SKIN CANCER CLASSIFICATION USING DEEP LEARNING MODELS

Vishwas Puri
Student# 1010174669
vishwas.puri@mail.utoronto.ca

Rayan Ahsan
Student# 1010140820
rayan.ahsan@mail.utoronto.ca

Adyan Hossain
Student# 1010150373
adyan.hossain@mail.utoronto.ca

Kashan Ahmed
Student# 1010090072
kashan.ahmad@mail.utoronto.ca

ABSTRACT

This document outlines Group 29's proposed final project for APS360. The goal of the project is to design a deep learning model that classifies various types of skin lesions. The proposal discusses the motivation behind addressing skin cancer classification, the dataset being used, our project's architecture, and project management. This document follows the structure and expectations outlined in the "APS360 Project Proposal Handout and Rubric". —-Total Pages: 7

1 INTRODUCTION

Skin cancer is one of the most common malignancies in the world, with approximately 80,000 cases diagnosed annually in Canada Canadian Skin Cancer Foundation (2025). Early and accurate skin cancer recognition significantly improves patient outcomes, yielding a five-year survival rate of 99% Skin Cancer Foundation (2025). However, limited dermatologist access, particularly in rural regions, highlights a persistent gap in timely diagnosis. Furthermore, the most common method of visual examination by dermatologists yields an accuracy of approximately 60% Marks (2002). In this context, ML-based systems hold significant potential to support dermatologists in the diagnostic process for skin cancer Pathan et al. (2022). This project proposes a deep-learning model for classifying dermoscopic images sourced from the HAM10000 dataset into seven skin lesion categories Codella et al. (2021). Deep learning, specifically convolutional neural networks (CNNs), are well-suited due to their ability to learn complex and hierarchical representations directly from raw images. Compared to other traditional models, such as multinomial logistic regression, CNNs are a justified selection for this visual recognition task because their layered architecture enables them to capture both low-level and high-level visual features, such as colour variation and border irregularity which are prominent visual cues commonly used in dermatological diagnosis Al-Antari et al. (2024). Ultimately, this project is a foundational step toward developing end-to-end deployable applications that can be integrated into clinical workflows to assist dermatologists in improving diagnostic process efficiency.

2 ILLUSTRATION

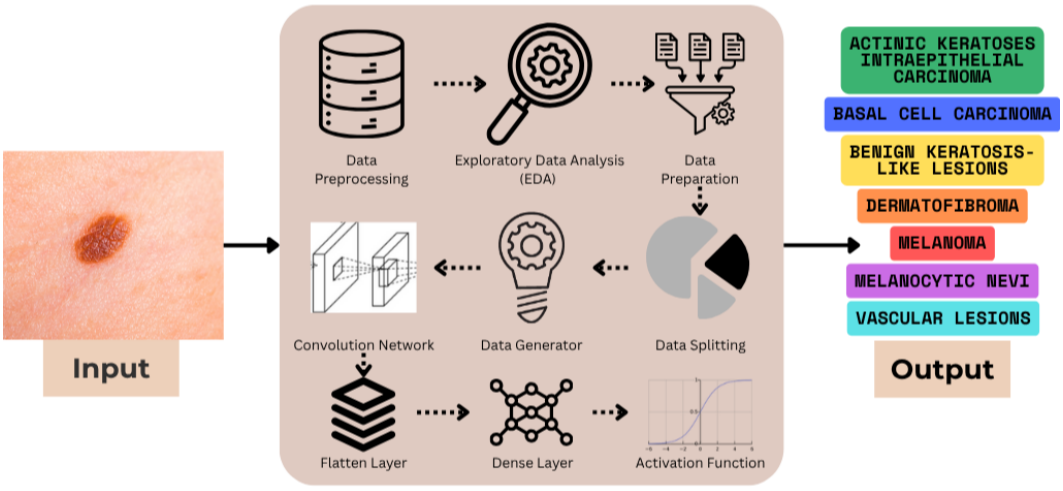


Figure 1: Flowchart displaying our Project's pipeline

3 BACKGROUND RELATED WORK

This background section provides context on related works that influenced our project's pipeline and architecture.

3.1 ENHANCED SKIN CANCER DIAGNOSIS USING OPTIMIZED CNN ARCHITECTURE AND CHECKPOINTS FOR AUTOMATED DERMATOLOGICAL LESION CLASSIFICATION

This recent study proposed a hybrid classification framework integrating CNNs with XGBoost for classifying dermoscopic images from the HAM10000 dataset. The data preprocessing and cleaning involved artifact removal and normalization of the images. The hybrid model achieved an accuracy of 97.86%, precision and recall of 97.9%, and an F1 score of 97.8%, demonstrating strong performance across multiple classification metrics. The authors concluded that while their model outperformed traditional approaches, the limited size and diversity of the dataset used limited how well the model generalizes to different populations Al-Antari et al. (2024).

3.2 SKIN LESION CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK WITH NOVEL REGULARIZER

This paper proposed a CNN binary classifier to differentiate between malignant and benign skin lesions using the ISIC dataset. The architecture consisted of two convolutional layers followed by pooling, dropout, flattening, and a fully connected layer. A novel regularization method based on the weight matrix's standard deviation was applied to improve generalization, with an optimal regularization parameter of 0.02. The model achieved a maximum average accuracy of 97.49% and demonstrated strong classification ability with good AUC scores. The authors identified the computationally intensive tuning process for the regularization parameter as a limitation in optimization Harangi (2019).

