Skin Disease Image Classification

Manjiri Bhandarwar MB8070@NYU.EDU
Hannah Phung HMP3554@NYU.EDU
Yen Le YTL2008@NYU.ED

Milestone 1

1. Motivation

Accurate diagnosis of skin disease requires years of dermatological experience, thus misdiagnosis leading to inappropriate treatment that could exacerbate one's condition is common. To address this issue, it is vital to create a reliable tool to assist dermatologists in making accurate judgements. From our literature review, we learned that current deep learning models for skin disease detection have high accuracies, however, they may not generalize well to new unseen datasets due to small datasets, different domains, model complexity, and lack of a standard dermatological evaluation criteria. Additionally, most of the papers fail to include ethical considerations, such as the performance of the models across different age groups and sexes. Thus, the purpose of our project is to try various CNNs models, as that is used most widely in the past, and test their performances on multiple datasets to find a generalizable model. As most deep learning models are black box models, we also aim to use techniques to improve the interpretability of our models and evaluate their safety, reliability and unbiasedness. The success of this research would mean better, faster, and cheaper diagnosis of skin diseases, making healthcare more accessible to people of all socioeconomic classes.

2. Related Works

Verma, et al., (2019) used a ensemble model of five data mining methods, including Classification and Regression Trees (CART), Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF) and Gradient Boosting Decision Tree (GBDT) to classify six types of skin diseases using a dataset from the UCI machine repository. They employed data-driven methods for patient record selection and normalization techniques for data preprocessing. Their model achieves a high accuracy of 98.64%, and an all average recall, precision, and F1 score of 0.99, and outperforms the individual models. The paper achieves its goal of finding a superior machine algorithm for skin disease prediction, building on past literature reviews and previous methodologies. However, the paper has limitations, including a lack of code and data link for replication, a small dataset of 355 samples potentially leading to overfitting and generalizability issues, unclear handling of missing values, and a lack of data augmentation. (Appendix A)

Jain et al., (2022) uses an Optimal Probability-Based Deep Neural Network (OP-DNN) to classify 4 different types of skin diseases using 3400 images from the ISIC database. The researchers did thorough data-preprocessing. They removed unwanted contents in the image, such as hair, using noise removal median filter, contrast enhancement by means of HE-histogram equalization, and domain-specific morphological operation. They also performed feature extraction based on color and texture to get the minimized, important and non-repetitive representation of images. This pillar step of their methodology is using the OP-DNN along with whale optimization, as it is faster than other methods, to obtain optimal weights while training. The model performed very well: 95% of accuracy, 0.97 of specificity, and 0.91 of sensitivity, and

compared to the performance of KNN and ANN, OP-DNN performs significantly better. The paper's limitations are using a small portion of the ISIC database (3400/23906), lack of data augmentation and a very high accuracy which combined may indicate lack of generalizability.

Wei et al., (2023) improved $DenseNet201$ and $ConvNeXt_L$ and combined them to create a 7 skin disease classifier on a Public HAM10000 dataset of 10015 images. Data augmentation (horizontal and vertical flipping, brightness adjustments, Cutmix, Augmix, and Random Erasing etc.) were applied to balance the data. Data was normalized to a [0,1] range and resized to the same dimensions. Efficient Channel Attention (ECA) and the Gated Channel Transformation (GCL) attention modules were added to the core blocks of $DenseNet201$ and $ConvNeXt_L$, respectively, to improve the feature extraction. On the public HAM10000 dataset, the model achieved a high accuracy of 95.29% and an F1-score of 89.99% and outperformed the individual models by 3-4%. Again, we see a high accuracy which raises concerns of generalizability. (Appendix B)

Sreekala et al., (2022) used an Enhanced Convolutional Neural Network (ECNN) to classify 3100 Dermoscopy images collected from PH2 and ISIC images. Images are transformed to grayscale and noise is removed. Structural Co-Occurrence Matrix (SCM), and Spectral Centroid Magnitude are used for feature extraction then the images are passed as input to the convolution layer. This way, the convolution layer takes in images without noise, and then put it into Max-Pooling to reduce overfitting. The model has a high accuracy of 97%, which is better than other models with different combinations of. Sensitivity and specificity were mentioned in the paper but the values were not given. (Appendix C)

Inthiyaz et al., (2023) used a pretrained model found on Kaggle to do transfer learning. The model makes use of CNN, ResNet50 (which has been trained on ImageNet database), and SVM for classification of 4 types of skin disease from the Xiangya-Derm dataset of 150,223 images. The photos were labelled by three expert dermatologists, normalized using SckImages to be 227x227 pixels. The results show that this proposed model extracts the best feature from skin images, and then classifies using Softmax Classifiers to attain a detection accuracy of 87%. Compared to the other models this is the lowest accuracy but largest dataset is used, which may signify that this model is more generalizable than the aforementioned ones.

