

APS360 PROJECT PROPOSAL: SKIN DISEASE DETECTION

Jacky Li

Student# 1011271678

jakkii.li@mail.utoronto.ca

Jordan Cui

Student# 1011026916

jordan.cui@mail.utoronto.ca

Lawrence Ding

Student# 1011439025

larryzm.ding@mail.utoronto.ca

Soham Shorey

Student# 1010845169

soham.shorey@mail.utoronto.ca

—Total Pages: 9

1 INTRODUCTION

This project proposal outlines preliminary plans for a CNN-based machine learning model to detect common and rare skin conditions, from eczema to psoriasis to melanoma.

1.1 THE NEED FOR SKIN DISEASE DETECTION

Skin conditions rank among the most widespread and diverse health concerns in Canada and globally, impacting more than a third of the world’s population (Li et al., 2024). Conditions like acne and psoriasis continue to affect millions of Canadians, while the incidence of more severe diseases like melanoma is steadily rising. According to the Canadian Dermatology Association (CDA), approximately 20% of Canadians experience acne, and 1 million have been diagnosed with psoriasis (Canadian Dermatology Association, 2025b). Additionally, melanoma rates have increased annually by 2.2% in men since 1984 and by 2.0% in women since 1990 (Canadian Dermatology Association, 2025a). The increasing frequency of skin disease strains healthcare systems around the world. In the United States, 84 million Americans consulted a physician for skin disease in 2013—a quarter of their population (American Academy of Dermatology, 2025). A tool that can accurately classify skin diseases from images could alleviate this strain by encouraging people with potentially serious conditions to seek timely medical attention and by assisting physicians in the diagnostic process.

1.2 A MACHINE LEARNING SOLUTION

Machine learning models based on convolutional neural networks (CNNs) are highly effective for image classification tasks due to their ability to extract spatial features like edges, shapes, and textures through trainable filters (Huang, 2023). Skin disease detection is inherently an image classification task, making it a strong candidate for a machine learning approach. Furthermore, CNN-based models that detect skin diseases from images already exist (see Section 3), and artificial intelligence assistance has been shown to increase diagnosis agreement by 10% for dermatologist panels and 12% for nurse practitioners (Liu et al., 2021), reinforcing the viability of machine learning for this application.

2 ILLUSTRATION OF PROJECT PIPELINE

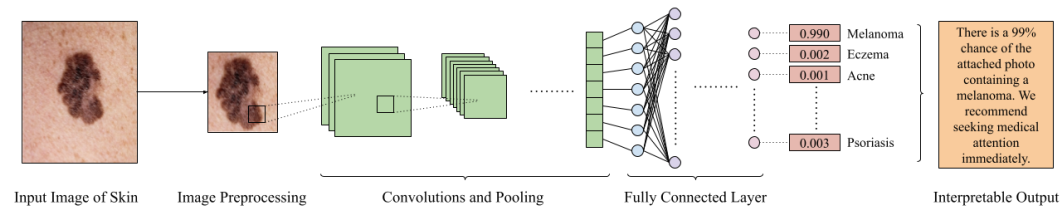


Figure 1: Overall project pipeline. Dermoscopic, clinical, and patient-taken images of skin conditions are extracted from datasets and preprocessed to resize and remove features. These modified images are passed into a pre-trained CNN to extract image features, which are then inputted into a fully connected classification network, which will output the predicted skin disease and next steps for the patient.

3 BACKGROUND & RELATED WORK

Machine learning models for skin disease detection have been widely studied and are now available to consumers through websites and mobile applications.

3.1 DEEP CONVOLUTIONAL NEURAL NETWORK TO DETECT SKIN CANCER

Several studies focus on detecting skin cancers from images. For example, a landmark 2017 study compared the performance of two convolutional neural network (CNN) models—trained on over 120,000 images to distinguish between keratinocyte carcinomas and benign seborrheic keratoses, as well as malignant melanomas and benign nevi—with the diagnostic accuracy of 21 board-certified dermatologists (Esteva et al., 2017).

Unlike earlier models limited to standardized dermoscopy and histology images, this model was trained to handle dermoscopy and smartphone-quality images, improving usability for everyday users. The model is a deep convolutional neural network, specifically Inception v3, designed by Google for large-scale image recognition. The model achieved higher sensitivity and specificity than the average of the 21 dermatologists.

