged size, irregular shape, and an elevated spot.

When working with a small dataset, especially in the context of deep learning for medical image analysis, several strategies can be employed to mitigate the challenges and improve model performance. We utilize transfer learning to leverage the knowledge acquired by pre-trained models on large datasets, such as ImageNet. This approach can significantly enhance performance when applied to small datasets. Additionally, we employ data augmentation techniques such as rotation, flipping, scaling, color jittering, and adding noises to artificially expand the size and diversity of the training dataset, which helps in preventing overfitting. These strategies collectively improve the model's ability to generalize from limited data and achieve better classification accuracy.

The dataset is divided into a training set (80%) and a testing set (20%), selected randomly. This split ensures enough data for the model to learn from while retaining a separate subset for evaluating model performance. The 80-20 split is a standard practice in machine learning, balancing the need for ample training data with the necessity of having an unbiased test set to assess model generalization. The model then attempts to identify which illness is which by learning from comprehending the visual of each one using the training set. While the testing set was used to determine how well the model preformed.

Table 1. Dataset Distribution

Model	Training Data Sum	Test Data Sum	Total
Melasma (ML)	28	7	35
Cafe-au-lait Spots (CS)	94	24	118
Congenital Nevus (CN)	40	10	50
Malignant Melanoma (MM)	102	26	128
Total	264	67	331

Given the imbalanced dataset distribution observed in Table 1, our evaluation focuses on metrics that provide deeper insights beyond accuracy. We use assessment methods like the confusion matrix, which offers a thorough analysis of the predictions produced by a classification model and displays the quantity of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) in each class. Area Under the ROC Curve are also known to offer more nuanced assessments of model performance with imbalanced datasets [8].

3.2. Convolutional Neural Network

Convolutional Neural Network is a type of feedforward neural network capable of extracting features from data that has convolutional structures. CNN does not need manual feature extraction, in contrast to conventional feature extraction techniques. CNN's architecture draws inspiration from visual perception. A CNN kernel is a variety of sensors that may respond to a wide range of stimuli; an artificial neuron is like a real neuron in that activation functions replicate the process by which neural electric signals that cross a particular threshold are transmitted to the next neuron. Specifically, loss functions and optimizers are designed to train the CNN system to understand our expectations. CNN has several benefits over ordinary artificial neural networks. 1) Local relationships. 2) Weight sharing: This reduces parameters and speeds up convergence by connecting each neuron to a smaller set of neurons instead of all the neurons in the preceding layer. Further reduction of parameters might occur when a set of links have the same weights. 3) Dimensionality decreases by down sampling. By using the idea of image local correlation, a pooling layer may down-sample a picture and save relevant information while doing so. Eliminating insignificant characteristics might also help to lower the total number of parameters. CNN has emerged as one of the most representative algorithms in the field of deep learning because to these three alluring features [9].

To create a CNN model, four components are often needed. Convolution is an important step in the feature extraction process. Feature maps are the result of a convolution. Setting a convolution kernel to a certain size results in information loss at the boundary. Padding is used to augment the input with a zero value, which indirectly modifies the size. In addition, stride is utilized to regulate the convolving density. The density decreases with increasing stride length. Following convolution, feature maps include many features that might lead to an overfitting issue. Therefore, pooling—also known as down-sampling—is suggested to eliminate repetition. This includes average and maximum pooling [10].

3.3. Deep Learning Algorithms

3.3.1. GoogleNet Architecture

GoogLeNet, project that the Google team created in 2014, and that same year, it won the ImageNet classification job competition [11]. The following are some of GoogLeNet's benefits. (1) To incorporate features from various sizes, an inception structure is created. (2) The model dimension is decreased using a 1 x 1 convolution kernel, which lowers the number of necessary parameters. (3) To aid in the sample training procedure, two additional classifiers are included. (4) To further minimize the number of needed parameters, fully linked layers are replaced by an average pooling layer [12].

The GoogLeNet model architecture uses 32 layers altogether and numerous Inception modules. To obtain high accuracy while lowering computational cost, utilizing the concept of Inception modules, GoogLeNet superimposes the average and maximum pooling layers over them. At various scales, it is capable for the model to extract features efficiently. [13] The primary disadvantage of GoogLeNet is its complexity, which makes it challenging to comprehend and use [14].