Our literature review's findings is consistent with the insights from Zhang et al's paper (2021) on opportunities and challenges in deep learning for skin disease classification. Popular deep learning models for our problem are CNNs, RNNs, and GANs. Most of the research papers have thourough data preprocessing, feature extraction, and data augmentation which are critical components. For most models we see very high accuracies when small datasets are used. This may indicate the common problem of overfitting and lack of generalization to unseen data. Additionally, all papers fail to consider ethical implications, which is very crucial especially for black-box deep learning models.

3. Methodology

This is a general outline of our methodology and the specific techniques will be narrowed down as we progress. **Preprocessing**: Remove noise (e.g. using median filter), and potentially transform RGB images to grayscale. Resize, normalize, and use data augmentation methods (e.g. contrast and brightness changes, flipping, Cutmix, Augmix, etc.) to balance the dataset. Morphological operations used to remove unwanted features like hair. **Feature extractions**: feature extraction based asymmetry, border structure, color variation, lesion diameter, texture through methods such as SCM, GCL, ECA. **Model building and classification**: Try different CNN models: $ResNet50$, $DenseNet201$, and $ConvNeXt_L$. Makes use of $ResNet50$, which has been trained on

ImageNet database (transfer learning). Try models introduced in lecture and homework such as AlexNet, GoogleNet, VGG. Follow papers' improvements on existing models to observe performance enhancement: Fusion of models, adding attention mechanism to improve performance. Fine tuning hyperparameters and add mechanisms to help with overfitting/model generalization and help make computation faster. **Evaluation**: Use accuracy, specificity, sensitivity, precision, recall. Use SHAP values for ethical evaluation and to increase interpretability. Compare performance of different models to identify the best one. Use the model on new datasets to find the one that is the most generalizable.

4. Dataset

We plan to use the HAM1000 dataset which consists of 10015 dermatoscopic images from different populations as our primary dataset, acquired and stored by different modalities. We decided to use this dataset because it is the largest out of all the datasets publicly available, which can help prevent overfitting and increase generalizability. Additionally, this dataset contains 7 types of skin dieases, which is the highest compared to other datasets (4-6). Moreover, since the data is standardized, we will not have to spend too much time pre-processing data. This will allows us to use more time in training our model, and exploring more ways we can improve our model. Moreover, due to its standardization, there will be extensive pre-trained models that can be applied to it. The visualizations of the characteristics of the dataset are in the Appendix D and our github repository. We will use the datasets used in the other papers as our secondary datasets used to test generalizability.

5. Work Plan

Each member has read 2 research papers each and we all came together to compare our findings, then we identified common weaknesses and tailored our goal and methodology accordingly. Hannah has set up the github, Yen did the preliminary data analysis, and Manjiri condensed everyone's findings in the report. Weeks 9-11: data preprocessing and feature extraction will be done together, then we will finish implementing and evaluating the different CNNs (Yen: $ResNet50$, Manjiri: $DenseNet201$, Hannah: $ConvNeXtL$). We will help each other for each step and have multiple weekly check ins. Week 12-14: Additional analysis by combining models, and applying models on secondary datasets.

Appendix A: Verma, Anurag K., et al. Classification of Skin Disease Using Ensemble Data Mining