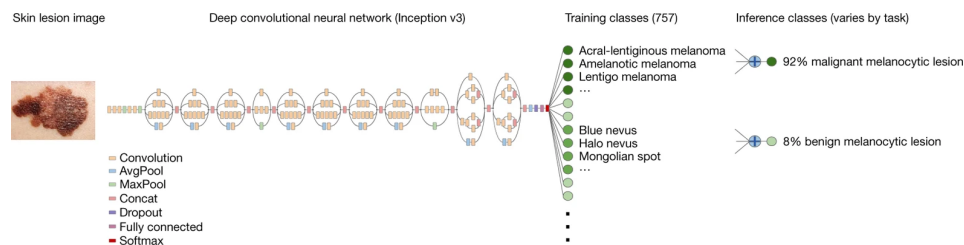


Figure 2: Google Inception v3 skin cancer classification model architecture (Esteva et al., 2017).

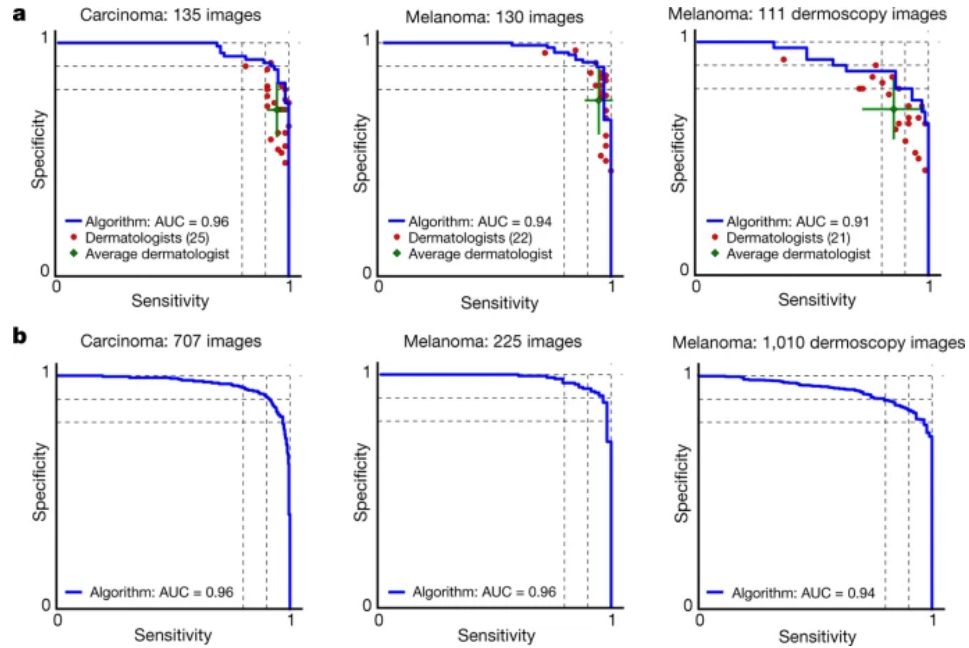


Figure 3: AUC-ROC curve of skin cancer classification performance for Google Inception v3 model (Esteva et al., 2017).

3.2 MICROSOFT RESNET FOR DETECTING SKIN CONDITIONS

Another study used the Microsoft ResNet-152 CNN model, trained on 19,398 images, to assess the probability of 12 distinct skin conditions and to differentiate between benign and malignant cases. The model achieved AUC scores ranging from 0.82 to 0.96, comparable to healthcare professionals (Han et al., 2018).

3.3 VARIATIONS OF RESNET MODELS

Variations of ResNet have been implemented frequently in studies. For example, a 2025 study uses a self-reliant residual network (SR-ResNet) to identify melanoma (Radhika et al., 2025). It uses Zoutendijk's Method, a nonlinear optimizer that selects optimal step size and direction to stabilize convergence and prevent gradient vanishing and overfitting. The model achieved an accuracy of 94.79%, precision of 93.09%, and recall of 96.62%, outperforming traditional deep learning models (Radhika et al., 2025).

