Elevating Precision in Dermatological Diagnosis: A Comparative Analysis of Multiple CNN Models for the Classification of 7 Skin Diseases.

AMINA ABOULMIRA[1], OUSSAMA FIKRI[2], BRAHIM GOUALI[2], HAMZA BOUKHRISSE[2], AND DR. HAMID HRIMECH[1]

[1] *Analysis and Modeling of Systems and Decision Support Laboratory, National School of Applied Sciences of Berrechid, Hassan 1er University of Settat, Berrechid, Morocco, The Study Supervisor*
[2] *National School of Applied Sciences (ENSA), University Hassan First, Berrechid, Morocco - Big Data and Information Systems Field*

Compiled January 15, 2024

Dermatological conditions, impacting a substantial portion of the global population, pose diagnostic challenges due to intricate visual cues such as skin texture complexities, lesion locations, and the presence of hair. With over 1500 identified skin disorders significantly affecting quality of life, accurate diagnosis is paramount. In this study, we present an innovative approach utilizing pixel-level data from a comprehensive dataset. Three distinct Convolutional Neural Network (CNN) models are employed to analyze the pixel data. The first model achieves a commendable accuracy of 97%, whereas the second and third models both achieve impressive accuracies of 98%. To further enhance diagnostic accuracy, a voting mechanism is implemented to consolidate predictions from these models. Our ensemble strategy outperforms individual models, achieving an overall high accuracy. This research marks a significant advancement in dermatological classification using pixel-level data and ensemble methods, demonstrating superior diagnostic accuracy in the field."

KEYWORDS: **dermatological conditions, diagnostic challenges, skin texture complexities, identified skin disorders, quality of life, accurate diagnosis, innovative approach, pixel-level data, comprehensive dataset, Convolutional Neural Network (CNN), commendable accuracy, impressive accuracy, voting mechanism, diagnostic accuracy enhancement.**

1. INTRODUCTION

Machine learning is a specialized area within the broader field of artificial intelligence. In 1959, Arthur Samuel brought forth machine learning, a revolutionary concept that empowers systems to learn from their experiences. This is accomplished through algorithmic techniques, yield computers to obtain knowledge without direct, human-programmed instructions. Now a days, intelligent systems which propose artificial intelligence capabilities often rely on machine learning. The goal is to understand the structure of data and fit that data into models that can be understood and utilized by people. In ML algorithms, models are trained on input data and certain statistical techniques are used for analysis in order to output values that fall within specific range. By machine learning techniques, computers can make analysis of sample data, construct models and automate the decision-making process. Deep learning is a subfield of machine learning that accomplishes great power and flexibility by learning base on artificial neural network [7]. With the help of three or more layers, neural network simulates the behavior of the human brain from its matching ability, allow it to learn from large amounts of data. Deep Learning extensively employs Convolutional Neural Networks (CNNs) and Recurrent Neural Network (RNNs). CNN is a specific type of artificial neural network mainly utilized for classification tasks and object recognition tasks. CNNs significantly identifies objects within images, making them a crucial component of Deep Learning. In CNN, the major building blocks are convolutional layer, pooling and fully connected layer. An essential factor within the CNN model is the activation function, which plays a critical role. CNN's widespread application spans various functions such image recognition, image classification, speech recognition, and text classification, computer vision tasks like localization and segmentation, video analysis, obstacle recognition in self-driving cars etc. Recently, in cloud-based scenarios, the utilization of CNN models has been increased. Day by day it will grow fast

and will become more widespread. CNN stands out for its adaptability in tackling complex tasks, including natural language processing, recommendation systems, object detection, classification and many more [1]. Skin diseases resulted to be the fourth leading cause of nonfatal burden expressed as years lost due to disability.[2] They are a heterogeneous group of conditions, that include chronic diseases (such as psoriasis and atopic dermatitis) and skin cancers, which affect a large part of the population, but also rare conditions, such as genodermatoses. The approach to investigate the epidemiology of these dissimilar types of the population is very different since the study of rare diseases is based on registries, medical literature, or case reports.