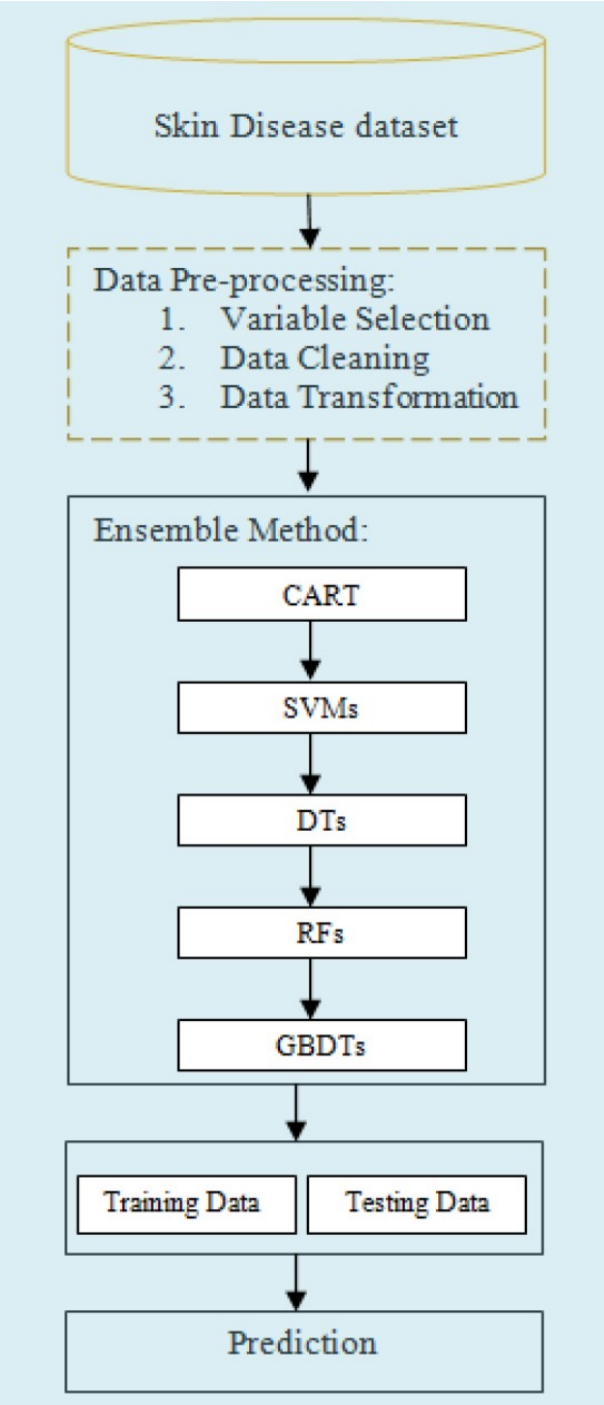

Figure 1: Method

Table 5

Output of Evaluating Ensemble Method

accuracy_score	**98.64%**				
confusion_matrix	[[24 0 0 0 0 0]				
	[0 10 0 0 0 0]				
	[0 0 11 0 0 1]				
	[0 0 0 13 0 0]				
	[0 0 0 0 11 0]				
	[0 0 0 0 0 4]]				
classification_report		precision	recall	f1-score	support
	cronic dermatitis	1.00	1.00	1.00	24.00
	lichen planus	0.91	1.00	0.95	10
	pityriasis rosea	1.00	1.00	1.00	11.0
	pityriasis rubra pilaris	1.00	0.93	0.96	14.0
	psoriasis	1.00	1.00	1.00	11
	seboreic dermatitis	1.00	1.00	1.00	4.0
	avg / total	0.99	0.99	0.99	74.0

Open in a separate window

Figure 2: Result

Appendix B: Wei, Mingjun, et al. A Skin Disease Classification Model Based on DenseNet and ConvNeXt Fusion

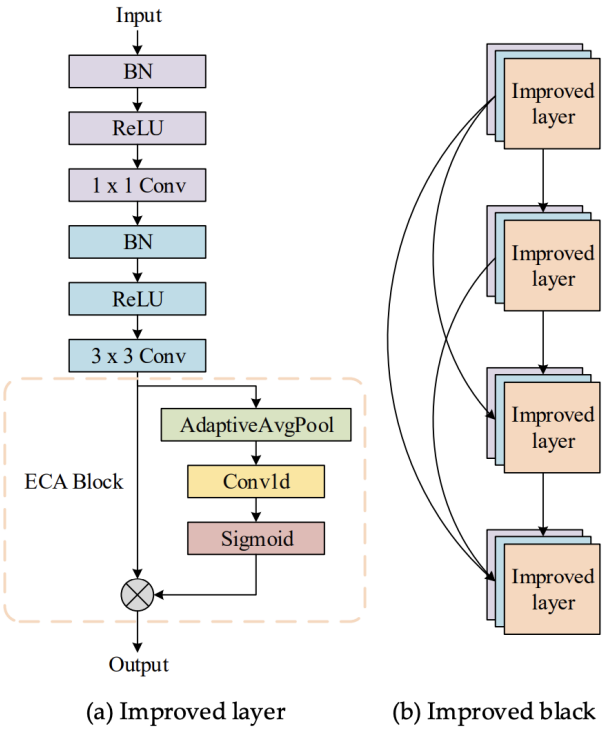

Figure 1. The structure of the improved DenseNet block.

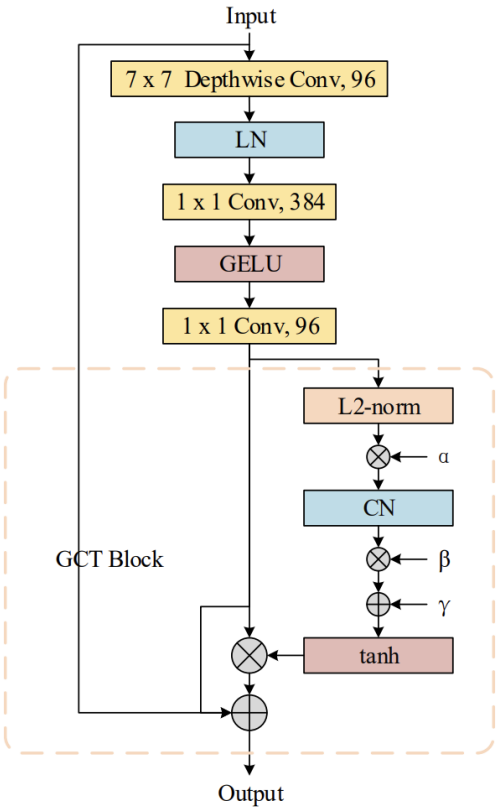

Figure 2. The structure of the improved ConvNeXt block.