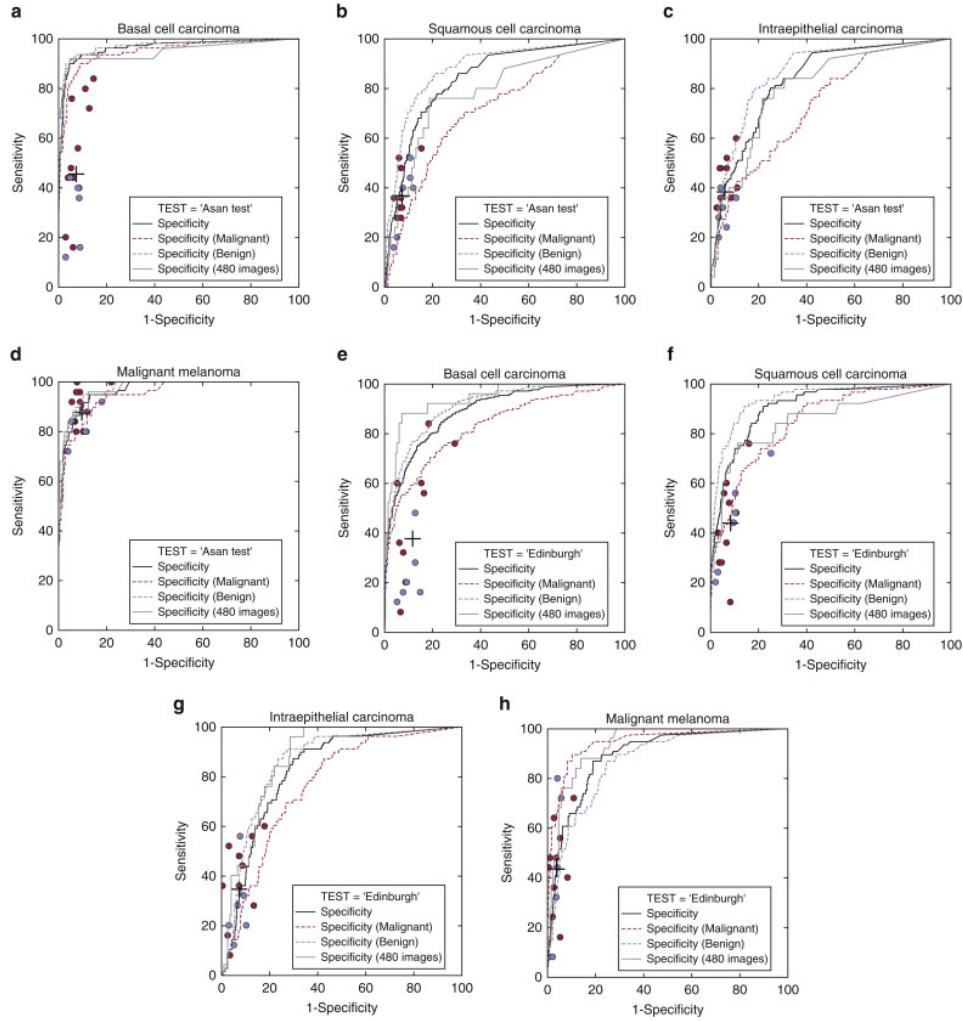


Figure 4: AUC-ROC curves of classification performance of eight skin cancers for SR-ResNet based model (Radhika et al., 2025).

3.4 COMPARISONS BETWEEN VARIOUS CLASSIFICATION MODELS

Beyond skin cancer, studies have used models to classify a wider range of skin diseases. For instance, a 2019 study evaluated five different models for classifying three types of skin diseases: Acne, Lichen Planus, and Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis (SJS-TEN), including a Random Forest (RF), Kernel Support Vector Machine (SVM), and CNN (Bhadula et al., 2019). As before, the CNN was the most successful, achieving high 90s precision and recall scores (Bhadula et al., 2019). Similarly, a 2025 study classifies both skin lesions and cancers using a convolution self-attention network (CSAN), where convolutional layers extract image features, and a Light Gradient Boosting Machine (LGBM) uses decision trees for final classification (Kalpana et al., 2025). Discrete Mycorrhized Optimization (DMS) optimizes hyperparameters like learning rate and number of trees (Kalpana et al., 2025). Interestingly, in addition to removing noise and resizing images in data preprocessing, the researchers also remove hair to highlight skin. These studies underscore the utility of machine learning, particularly CNNs, in skin disease detection.