3.3.2. MobileNet

Google released MobileNet [15], an open-source CNN class designed with mobile devices in mind, as a starting point for classifier training using a lightweight model. Convolutional neural networks, or CNNs, are made for mobile and embedded devices. There are several variations of the MobileNet model architecture, and most of them include dozens of layers [16]. Deep separable convolutions are used by MobileNet to minimize computational complexity while retaining excellent accuracy. Typically, the model consists of fully Convolutional layers, convolutional deep separable layers, and linked layers. When there are few computer resources available, the MobileNet design works well. In mobile apps, it is often utilized. Accuracy may be somewhat sacrificed by MobileNet in comparison to other models. Performance might not be as excellent as other deeper models for some complicated jobs.

3.3.3. DenseNet201

The InceptionV3 model's architecture DenseNet201: Developed in collaboration by Cornwell University, Tsinghua University, and Facebook AI Research (FAIR), is the third version of the CNN architecture known as densely linked convolutional networks, or DenseNet, which was released by CVPR in 2017 [69,77] [17]. Fully connected layers densely connected blocks, and convolutional layers make up the 201 layers of the DenseNet201 model architecture. Every layer in DesseNet201 is directly connected to every other layer, emphasizing dense connections. Better feature reuse and parameter reduction are made possible by DesseNet201's dense connectivity. It is more computationally complex than other models, performs well in tasks like image classification, and might take longer to train [18].

3.3.4. InceptionResNetV2

A convolutional neural architecture called Inception-ResNet-v2 [19] upon the Inception family of architectures by adding leftover connections, which substitute the filter concatenation stage of the Inception architecture. The InceptionResNetv2 model architecture, which piles the inception modules and utilizes residuary connections inside these modules, is composed of around 572 levels. Using residual connections, InceptionResNetV2 outperforms GoogLeNet and ResNet and achieves exceptional accuracy across a range of computer vision tasks. Due to its relatively complex architecture, InceptionResNetV2 requires a lot of processing power to train.

3.3.5. DCNN(YOLO)

The YOLO algorithm's fundamental idea is to use an end-to-end convolutional neural network to directly anticipate the target's category and location [20]. The YOLO algorithm divides the input image into $S \times S$ grid cells, and the task assigned to each grid is to forecast the target whose center point is inside it. The label category of the grid containing the target center point is the target category in the image labeling stage, whilst the label categories of other grids are referred to as the background label category [21].

3.4. Flowchart

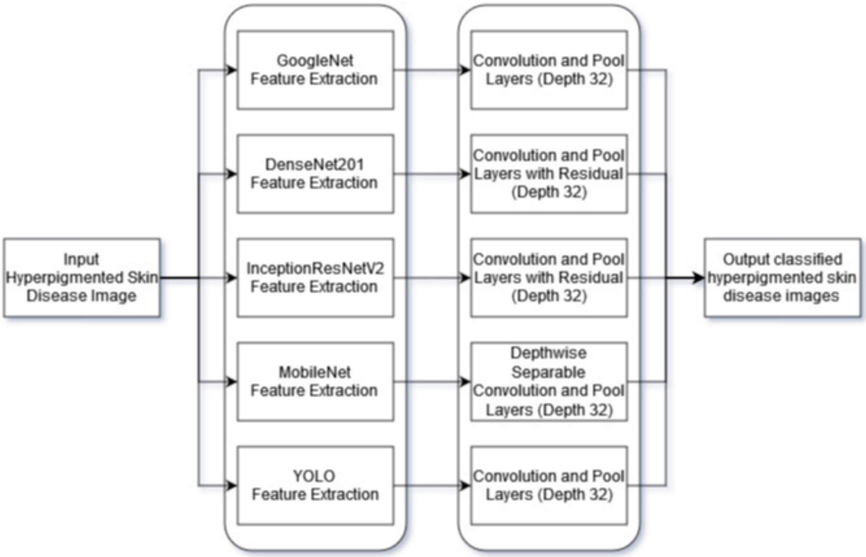

Fig. 1. Research process flowchart

The models, pre-trained on the ImageNet dataset, are used for feature extraction. The top classification layer is excluded, and the model is set to be trainable to allow fine-tuning on the specific dataset. After the feature extraction layers, additional convolutional and pooling layers with a depth of 32 are added. Smaller batch sizes introduce more noise in the gradient estimates, which can help in escaping local minima and finding better generalization solutions. Batch sizes like 32 provide a good level of noise, leading to better generalization compared to very large batch sizes which might lead to overfitting. These layers include residual connections and depth wise separable convolutions to further refine the extracted features. The model is assembled using the Adam optimizer with a categorical cross-entropy loss and a default learning rate of 0.0001. There are 50 training epochs for the model.