Figure 1: Structures of individual DenseNet and ConvNeXt

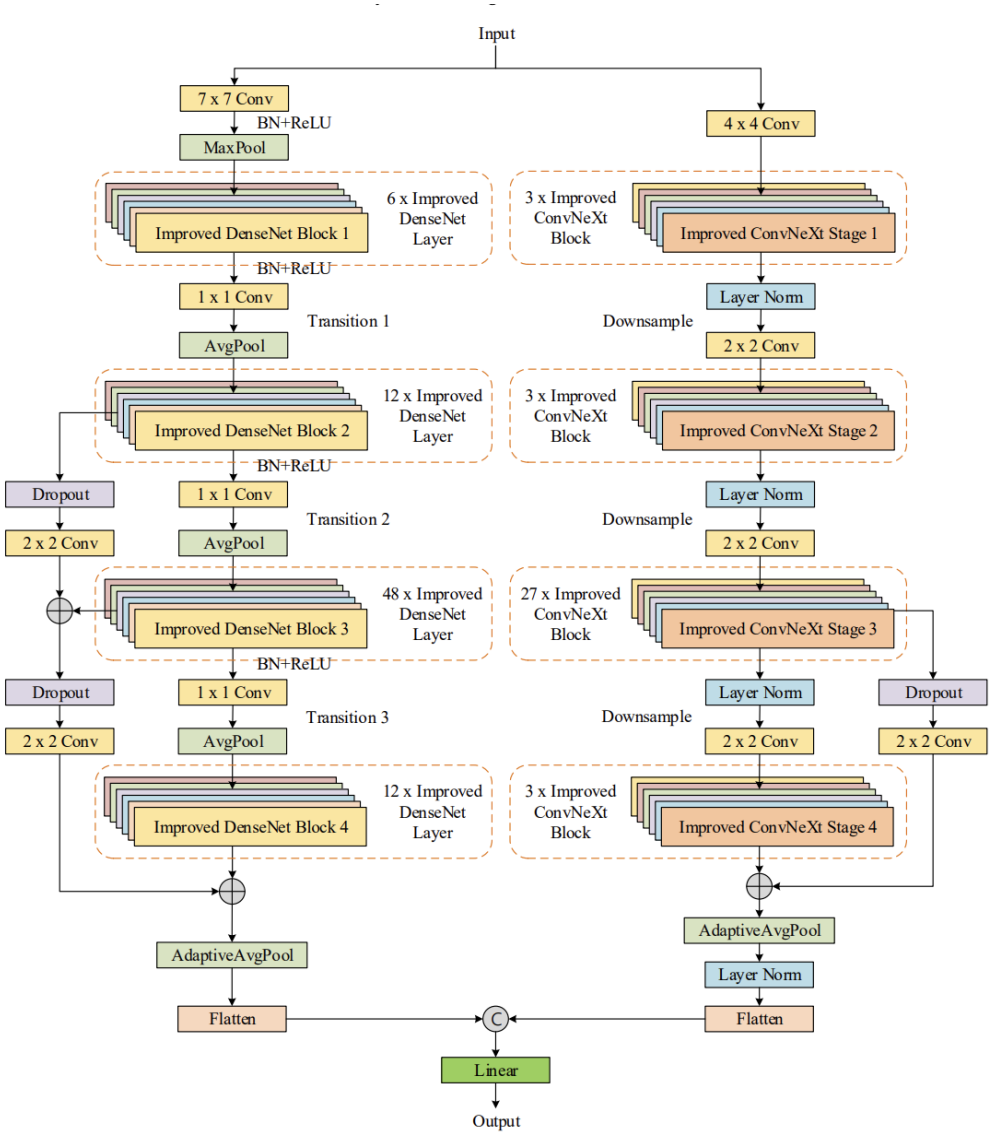

Figure 3. The full structure of the proposed model.

Figure 2: Structures of combined model

Appendix C: Sreekala, Keshetti, et al. Skin Diseases Classification Using Hybrid AI Based Localization Approach

Figure 2 implements the proposed approach for the skin diseases classification, and our proposed method implements the median filter for the preprocessing technique, the preprocessing filter is one the most crucial technique in the image processing system.