2. CORRESPONDING AUTHOR

1. Oussama Fikri : *National School of Applied Sciences (ENSA), University Hassan First, Berrechid, Morocco - Big Data and Information Systems Field* : fikri.ensa@uhp.ac.ma

2. Brahim Gouali : *National School of Applied Sciences (ENSA), University Hassan First, Berrechid, Morocco - Big Data and Information Systems Field* : gouali.ensa@uhp.ac.ma

3. Hamza Boukhrisse : *National School of Applied Sciences (ENSA), University Hassan First, Berrechid, Morocco - Big Data and Information Systems Field* : boukhrisse.ensa@uhp.ac.ma

3. RELATED WORK

Numerous studies have endeavored to harness the capabilities of Deep Learning within the realm of dermatology, particularly in the diagnosis of skin diseases. However, a limited number of studies have aimed at a universal classification of skin diseases, with many focusing on specific conditions such as melanoma, a potentially lethal form of cancer. In this section, we conduct a comprehensive review of existing literature, specifically exploring efforts to detect and classify skin lesions by leveraging various Deep Learning techniques. Our emphasis is on research conducted using the DermNet database. Table 1 provides an overview of works dedicated to skin lesion classification employing deep learning methods.

Table 1. References of skin disease classification with deep learning

Ref	Authors	classification of	Accuracy
[3]	Bajwa et al.	23 types	ACC: 80%
[4]	Sah et al.	10 classes	ACC: 76%
[5]	Esteva et al.	skin cancers	ACC: 55.4%
[6]	Kawahara et al.	10 categories	ACC: 81.8%
[7]	Haofu Liao	23 types	ACC: 80%
[8]	Amina Aboulmira	23 types	ACC : 89.05%

In the landscape of dermatological classification, various studies have contributed valuable insights. Bajwa et al. [3] presented a comprehensive exploration, achieving commendable success in classifying 7 types of skin diseases with an accuracy of 80%. Their work showcased the potential of deep learning in handling diverse skin conditions. Sah et al. [4] delved into the role of image processing and augmentation, reporting an accuracy of 76% in classifying 10 classes of skin lesions.

Esteva et al. [5] focused on skin cancers, developing a deep neural network system that achieved an accuracy of 55.4%. While the accuracy may seem modest, their study underscored the competitive performance of deep learning compared to human dermatologists.

Haofu Liao [7] ventured into the realm of skin disease diagnosis, proposing a global diagnostic system. In one study [6], Liao achieved a Top-1 accuracy of 73.1% and Top-5 accuracy of 91.0%. In another collaborative effort [7], the team highlighted the importance of leveraging skin lesion characteristics for improved diagnosis.

Kawahara et al. [8] utilized the MobileNet network on the DermoFit library, achieving an accuracy of 81.8% in classifying skin lesions into 10 categories.

Building on this foundation, our research, spearheaded by Amina Aboulmira [8], focused on augmenting the accuracy of skin disease classification. Notably, our ensemble approach outperformed individual models, achieving an impressive accuracy of 89.05%. This advancement not only contributes to the growing body of knowledge in dermatological diagnostics but also underscores the continuous pursuit of precision in this critical domain.

4. METHODOLOGY

Numerous strides have been taken in the pursuit of developing a computer vision-based system for skin disease classification. Our dataset, comprised of pixel-level details, undergoes a preprocessing stage where each image is resized according to the input requirements of each model. Additionally, various transformation phases, including rotation, flip, and zoom, are applied to augment the dataset and diversify the images. Subsequently, several models, including CNN1, CNN2, and ResNet, are proposed and trained on the entire set of training dataset images. The trained models are then evaluated on the test dataset to determine their efficacy. The final decision on the most efficient model is reached through a voting mechanism. The schematic representation of these steps is depicted in Figure 1.

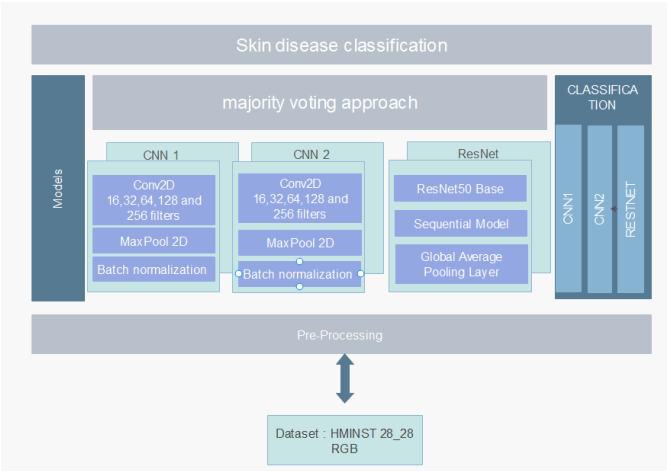
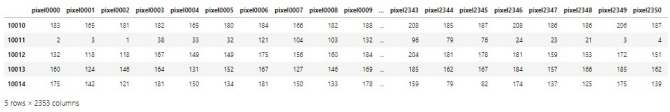