3.5. Statistical Analysis

3.5.1. Normalization

The automated algorithms' performance is impacted by the images' non-uniform intensities. Different kinds of images have been used to create several normalization methods for high performance [22], [23]. They might, however, result in significant computing expenses. As a result, the suggested method includes an effective normalizing technique. The following are the formulas for normalization and standardization:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

$$x'' = \frac{x' - \overline{x}}{\sigma} \tag{2}$$

The normalized value, the characteristic value, the lowest and maximum values under the characteristic, and the characteristic value are represented, respectively, by the variables X, X′, min(X), max(X), and in formulas 1 and 2 [24]. Under the feature, they show the standardized result, the variance, and the mean value.

3.5.2. Evaluation

The four deep learning systems' diagnoses of PSLs were evaluated according to their accuracy, sensitivity, specificity, and precision. The following formulas are utilized to calculate these measurements:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{FN}} \tag{5}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{6}$$

where TP, FP, FN, and TN refer to true positive, false positive, false negative, and true negative respectively. These measurements offer a thorough comprehension of how well the models categorize the various skin diseases.

3.5.3. Confusion Matrix

Confusion matrix provides a clear visual depiction of the quality of the deep learning classification model [25]. Sometimes called an error matrix, possibility table, or confusion matrix, these terms are used within machine learning and artificial intelligence. Utilizing the confusion matrix allows for visual representation results of algorithm performance, particularly in supervised learning. Matching matrices are commonly utilized in unsupervised learning. The discrepancies between the results of the classification and the actual values may be used to illustrate the accuracy of the classification results in a confusion matrix. Each measured pixel is classified by comparing its position with the matching status and classification in the identified picture. This process creates the confusion matrix.

3.5.4. Sensitivity Specificity Curve

AUC (Area Under Curve) is an indicator that's used to assess how well a deep learning classification model's performance is doing. One performance indicator used to assess the calibre of students is the area under the ROC curve, or AUC [26]. AUC value is frequently utilized as the assessment criteria since ROC curves are not always able to clearly show which classifier is superior. It is advisable to use the classifier with a higher AUC value. AUC = 1 denotes a perfect classifier; 0.5 < AUC <1 indicates superiority over the random classifier; 0 < AUC <0.5 indicates inferiority over the random classifier. A ROC curve's ordinate indicates sensitivity (specificity = TN/(TN + FP)), while its abscissa shows specificity (sensitivity = TP/(TP + FN)).

4. Results

4.1. The area under ROC curve result

Table displays the AUC region and accuracy of the training set for five distinct pretrained models. The training set's corresponding accuracy rates were 93.8%, 100%, 100%, 98.77% and 97,43% after 50 iterations of the GoogleNet, MobileNet, DenseNet201, InceptionResNetV2, and YOLO respectively. The test set data's accuracy rates were 87.18%, 79.49%, 87.18%, 89.74%, and 97.56%, in that order. The AUC curves have the following relative areas: 0.90, 0.98, 0.99, 0.94, and 0.97.

Table 2. The result for each model's accuracy

Model	**Train Accuracy (%)**	**Test Accuracy (%)**	**AUC**
GoogleNet	93.80	87.18	0.90
MobileNet	100	79.49	0.98
DenseNet201	100	87.18	0.99
InceptionResNetV2	98.77	89.74	0.94
YOLO	97.43	97.56	0.97

Based on Table 2, it can be concluded that, some models have excellent training accuracy but fail to generalize well to unseen data, indicating potential overfitting issues.

GoogleNet, although a powerful model, may underperform due to its architectural complexity. It has multiple branches and convolutions that require extensive training data to fine-tune. The moderate test accuracy suggests potential overfitting, where the model performs well on the training data but struggles to generalize to unseen data. Similarly, InceptionResNetV2 combines the strengths of Inception and ResNet architectures, leading to high training accuracy and good test accuracy. However, its slightly lower test accuracy compared to other models suggests it might not generalize as well to new data. This could be due to the model's complexity, which requires careful tuning of hyperparameters and regularization techniques to prevent overfitting.

MobileNet is designed for efficiency and speed, often at the expense of accuracy. It results in this model not fitting a tiny dataset. The perfect training accuracy combined with a significantly low-test accuracy indicates severe overfitting. MobileNet's lightweight architecture might not capture complex patterns in small dataset as effectively as heavier models, leading to lower performance on the test set.

DenseNet201 achieves high training accuracy and a high AUC, suggesting that it models the data well. However, its test accuracy indicates some overfitting. DenseNet's dense connectivity pattern can lead to high computational costs and require extensive regularization to avoid overfitting, which may not have been adequately addressed in this setup.

The results indicate that YOLO outperform every other model with the most balance performance in both training and test accuracy. Excellent classification outcomes for CS skin pigmentation illness have been demonstrated by the majority of classification methods. In summary, the analysis revealed that the YOLO architecture was superior at detecting objects, considerable generalization power and increased accuracy with a small dataset.