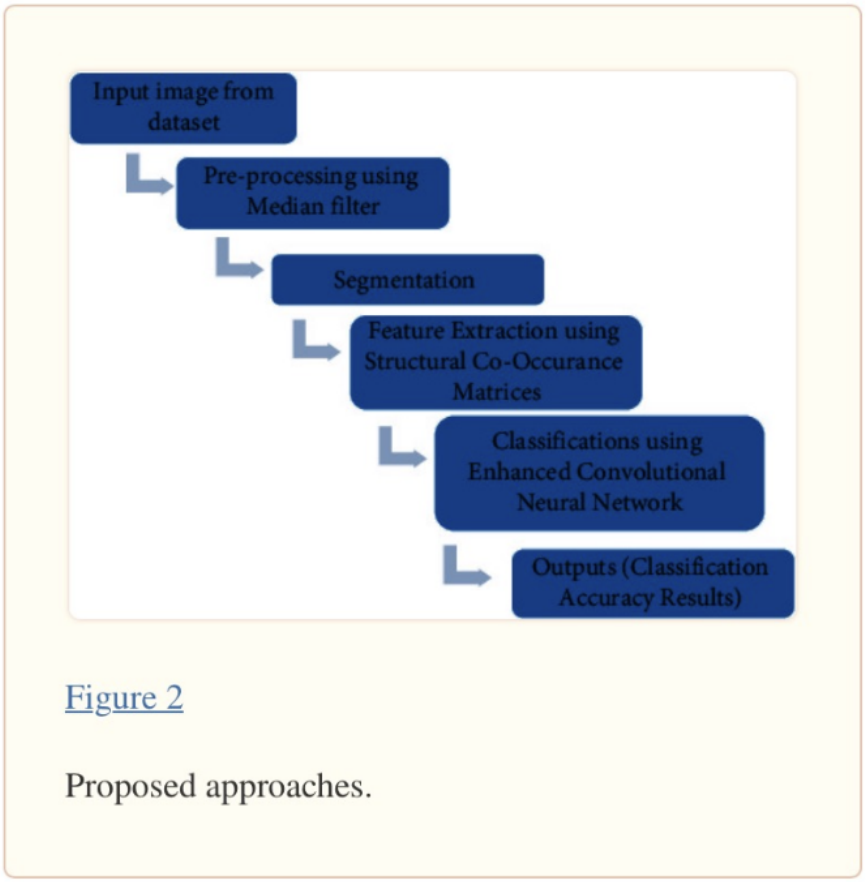

Figure 2

Proposed approaches.

Figure 1: Proposed Approaches

Figure 5 implements the comparison results for our proposed approach.

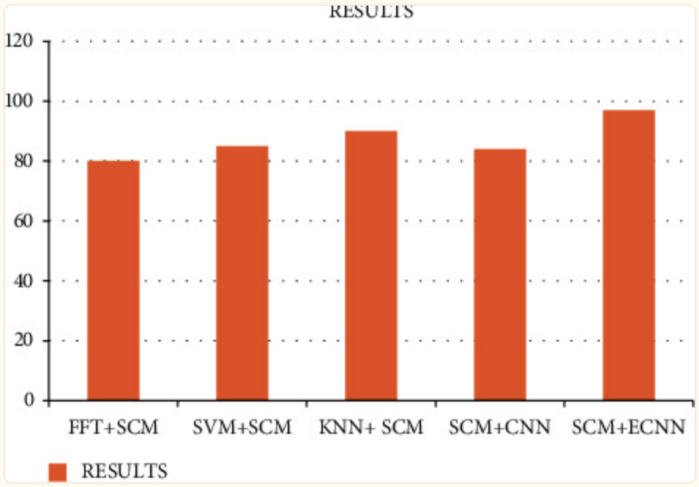

Figure 5

Results.