3.5 WEBSITES AND MOBILE APPLICATIONS

Machine learning models that detect skin conditions are readily available on websites and mobile applications. AI Dermatologist, First Derm, and DermScanAI all allow users to upload images of their

affected skin at differing prices (AI-Derm, 2025; First Derm, 2025; DermScanAI, 2025). DermScanAI specifically uses an EfficientNetB0-based network, predicting nine common skin conditions with an accuracy of 87.2% (Vadivelraju et al., 2025). Smartphones and clinical apps also exist for skin cancer detection. For example, SkinVision detected premalignant and malignant skin lesions with 86.9% sensitivity and 70.4% specificity (Sangers et al., 2022). While accuracy could improve for these models—especially to account for diverse skin tones (see Section 7)—these tools help patients self-assess before visiting a doctor and aid clinicians in diagnosis.

4 DATA PROCESSING

Online databases of common and rare skin conditions involve different imaging techniques. This includes dermoscopic (the most standardised), clinical, and 3D TBP (full body imaging). For the purpose of this project, the dataset must be diverse in the types of imaging technique, which allows the model to generalize. Data augmentation can also be employed to imitate the various imaging techniques. Currently, the objective is to train a model that can classify images taken by medical professionals; however, the objective can be extended to detect images taken by personal cameras, such as the ones in phones.

The following list includes potential datasets that can be used to train the model:

- **HAM10000**: Human Against Machine with 10000 Images (Ma et al., 2018)
 - Public dataset released for the purpose of training deep learning models.
 - Images are collected from different populations and modalities, which allows the model to classify people of all skin colors.
 - The dataset is hosted on Kaggle, which allows for direct download in Python (Kästle, 2018)

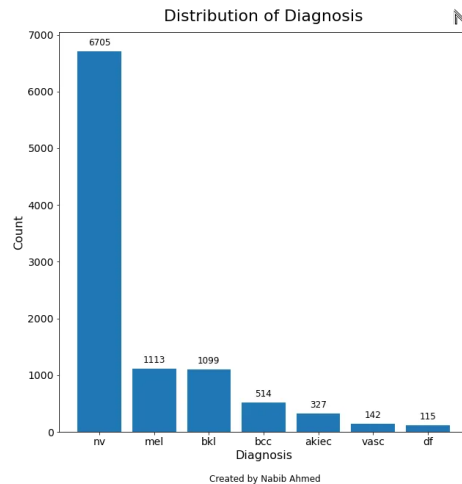


Figure 5: Distribution of skin condition classes in HAM10000 dataset (Kästle, 2018)

- **BCN20000** (Combalia et al., 2019)
 - Contains 19,424 dermoscopic images captured under natural (unconstrained) settings in Hospital Clinic Barcelona
 - Provides generalization of lesions that are bigger than the capture range of the dermoscope.
 - The dataset is publicly available on the ISIC archives (International Skin Imaging Collaboration, 2025)

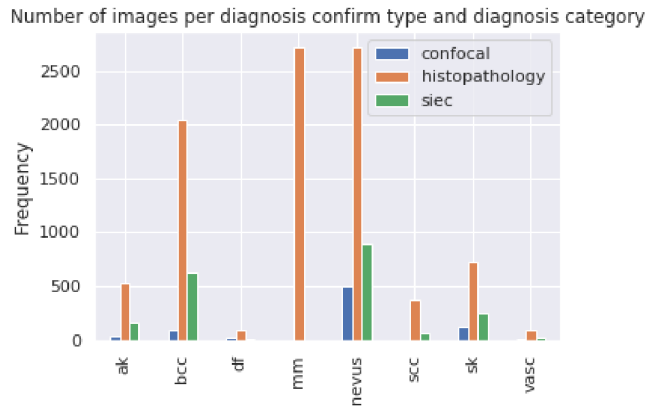


Figure 6: Class distribution and image type in BCN20000 dataset (International Skin Imaging Collaboration, 2025)