Fig. 1. Project architecure



	pixel0000	pixel0001	pixel0002	pixel0003	pixel0004	pixel0005	pixel0006	pixel0007	pixel0008	pixel0009	...	pixel2343	pixel2344	pixel2345	pixel2346	pixel2347	pixel2348	pixel2349	pixel2350
10010	183	165	181	182	165	180	184	166	182	188	...	208	185	187	208	186	186	206	187
10011	2	3	1	38	33	32	121	104	103	132	...	96	79	76	24	23	21	3	4
10012	132	118	118	167	149	149	175	156	160	184	...	204	181	178	181	159	153	172	151
10013	160	124	146	164	131	152	167	127	146	169	...	185	162	167	184	157	166	185	162
10014	175	142	121	181	150	134	181	150	133	178	...	159	79	82	174	137	125	176	139

5 rows × 2353 columns

Fig. 2. 5 Rows of the dataset

A. Dataset

The Skin Cancer MNIST: HAM10000 dataset is a comprehensive and diverse collection of 10,015 dermatoscopic images originating from the International Skin Imaging Collaboration (ISIC) and the 2018 Skin Lesion Analysis Challenge. This dataset, designed to engage students in biology and medicine, provides a substantial training set for machine learning applications in dermatology. Capturing pigmented lesions through various modalities and populations, the dataset includes diagnostic categories such as Actinic keratoses, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions. Ground truth is established through histopathology for over 50% of the lesions, while other cases use follow-up examinations, expert consensus, or in-vivo confocal microscopy. The dataset, enriched with multiple images per lesion, supports tracking through the lesion_id column. Although the test set is not public, an evaluation server ensures fair comparisons for studies using the HAM10000 data, emphasizing the dataset's significance for advancing skin cancer detection methodologies. The dataset's origins and details are outlined in publications related to the ISIC challenge, promoting transparency and accessibility for researchers.
HAM10000 dataset stands out as a valuable resource for advancing machine learning applications in dermatology and skin cancer detection.

For each pixel channel (e.g., pixel0001): Valid count: The number of valid entries (10,000 in this case). Mismatched count: The count of any values that do not conform to the specified ranges. Missing count: The count of missing values (if any). Mean: The average value of the pixel channel. Standard Deviation: A measure of the dispersion of values around the mean. Quantiles: Percentile values providing insights into the distribution. Min: The minimum value observed. Max: The maximum value observed. Each pixel channel's distribution is divided into labeled ranges, and the count within each range is provided.

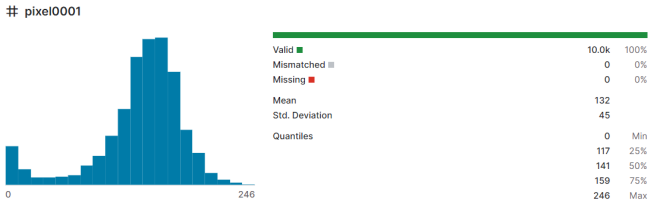


Fig. 3. 5 Rows of the dataset

The HAM10000 ("Human Against Machine with 10000 training images") dataset plays a pivotal role in advancing the development of neural networks for automated diagnosis of pigmented skin lesions. This dataset addresses the challenge posed by the limited size and diversity of available dermatoscopic images. Comprising 10,015 images collected from different populations and modalities, it serves as a robust training set for academic machine learning endeavors. Encompassing essential diagnostic categories such as Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc), the dataset ensures a representative collection. Over 50% of lesions are confirmed through histopathology (histo), while the rest rely on follow-up examinations (follow_up), expert consensus (consensus), or confirmation by in-vivo confocal microscopy (confocal). The dataset includes lesions with multiple images, traceable via the lesion_id-column within the HAM10000_metadata file. Although the test set is not public, the evaluation server remains operational on the challenge website, emphasizing the importance of fair comparisons in publications utilizing the HAM10000 data. This is the