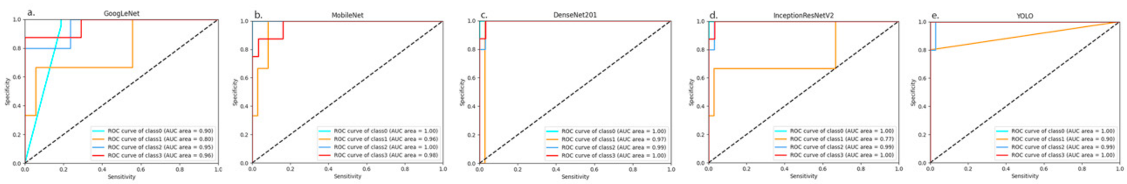

Fig. 2 Area under the ROC Curve of 6 CNN algorithm model for hyperpigmented skin disease classification. The AUC for skin disease prediction of GoogleNet(a), MobileNet(b), DenseNet201(c), InseptionResNetV2(d), and YOLO(e).

4.2. The Confusion Matrix Result

The confusion matrices for 4 classification tasks and five classical CNNs are shown in Fig. 3, respectively. As seen in the graphic, each confusion matrix's elements (i and j) indicate the likelihood that class j(Predicted label) will be predicted when class i(True label) is the actual case. Where CS, CN, MN, and ML are represented by the numbers 0, 1, 2, 3, respectively. The only element that the YOLO model to be selected for detecting hyperpigmentation disease was the confusion matrix.

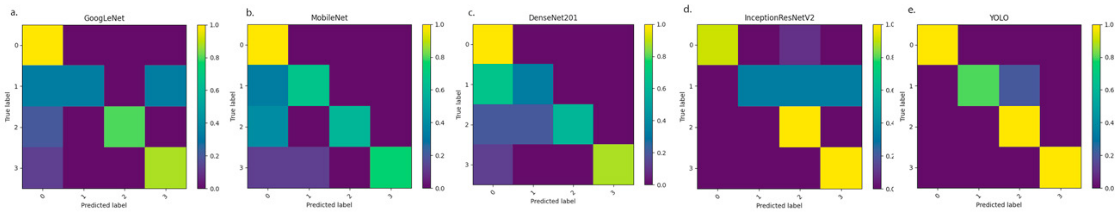

Fig. 3. Confusion matrix of all 6 models: GoogleNet(a), MobileNet(b), DenseNet201(c), InseptionResNetV2(d), and YOLO(e).

5. Conclusion

The performance of various deep learning algorithms—including GoogleNet, MobileNet, DenseNet201, InceptionResNetV2, and YOLO—in the classification of hyperpigmented skin conditions was assessed in this comparative study.

Results: MobileNet and DenseNet201 exhibit signs of overfitting, indicated by perfect training accuracy but lower test accuracy. This suggests the need for better regularization and more diverse training data. GoogleNet and InceptionResNetV2, while powerful, might suffer from their architectural complexity, requiring more extensive hyperparameter tuning and regularization. MobileNet, optimized for efficiency, may sacrifice some accuracy, leading to lower test performance with low dataset compared to more complex models. YOLO is the best model for this classification task, which was shown by the most stable training accuracy, test accuracy, and AUC score. YOLO's single-shot approach helps it balance accuracy and speed, making it highly effective for this classification task.

The work highlights the potential of deep learning models to diagnose hyperpigmented skin disorders accurately, giving dermatologists a useful tool. To improve the accuracy and efficacy of these models in clinical settings, in particular, further research and optimization are necessary, as seen by the disparities in performance between the models.

In order to improve the models' capacity for generalization, future work will concentrate on growing the dataset, adding more varied skin conditions, and improving the models themselves. Furthermore, investigating hybrid models and incorporating them into clinical decision support systems may enhance patient outcomes and diagnostic precision even further.

References

[1] A. M. Thawabteh, A. Jibreen, D. Karaman, A. Thawabteh, and R. Karaman, “Skin Pigmentation Types, Causes and Treatment—A Review,” *Molecules*, vol. 28, no. 12. MDPI, Jun. 01, 2023. doi: 10.3390/molecules28124839.

[2] B. Zhang *et al.*, “Opportunities and Challenges: Classification of Skin Disease Based on Deep Learning,” *Chinese Journal of Mechanical Engineering (English Edition)*, vol. 34, no. 1. Springer, Dec. 01, 2021. doi: 10.1186/s10033-021-00629-5.