- **DERM12345** (Yilmaz et al., 2024)
 - Diverse dataset comprising 12,345 dermatoscopic images with 40 subclasses of skin lesions, collected in Turkey.
 - The dataset consists of skin types from the transition zone between Asia and Europe.
 - The dataset access information is available in the article (Yilmaz et al., 2024)
- **ISIC Challenge Datasets** (Rotemberg et al., 2021)
 - The ISIC challenge dataset consists of skin disease images provided to contestants of the ISIC challenge from 2016 to 2020.
 - The dataset is already sorted into training, validation, and testing, which reduces the workload.
 - Citations and sources of the datasets are available on the ISIC website (Rotemberg et al., 2021).

If required, additional datasets may be incorporated in the future due to the needs of the project. The datasets will be combined and sorted into training, validation, and testing. This will require all of the images to be adjusted to the same size and format. In addition, the labels will have to be standardized, especially given that the different datasets use different classifications for the diseases. Based on the information provided from the background, the standardized label approach will be image segmentation, where a point or polygonal mask is applied over the lesion (region of interest/disease). The color of the mask will determine the class/disease.

As mentioned before, data augmentation will be used to increase training data and help the model to generalize. Techniques such as brightening/dimming, cropping, zooming in/out, distortion, and gaussian noise will be used to imitate the different imaging devices people and clinics may have access to.

5 ARCHITECTURE

The model architecture will be a Convolutional Neural Network (CNN) with multiple convolutional layers. A CNN can extract spatial features, such as edges, textures, and patterns, using learnable

filters, making them ideal for image classification tasks like skin disease detection. Due to computational and time constraints, transfer learning is employed with the pre-trained EfficientNet-B3 or ResNet as the feature extractor, making only the classifier training necessary.

The custom classifier that follows the pre-trained network includes a global average pooling layer to reduce feature dimensionality, multiple fully connected layers using ReLU activation functions, dropout layers to prevent overfitting, and a final output layer with softmax or sigmoid activation depending on the desired output.

The pre-trained backbone will be frozen to train the classifier on the features extracted from the skin images. Once the classifier stabilizes, selected layers of the backbone may be unfrozen for fine-tuning with a reduced learning rate. Cross-entropy loss and Adam will be used as the loss function and optimizer.

6 BASELINE MODEL

A good baseline model to use for comparison is a model that utilizes k-nearest neighbour (KNN) or support vector machine (SVM). Numerous skin disease classifiers have been made using these algorithms, and many of them have achieved strong results with accuracies as high as 98% (Sumithra et al., 2015; Hatem, 2022; Bakheet, 2017; Das et al., 2013). However, these models typically use highly pre-processed images with a small dataset (typically less than 1000 samples). Moreover, the datasets used do not contain images of healthy skin, completely eliminating the possibility of type 1 (false positive) and type 2 (false negative) errors.

These types of models can be replicated using an implementation of KNN or SVM with the expected classifications and feeding in the dataset. Given that the dataset being used is a combination of various datasets whose images may not be ideally processed, the baseline model is expected to not perform as well. This provides a good opportunity to test if a CNN-based machine learning model would be able to perform better in a more realistic setting.

7 ETHICAL CONSIDERATIONS

Machine learning models for skin disease detection raise ethical concerns in both data collection and model use.

Training and validating these models requires labelled images of skin. Datasets that contain private healthcare information like skin images must adhere to patient privacy standards, including patient consent and research ethics board approval. For instance, DermNet, a leading organization for dermatology, requires images not to be identifiable (i.e. no face) and evidence of patient consent before addition to their library (DermNet New Zealand Trust, 2025).

Additionally, existing datasets contain a disproportionately large number of images representing lighter skin tones compared to darker skin tones. This imbalance results from lower skin cancer rates in darker-skinned individuals and unequal access to healthcare (Alipour et al., 2024). It impairs model performance for patients with darker skin, potentially reinforcing pre-existing healthcare disparities across racial and ethnic groups. To mitigate this issue, some skin datasets, such as Skin-Con, include Fitzpatrick skin type labels while others like DDI include equal numbers of each skin type (Daneshjou et al., 2022a;b).