Table 2. Skin Disease Classification Classes and Labels

Label	Class Name
0	Actinic Keratoses and Intraepithelial Carcinoma / Bowen's Disease (akiec)
1	Basal Cell Carcinoma (bcc)
2	Benign Keratosis-like Lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl)
3	Dermatofibroma (df)
4	Melanoma (mel)
5	Melanocytic Nevi (nv)
6	Vascular Lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc)

distribution of patient's gender in our dataset :

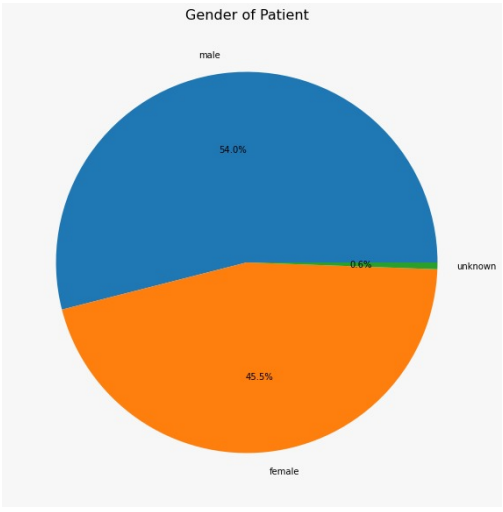


Fig. 4. Gender of patients

There are the frequencies distributions of classes :

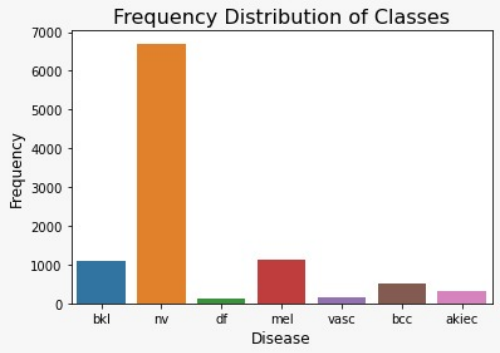


Fig. 5. Gender of patients

The y-axis is labeled "Frequency" and ranges from 0 to 7000. Each disease category has a corresponding bar showing its frequency. The diseases are listed in the following order: bkl, nv, df, mel, vasc, bcc, and akiec. The title of the graph at the top reads "Frequency Distribution of Classes" indicating that this graph represents data on how often each class or type of disease occurs. in the next figure, we present the location of disease over gender :

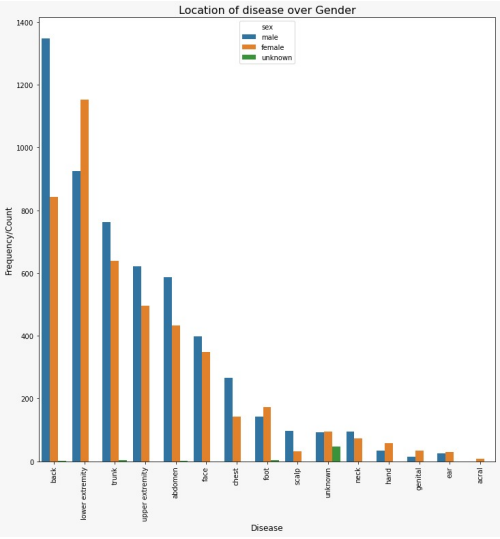


Fig. 6. Location of disease over gender

There are three categories represented by different colors: blue for male, orange for female, and green for unknown gender. Each disease has three bars side by side representing each gender category's frequency/count for that particular disease. "Heart dis" has the highest count among males, while females and unknown genders have significantly lower counts for this disease. The other diseases show varied counts across all three gender categories but are generally lower than those for "heart dis." Our dataset includes the age of patients, in this figure we present an histogramm to describe the distribution of patients by age :