[3] Y. Liu *et al.*, “A deep learning system for differential diagnosis of skin diseases,” *Nat Med*, vol. 26, no. 6, pp. 900–908, Jun. 2020, doi: 10.1038/s41591-020-0842-3.

[4] J. Zhang, Y. Xie, Q. Wu, and Y. Xia, “Medical image classification using synergic deep learning,” *Med Image Anal*, vol. 54, pp. 10–19, May 2019, doi: 10.1016/j.media.2019.02.010.

[5] P. Tschandl *et al.*, “Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study,” *Lancet Oncol*, vol. 20, no. 7, pp. 938–947, Jul. 2019, doi:

10.1016/S1470-2045(19)30333-X.

[6] DermNet NZ, “Dermnet,” Kaggle. Accessed: Feb. 19, 2024. [Online]. Available: https://www.kaggle.com/datasets/shubhamgoel27/dermnet

[7] N. Ibrahim, “cafe-au-lait-spots,” GitHub. Accessed: Apr. 01, 2024. [Online]. Available: https://github.com/NoelleI/cafe-au-lait-spots/tree/master

[8] A. M. Carrington *et al.*, “Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation,” *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 1, pp. 329–341, Jan. 2023, doi: 10.1109/TPAMI.2022.3145392.

[9] V. Reddy Allugunti, “A machine learning model for skin disease classification using convolution neural network,” *~ 141 ~ International Journal of Computing, Programming and Database Management*, vol. 3, no. 1, pp. 141–147, 2022, [Online]. Available: https://dermnetnz.org/.

[10] M. W. P. Maduranga and D. Nandasena, “Mobile-Based Skin Disease Diagnosis System Using Convolutional Neural Networks (CNN),” *International Journal of Image, Graphics and Signal Processing*, vol. 14, no. 3, pp. 47–57, Jun. 2022, doi: 10.5815/ijigsp.2022.03.05.

[11] C. Szegedy *et al.*, “Going Deeper with Convolutions,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.4842

[12] N. Yang, Z. Zhang, J. Yang, Z. Hong, and J. Shi, “A Convolutional Neural Network of GoogLeNet Applied in Mineral Prospectivity Prediction Based on Multi-source Geoinformation,” *Natural Resources Research*, vol. 30, no. 6, pp. 3905–3923, Dec. 2021, doi: 10.1007/s11053-021-09934-1.

[13] S. Sa’idah, I. Putu, Y. Nugraha Suparta, and E. Suhartono, “Modification of Convolutional Neural Network GoogLeNet Architecture with Dull Razor Filtering for Classifying Skin Cancer,” 2022.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, “Rethinking the Inception Architecture for Computer Vision.”

[15] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.04381

[17] M. Gharaibeh *et al.*, “Early Diagnosis of Alzheimer’s Disease Using Cerebral Catheter Angiogram Neuroimaging: A Novel Model Based on Deep Learning Approaches,” *Big Data and Cognitive Computing*, vol. 6, no. 1, Mar. 2022, doi: 10.3390/bdcc6010002.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks.” [Online]. Available: https://github.com/liuzhuang13/DenseNet.

[19] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.” [Online]. Available: www.aaai.org

[20] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A Review of Yolo Algorithm Developments,” in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 1066–1073. doi: 10.1016/j.procs.2022.01.135.

[21] H. Ding *et al.*, “Automatic identification of benign pigmented skin lesions from clinical images using deep convolutional neural network,” *BMC Biotechnol*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12896-022-00755-5.

[22] E. Goceri, “Fully Automated and Adaptive Intensity Normalization Using Statistical Features for Brain MR Images,” *Celal Bayar Üniversitesi Fen Bilimleri Dergisi*, vol. 14, no. 1, pp. 125–134, Mar. 2018, doi: 10.18466/cbayarfbe.384729.

[23] E. Goceri, “INTENSITY NORMALIZATION IN BRAIN MR IMAGES USING SPATIALLY VARYING DISTRIBUTION MATCHING,” 2017.

[24] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, May 27, 2015. doi: 10.1038/nature14539.

[25] E. Passia *et al.*, “OP0059 AUTOSCORA: DEEP LEARNING TO AUTOMATE SCORING OF RADIOGRAPHIC PROGRESSION IN RHEUMATOID ARTHRITIS,” *Ann Rheum Dis*, vol. 79, no. Suppl 1, pp. 38–39, Jun. 2020, doi: 10.1136/annrheumdis-2020-eular.3461.

[26] J. Huang and C. X. Ling, “Using AUC and Accuracy in Evaluating Learning Algorithms.” [Online]. Available: http://www.computer.org/publications/dlib