Beyond data collection concerns, a skin disease detection model also risks patients misinterpreting model outputs. For example, if the model incorrectly classifies a mild condition as severe, they may face unnecessary anxiety. Conversely, if the model fails to detect a serious condition, patients may delay or avoid medical advice and treatment, worsening their skin condition. Patients must understand that the model is only an aid—not a replacement for proper diagnosis. To mitigate these risks, models should prioritize high recall to minimize missed detections of serious conditions.

8 PROJECT PLAN

The team's main source of communication will be through a group chat, where discussions happen on a daily basis. All team members live within the GTA, which allows for in-person meetings.

Moving forward, the team will hold weekly meetings online/in-person to discuss project progress and next steps. At the meeting, each team member will be expected to report the tasks they have completed or the progress they have made to accomplish said task. In addition, new tasks may be assigned to team members depending on their availability and progress. All tasks, assignees, and progress will be tracked using monday.com, which is a web-based app for work management.

Figure 6 provides a rough summary of the tasks, assignees, and deadlines outlining the project. Since certain tasks may have the same deadline, priority will be assigned to each task so team members can prioritize more urgent tasks.

APS360 ▾

Main table ... +

New task ▾ 🔍 Search 👤 Person ⚙️ Filter ▾ ⬆️ Sort ⚙️ Hide 📁 Group by ...

▾ To-Do

<input type="checkbox"/>	Task	People	Status	Due date	Priority	+
<input type="checkbox"/>	Project Proposal	👤 +2	Working on it	Jun 13	High	
<input type="checkbox"/>	Dataloader for Datasets	👤	Not Started	Jun 20	Medium	
<input type="checkbox"/>	Download Pre-trained CNN	👤	Not Started	Jun 20	Medium	
<input type="checkbox"/>	Investigate other architect...	👤	Not Started	Jun 24	Low	
<input type="checkbox"/>	Data Augmentation	👤	Not Started	Jun 27	Medium	
<input type="checkbox"/>	Baseline Experiment	👤	Not Started	Jun 27	Low	
<input type="checkbox"/>	Fine-tune on Collected Data	👤	Not Started	Jul 4	Medium	
<input type="checkbox"/>	Performance Analysis	👤	Not Started	Jul 8	Medium	
<input type="checkbox"/>	Progress Report	👤 +2	Not Started	Jul 11	High	
<input type="checkbox"/>	Hyperparameter Tuning	👤	Not Started	Jul 25	Low	
<input type="checkbox"/>	Project Demo	👤 +2	Not Started	Aug 8	High	
<input type="checkbox"/>	Project Presentation	👤 +2	Not Started	Aug 8	High	
<input type="checkbox"/>	Final Report	👤 +2	Not Started	Aug 15	High	
<input type="checkbox"/>	+ Add task					

Jun 13 - Aug ...

▾ Completed

<input type="checkbox"/>	Task	People	Status	Due date	Priority	+
<input type="checkbox"/>	Project Topic	👤 +2	Done	Jun-5	Medium	
<input type="checkbox"/>	Background Research	👤	Done	Jun-8	Low	
<input type="checkbox"/>	Research Datasets	👤	Done	Jun-11	Medium	
<input type="checkbox"/>	Research Models	👤	Done	Jun-11	Medium	
<input type="checkbox"/>	Download Dataset (Statisti...	👤	Done	Jun-12	High	
<input type="checkbox"/>	Setup Github	👤	Done	Jun-12	Medium	
<input type="checkbox"/>	+ Add task					

Jun 5 - 12

Figure 7: Project task outline created using monday.com

The team will collaborate in their work on Github, while training may be done on Google Colab due to the lack of GPUs. Project files, such as Python scripts and Jupyter notebooks will be stored in the team's public Github repository. New branches will be created for specific tasks to avoid overwriting existing code. Once a task is complete, developers may open pull requests, then merge the completed code into master after code reviews.

9 RISK REGISTER

Potential risks for the project are listed below.

9.1 TEAM MEMBERS DROPPING THE COURSE

After a team discussion, we determined that the likelihood of a team member dropping the course is small, however, non-zero. In the case that a team member does drop the course, it is important that we have properly documented each stage of the project and explicitly parceled out tasks. In this way, the remaining team members can easily take over the responsibility of the team member who dropped the course. The general outline for this has been given in the Project Plan section and we will continue in a similar manner on a shared Google Doc and in our team Discord as the project progresses. Even without the risk of a team member dropping, proper documentation and task assignment is still of paramount importance to working efficiently as a team.