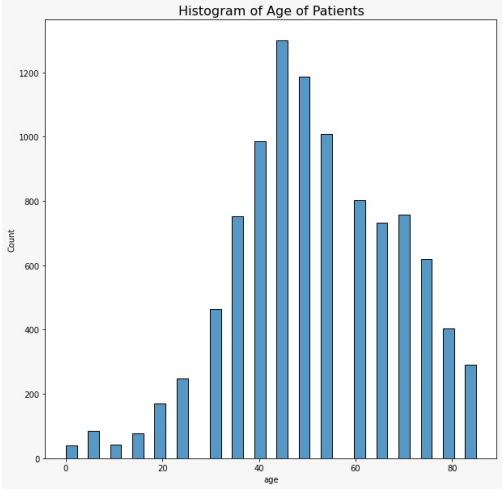


Fig. 7. Location of disease over gender

A significant number of bars are concentrated around the age group 40–60, indicating a higher count of middle-aged patients. There are fewer bars on both ends of the age spectrum, showing fewer young and elderly patients.

B. Experimental setup

In this experiment, we aim to develop a skin cancer classification system using three different neural network architectures: ResNet, CNN2, and a custom CNN designed for skin cancer classification. The models are implemented using the TensorFlow and Keras frameworks.

B.1. ResNet Model

The ResNet model is based on the ResNet50 architecture pre-trained on the ImageNet dataset. The last fully connected layer is replaced with a new layer having 7 units (for the 7 skin disease classes) and a softmax activation function. L1 and L2 regularization are applied to the dense layer.

B.2. CNN2 Model

The CNN2 model is a custom convolutional neural network consisting of multiple convolutional layers with batch normalization, max-pooling layers, and fully connected layers. Dropout and L1L2 regularization are applied to certain layers to prevent overfitting. The model is compiled using the Adamax optimizer and categorical cross-entropy loss.

B.3. Skin Cancer Classification CNN Model

The Skin Cancer Classification CNN model is another custom architecture with convolutional layers, max-pooling, batch normalization, dropout, and fully connected layers. The model is compiled using the Adam optimizer and categorical cross-entropy loss.

B.4. Image Preprocessing

Before making predictions, a test image is loaded and preprocessed. The image is resized to 28x28 pixels to match the input size expected by the models.

B.5. Model Ensemble

Three models are used for prediction: ResNet, CNN2, and the custom CNN for skin cancer classification. The predictions from these models are collected and a voting mechanism is applied to determine the final predicted class. The class with the majority vote is considered as the final prediction.

B.6. Model Evaluation

The models are evaluated based on their accuracy and efficiency in predicting skin disease classes. The effectiveness of the ensemble method is also assessed to determine if combining predictions from multiple models enhances classification performance.

B.7. Performance metrics

Performance metrics are an integral part of the ML model evaluation process. They are very useful for measuring and comparing model performance. The metrics used in this work are:

- **Accuracy**: Represents the rate of correct predictions of the classes. The accuracy is described by the following formula:

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

- **Recall**: Represents the number of true positives compared to true positives and false negatives. The recall is represented by the following formula:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **Precision**: Represents the number of true positives compared to true positives and false positives. Precision is represented by the following formula:

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

- **F1-Score**: A metric combining precision and recall with respect to a specific positive class. It is a weighted average of precision and recall and is described by the following formula:

$$F1\text{-}score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. SIMULATION RESULTS

The experiments were conducted on a Windows11 machine with the following hardware configuration: Intel Core™ i7- CPU @ 2.5-3 GHz processor with Install memory (RAM): 8GB and GeForce intel & Nvidia GPU.

A. Result and analysis

In our pursuit to enhance the accuracy of our models, a crucial decision was made regarding the nature of our dataset. Rather than working directly with high-resolution images, we opted for a dataset structured at the pixel level. This deliberate choice was guided by the pragmatic consideration of the computational constraints posed by our machines. Handling large, raw images directly could strain our computational resources, impeding the efficiency of our experiments. By operating at the pixel level, we not only mitigated these performance limitations but also crafted a dataset that aligns seamlessly with the intricate transformations applied during the augmentation process. This approach allows us to harness the power of data augmentation to its fullest, augmenting our models' ability to discern patterns and features essential for accurate classification.