Figure 2: Comparison of different combinations

Appendix D: HAM10000 Dataset

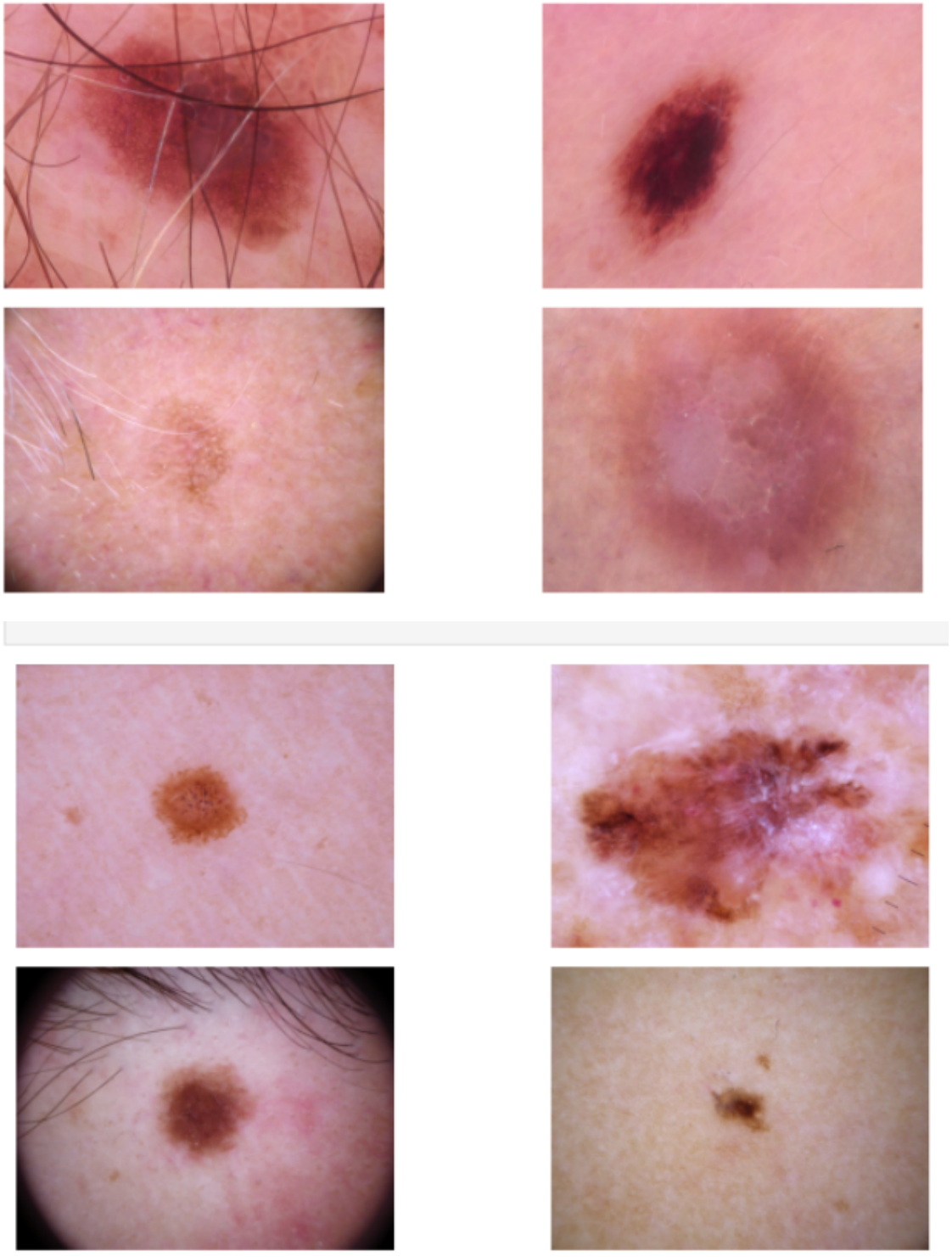

Figure 1: Skin Disease Images

	lesion_id	image_id	Diagnosis	Lession Confirmation	age	sex	localization
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear
...	...	...	...	...	...	...	...
10010	HAM_0002867	ISIC_0033084	akiec	histo	40.0	male	abdomen
10011	HAM_0002867	ISIC_0033550	akiec	histo	40.0	male	abdomen
10012	HAM_0002867	ISIC_0033536	akiec	histo	40.0	male	abdomen
10013	HAM_0000239	ISIC_0032854	akiec	histo	80.0	male	face
10014	HAM_0003521	ISIC_0032258	mel	histo	70.0	female	back

10015 rows × 7 columns

Figure 2: Dataset Attributes

Diagnostic categories in the realm of pigmented lesions:

- 1. Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec)
- 1. basal cell carcinoma (bcc)
- 1. benign keratosis-like lesions (solar lentigines / seborrheic keratoseslichen-planus like keratoses, bkl)
- 1. dermatofibroma (df)
- 1. melanoma (mel)
- 1. melanocytic nevi (nv)
- 1. vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc).