9.2 BUSY WORK SCHEDULES

Each of our team members have other commitments, for example, we are all involved in summer research programs. This can lead to conflicts with meeting times and work schedules. This issue is very likely to arise, in fact, in the process of working on this project proposal it has already become apparent. When encountering this issue, it is important that we maintain effective communication and work together around each other's schedules. As a team, we have already implemented this strategy and seen success from it. Team members communicated their availability for working on the project proposal and we coordinated tasks accordingly, allowing us to make steady progress despite our conflicting commitments.

9.3 UNEXPECTED TRAINING TIME

A potential risk we foresee is that training time may be longer than expected, which could result in missing important deadlines. This is a fairly likely scenario, given that training times are highly variable. However, this risk can be effectively mitigated with proactive planning. To address this, we will train a baseline model early in the process to establish a rough estimate of how long model training typically takes in our environment. This initial benchmark will help us better schedule future training sessions and avoid last-minute surprises. Additionally, we plan to schedule final model training at least one full day in advance of any deadlines to allow for grace time in case of unexpected delays or errors. This buffer will give us room to troubleshoot issues without compromising our timeline.

9.4 POOR MODEL PERFORMANCE

Due to the variety of skin diseases in our datasets, we find it fairly likely that our model fails to classify images with an acceptable accuracy (85+%) and recall, which is particularly important for medical diagnosis tools which should not miss positive cases. To minimize this risk, we will train the model with a combination of multiple datasets including the datasets mentioned in Section 4, use transfer learning to prevent overfitting, remove hair from images in preprocessing, and add extra noise to data. If the model has difficulty predicting minority classes, we can increase the weighting of minority classes in the training or generate artificial examples. If these steps are still insufficient, we can limit our model to predict skin conditions with greater data availability like eczema, acne, and melanoma.

10 LINK TO COLAB/GITHUB

Github Repository containing all source code is publicly available at <https://github.com/jakkiiri/APS360-Project>

REFERENCES

- AI-Derm. Ai-dermatologist: Skin scanner. <https://ai-derm.com/>, 2025. Accessed: Jun. 8, 2025.
- N. Alipour, T. Burke, and J. Courtney. Skin type diversity in skin lesion datasets: A review. *Current Dermatology Reports*, 13:198–210, Aug. 2024. doi: 10.1007/s13671-024-00440-0.
- American Academy of Dermatology. Burden of skin disease. <https://www.aad.org/member/clinical-quality/clinical-care/bsd>, 2025. Accessed: Jun. 7, 2025.
- S. Bakheet. An svm framework for malignant melanoma detection based on optimized hog features. *Computation*, 5(1):4, 2017. doi: 10.3390/computation5010004.
- S. Bhadula, S. Sharma, P. Juyal, and C. Kulshrestha. Machine learning algorithms based skin disease detection. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(2):4044–4049, 2019.
- Canadian Dermatology Association. Melanoma. <https://dermatology.ca/public-patients/diseases-conditions/skin-cancer/melanoma/>, 2025a. Accessed: Jun. 7, 2025.
- Canadian Dermatology Association. Skin conditions. <https://dermatology.ca/public-patients/diseases-conditions/skin-conditions/>, 2025b. Accessed: Jun. 7, 2025.
- M. Combalia et al. Bcn20000: Dermoscopic lesions in the wild. *arXiv preprint arXiv:1908.02288*, Aug. 2019. doi: 10.48550/arXiv.1908.02288.
- R. Daneshjou, M. Yuksekgonul, Z. R. Cai, R. Novoa, and J. Y. Zou. Skincon: A skin disease dataset densely annotated by domain experts for fine-grained debugging and analysis. In *Advances in Neural Information Processing Systems*, volume 35, pp. 18157–18167, Dec. 2022a.
- R. Daneshjou et al. Disparities in dermatology ai performance on a diverse, curated clinical image set. *Science Advances*, 8(32):abq6147, Aug. 2022b. doi: 10.1126/sciadv.abq6147.
- N. Das, A. Pal, S. Mazumder, S. Sarkar, D. Gangopadhyay, and M. Nasipuri. An svm based skin disease identification using local binary patterns. In *2013 Third International Conference on Advances in Computing and Communications*, pp. 208–211, Cochin, India, 2013. doi: 10.1109/ICACC.2013.48.
- DermascanAI. Dermascanai: Ai skin analysis tool. <https://dermascanai.com/#analysis>, 2025. Accessed: Jun. 8, 2025.
- DermNet New Zealand Trust. Clinical images. <https://dermnetnz.org/cme/teledermatology-skin-cancer/clinical-images>, 2025.
- A. Esteva, B. Kuprel, R. Novoa, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542:115–118, 2017. doi: 10.1038/nature21056.
- First Derm. Ai dermatology and teledermatology services. <https://www.firstderm.com/ai-dermatology/>, 2025. Accessed: Jun. 8, 2025.
- S. S. Han et al. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, 138(7):1529–1538, 2018.
- M. Q. Hatem. Skin lesion classification system using a k-nearest neighbor algorithm. *Visual Computing for Industry, Biomedicine, and Art*, 5(7), 2022. doi: 10.1186/s42492-022-00103-6.
- M. Huang. Image recognition with cnns: improving accuracy and efficiency. Medium, Feb. 2023. Accessed: Jun. 7, 2025.
- International Skin Imaging Collaboration. Bcn20000: Dermoscopic lesions in the wild. ISIC Archive, 2025. Dataset.