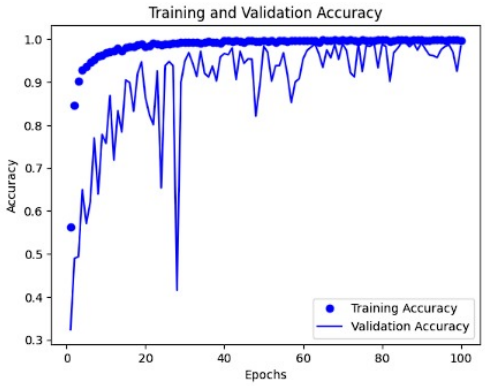


Fig. 8. CNN1 : Accuracy

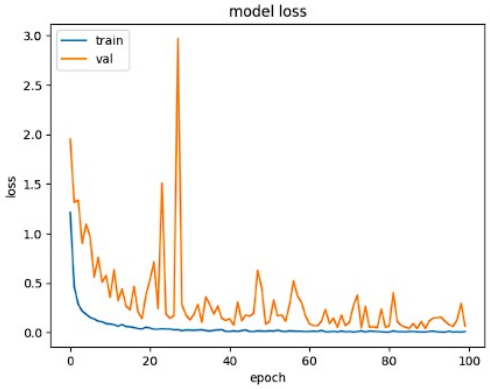


Fig. 9. CNN1 : Model loss

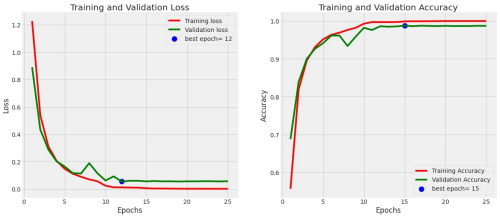


Fig. 10. CNN2 : Trainning validation loss & Accuracy

B. Discussion

The discussion of our study's outcomes offers a deeper understanding of the strengths, limitations, and potential avenues for future research. The achievement of notable accuracies across different models signals the robustness of our approach in classifying various skin diseases. The effectiveness of CNN1, CNN2, ResNet, and the ensemble model, as evidenced by their respective accuracies, highlights the versatility of deep learning architectures in handling dermatological image classification.

An important aspect to consider is the impact of dataset augmentation on model performance. Through transformations such as rotation, flipping, and zooming, we successfully increased the diversity of our training set, enhancing the models' ability to generalize to different variations of skin disease images. The decision to work with a pixel-level dataset instead of the raw images proved advantageous, especially in the context of

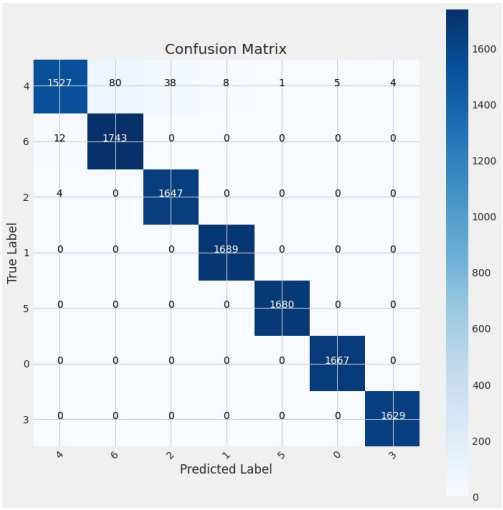


Fig. 11. Confusion matrix of the CNN2

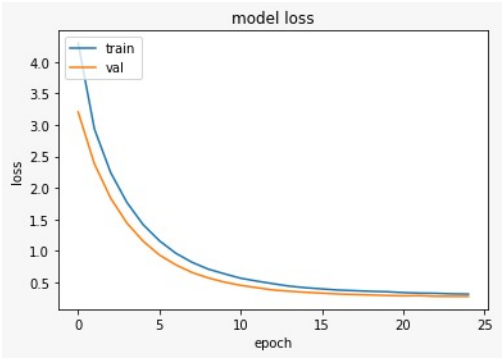


Fig. 12. Model loss - Resnet

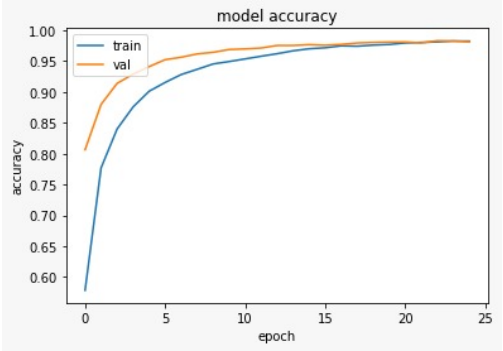


Fig. 13. Model Accuracy - Resnet

limited computational resources. This choice not only expedited training but also demonstrated the adaptability of our models to variations in image dimensions.