Figure 3: Skin Disease Classes

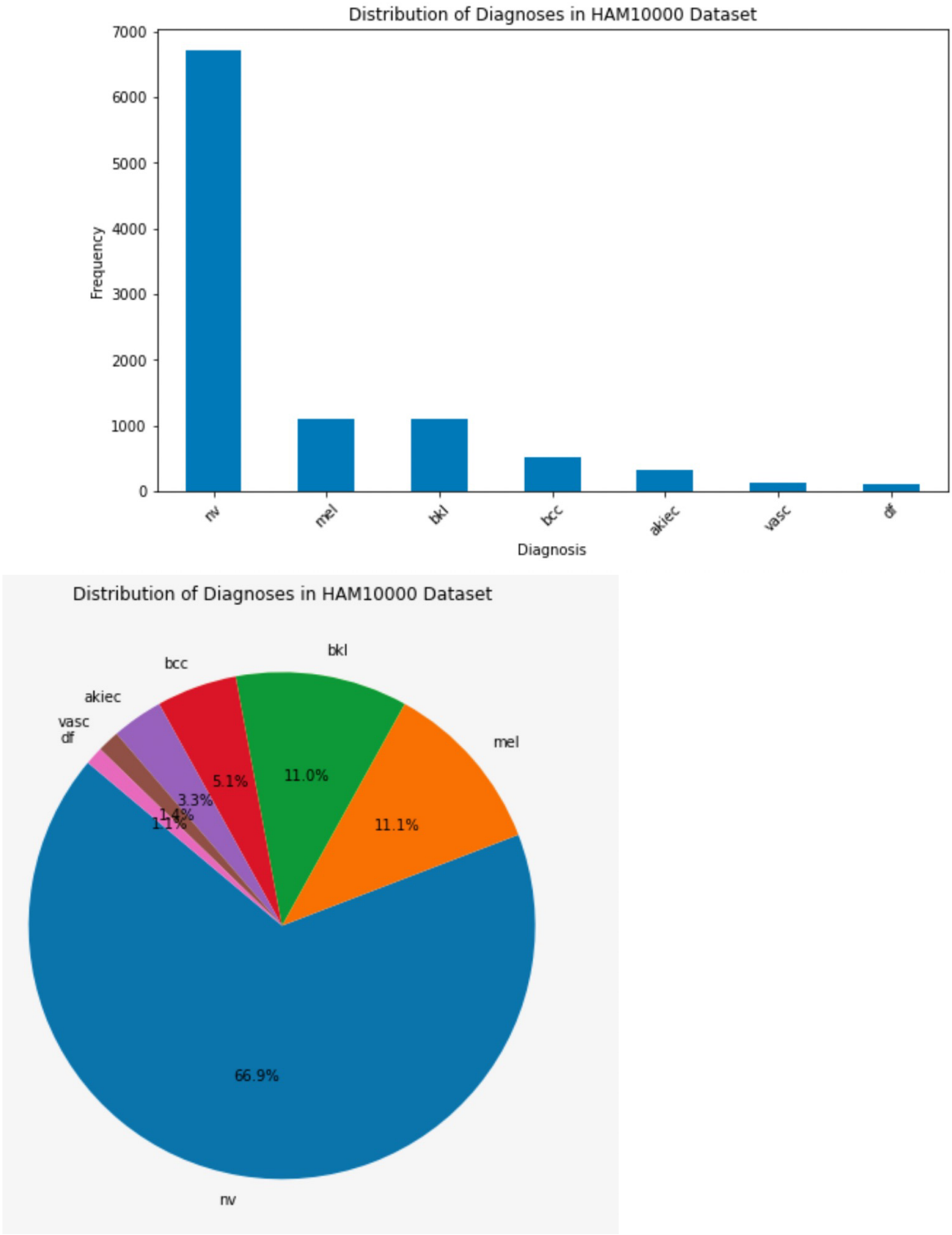

Figure 4: Distribution of 7 skin disease types

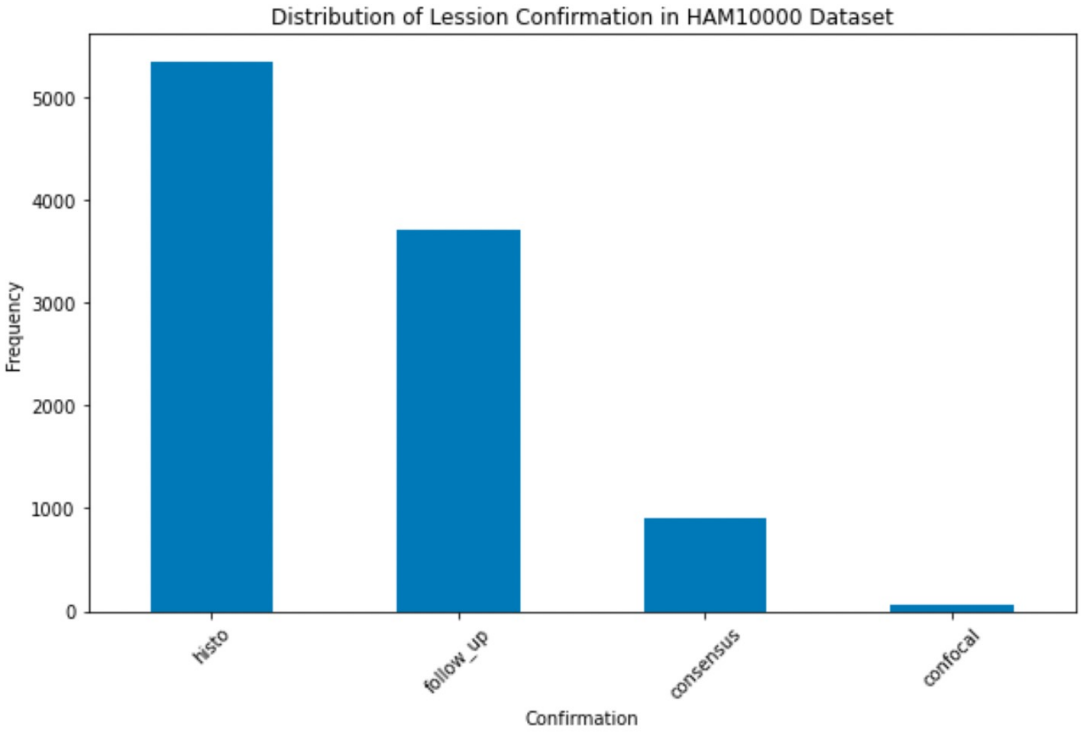

Figure 5: Distribution of Lesions Confirmations

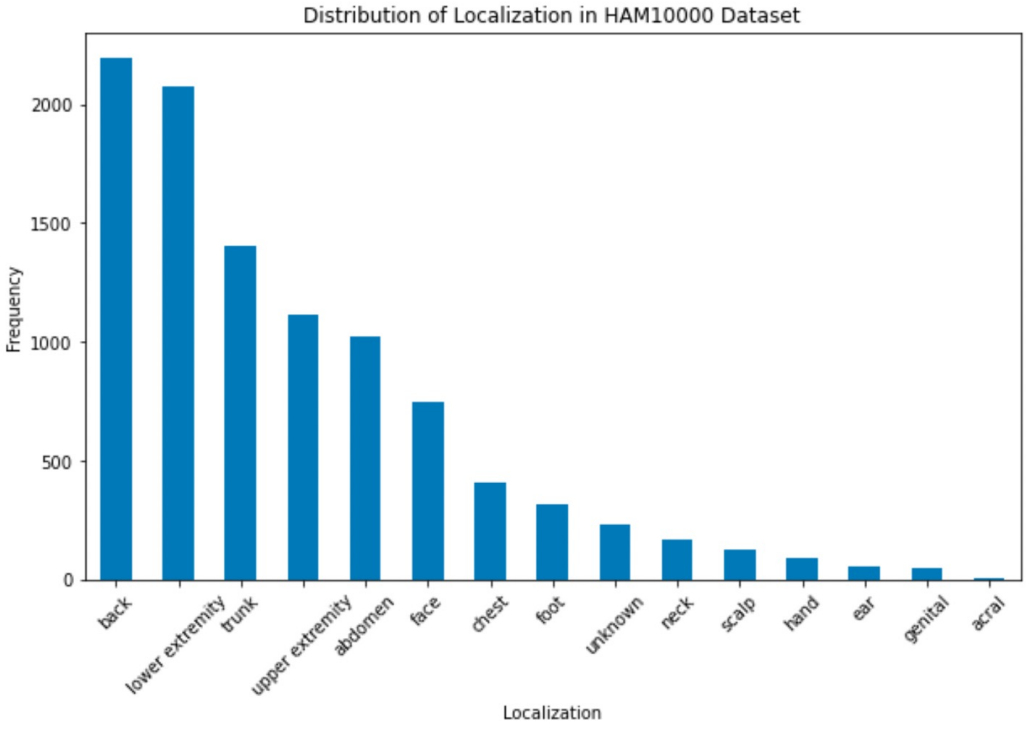

Figure 6: Distribution of Localizations

References

Inthiyaz, Syed, et al. Skin disease detection using deep learning. ScienceDirect, 2023. www.sciencedirect.com/science/article/abs/pii/S0965997822002629.

Jain, Arushi, et al. Multi-type Skin Diseases Classification Using OP-DNN Based Feature Extraction Approach. Springer Link, 12 Jan. 2022. link.springer.com/article/10.1007/s11042-021-11823-x.

Sreekala, Keshetti, et al. Skin Diseases Classification Using Hybrid AI Based Localization Approach. PubMed Central, 2022. www.ncbi.nlm.nih.gov/pmc/articles/PMC9444379/.

Verma, Anurag K., et al. Classification of Skin Disease Using Ensemble Data Mining Techniques. PubMed Central, 2019. www.ncbi.nlm.nih.gov/pmc/articles/PMC7021628/.

Wei, Mingjun, et al. A Skin Disease Classification Model Based on DenseNet and ConvNeXt Fusion. MDPI, 14 Jan. 2023. www.mdpi.com/2079-9292/12/2/438.

Zhang, Bin, et al.Opportunities and Challenges: Classification of Skin Disease Based on Deep Learning. SpringerOpen, 24 Nov. 2021. cjme.springeropen.com/articles/10.1186/s10033-021-00629-5.