- B. Kalpana, A. Senthilselvi, S. Muruganandam, and S. V. Kumar. Enhancing skin disease diagnosis: Light gbm-dms algorithm for accurate image classification. *Cognitive Computation*, 17(3):1–18, 2025.
- M. Kästle. Skin cancer mnist: Ham10000. Kaggle, 2018. Dataset.
- Q. Li, M. T. Patrick, S. Sreeskandarajan, J. Kang, J. M. Kahlenberg, J. E. Gudjonsson, Z. He, and L. C. Tsoi. Large-scale epidemiological analysis of common skin diseases to identify shared and unique comorbidities and demographic factors. *Frontiers in Immunology*, 14, Jan. 2024. doi: 10.3389/fimmu.2024.1344061.
- Y. Liu et al. Development and assessment of an artificial intelligence–based tool for skin condition diagnosis by primary care physicians and nurse practitioners in teledermatology practices. *JAMA Dermatology*, Apr. 2021. doi: 10.1001/jamadermatol.2020.5217.
- A. Ma, Y. Zhang, and B. Wang. Machine-readable metadata of biological sample data in the ncbi biosample database. *Scientific Data*, 5:180161, 2018. doi: 10.1038/sdata.2018.161.
- V. Radhika, A. Muthuchudar, and M. Lingaraj. Self-reliant residual network based deep learning framework for melanoma skin disease detection. *Journal of Theoretical and Applied Information Technology*, 103(10), May 2025.
- V. Rotemberg et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific Data*, 8(34):1–15, 2021. doi: 10.1038/s41597-021-00815-z.
- T. Sangers, S. Reeder, S. van der Vet, S. Jhingoer, A. Mooyaart, D. M. Siegel, T. Nijsten, and M. Wakkee. Validation of a market-approved artificial intelligence mobile health app for skin cancer screening: A prospective multicenter diagnostic accuracy study. *Dermatology*, 238(4): 649–656, 2022.
- R. Sumithra, M. Suhil, and D. S. Guru. Segmentation and classification of skin lesions for disease diagnosis. *Procedia Computer Science*, 45:76–85, 2015. doi: 10.1016/j.procs.2015.03.090.
- C. Vadivelraju, K. Elisheba, V. M. Bharath Kumar, P. Suraj, and M. H. Reddy. Derascan ai: Deep learning system for preliminary diagnosis of dermatological manifestations. *International Journal of Creative Research Thoughts (IJCRT)*, 13(5):406–413, May 2025.
- A. Yilmaz et al. Derm12345: A large, multisource dermatoscopic skin lesion dataset with 40 sub-classes. *Scientific Data*, 11:1302, 2024. doi: 10.1038/s41597-024-04104-3.