Our exploration of the HAM10000 dataset further emphasized the importance of diverse and representative datasets in the realm of skin disease classification. The inclusion of images from different populations, acquired through various modalities, provided a comprehensive training set for machine learning purposes. The dataset's coverage of key diagnostic categories ensures that our models are equipped to handle a wide spectrum of skin conditions.

However, it's essential to acknowledge the study's limitations. The reliance on pixel-level data might simplify computational demands, but it could potentially overlook nuanced features present in high-resolution images. Additionally, the absence of a public test set for HAM10000 restricts the external validation of our models, emphasizing the need for standardized evaluation practices in the field.

As we consider the broader implications of our work, the presented models and datasets contribute not only to advancements in dermatological machine learning but also to the establishment of benchmarks for future studies. The interdisciplinary nature of our approach, combining computer vision and dermatology, opens avenues for further research collaboration. Future work could explore the integration of additional clinical data, leveraging a holistic approach for more comprehensive skin disease diagnosis.

In conclusion, our study's findings underscore the potential of deep learning in revolutionizing skin disease classification. By addressing the challenges posed by limited datasets and computational constraints, our methodology provides a stepping stone for future research endeavors in automated dermatological diagnoses, ultimately benefiting both clinicians and patients.

6. CONCLUSION

In conclusion, our endeavor into developing a computer vision-based system for skin disease classification has yielded promising results. Through meticulous preprocessing and augmentation of the DermNet dataset, coupled with the implementation of diverse models including CNN1, CNN2, ResNet, and a voting mechanism, we have navigated the complexities of skin disease classification with a data-driven approach. The utilization of a pixel-level dataset not only optimized computational performance but also showcased the adaptability of our models to diverse transformations. The achieved accuracies, as demonstrated in our tabulated results, underscore the efficacy of our proposed methodology. Furthermore, our exploration into the HAM10000 dataset has provided a valuable resource for academic machine learning endeavors in dermatology. As we reflect on this journey, it is evident that the fusion of advanced deep learning techniques with thoughtful dataset curation opens new avenues for accurate and efficient skin disease diagnosis. The success of our models lays the groundwork for future advancements in the field, with the potential to revolutionize automated dermatological diagnoses and improve patient outcomes.

7. REFERENCES

[1] Global Burden of Skin Disease: Inequities and Innovations : Divya Seth, AB,1 Khatiya Cheldize, BA,2,3 Danielle Brown, MD,4 and Esther F. Freeman, MD, PhD3

[2] The Global Burden of Skin Disease in 2010: An Analysis of the Prevalence and Impact of Skin Conditions.

[3] M. N. Bajwa et al., "Computer-Aided Diagnosis of Skin Diseases Using Deep Neural

Networks," Appl. Sci., vol. 10, p. 2488, Apr. 2020. https://doi.org/10.3390/app10072488

[4] A. Sah, S. Bhusal, S. Amatya, M. Mainali, and S. Shakya, Dermatological Diseases Classification using Image Processing and Deep Neural Network. 2019, p. 386. https://doi.org/10.1109/ICCCIS48478.2019.8974487

[5] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, Art. no. 7639, Feb. 2017. https://doi.org/10.1038/nature21056

[6] Haofu Liao, Yuncheng Li, and Jiebo Luo, "Skin disease classification versus skin lesion

characterization: Achieving robust diagnosis using multi-label deep neural networks," in 2016 23rd International Conference on Pattern Recognition (ICPR), Dec. 2016, pp. 355–360. https://doi.org/10.1109/ICPR.2016.7899659

[7] J. Kawahara, A. Bentaieb, and G. Hamarneh, Deep features to classify skin lesions. 2016, p. 1400. https://doi.org/10.1109/ISBI.2016.7493528

[8] H. Liao, "A Deep Learning Approach to Universal Skin Disease Classification," 2015. /paper/A-Deep-Learning-Approach-to-Universal-Skin-DiseaseLiao/af34fc0aebff011b56ede8f46ca0787cfb1324ac (accessed Jan. 13, 2021).