3.3 SKIN LESION CLASSIFICATION USING HYBRID DEEP NEURAL NETWORKS

This study classified skin lesions by combining deep features extracted from AlexNet, VGG16, and ResNet-18, employing an SVM classifier on the ISIC 2017 validation set. This approach yielded AUCs of 83.83% for melanoma and 97.55% for seborrheic keratosis, outperforming individual models and demonstrating improved robustness through the integration of complementary representations across different CNN architectures. Future work was suggested to explore deeper CNNs like DenseNet and the use of localized image patches Esteva et al. (2017).

3.4 A COMPREHENSIVE STUDY ON SKIN CANCER DETECTION USING ARTIFICIAL NEURAL NETWORK (ANN) AND CONVOLUTIONAL NEURAL NETWORK (CNN)

This study emphasized the superiority of CNNs over traditional ANNs in identifying features critical to skin lesion classification. A hybrid system combining CNNs with NLP achieved 99.35% on training accuracy; however, test accuracy ranged between 66–83%, revealing robustness concerns. The authors highlighted key challenges pertaining to the datasets being inclusive of diverse lesion types and skin tones Brinker et al. (2019).

3.5 AUTOMATED SKIN LESION CLASSIFICATION USING ENSEMBLE OF DEEP NEURAL NETWORKS IN ISIC 2018: SKIN LESION ANALYSIS TOWARDS MELANOMA DETECTION CHALLENGE

This paper evaluated several advanced CNN architectures on the ISIC 2018 dataset, with PNASNet-5-Large achieving a notable validation score of 0.76. The study emphasized the efficiency of CNNs in skin lesion classification while noting the role of hyperparameters optimization and diverse datasets in improving the generalizability of these models Tajbakhsh et al. (2024).

4 DATA PROCESSING

The dataset used for this project is HAM10000 Codella et al. (2021), which consists of 10,015 dermoscopic images of pigmented skin lesions, categorized into seven diagnostic classes: melanoma, melanocytic nevi, basal cell carcinoma, actinic keratoses, benign keratosis-like lesions, dermatofibroma, and vascular lesions. Each image is accompanied by metadata, including lesion identifiers and corresponding diagnosis labels. This dataset serves as the foundation for training a deep learning model capable of performing multi-class classification. The preprocessing pipeline begins with merging and cleaning the metadata. Entries with missing or ambiguous labels are removed, and duplicate lesion IDs are filtered out to prevent data leakage between training and evaluation phases. Labels are then standardized to ensure consistent class representation across the dataset. As shown in Figure 2, the dataset is significantly imbalanced, with certain lesion types—such as melanocytic nevi—dominating the sample count, while others, like dermatofibroma or vascular lesions, are underrepresented. To address this, data augmentation techniques are applied to the training set. Common transformations, including horizontal flips, rotations, and zooming, are used to synthetically expand underrepresented classes and help the model generalize better. Finally, the dataset is partitioned into training, validation, and test sets using stratified sampling to preserve class distribution across splits. Approximately 80% of the data is used for training, 10% for validation, and 10% for final testing. An overview of the complete preprocessing pipeline is illustrated in Figure 1.

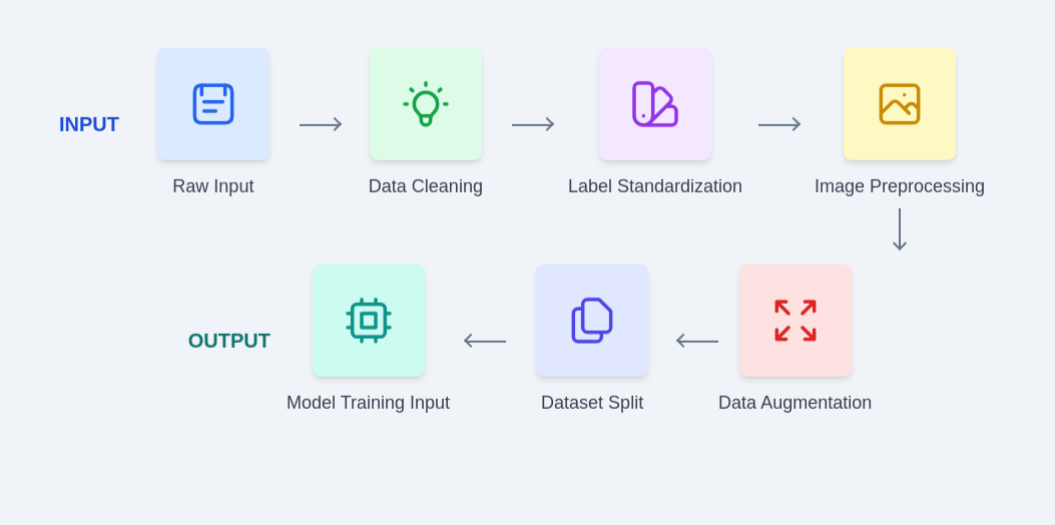


Figure 2: Display of preprocessing pipeline

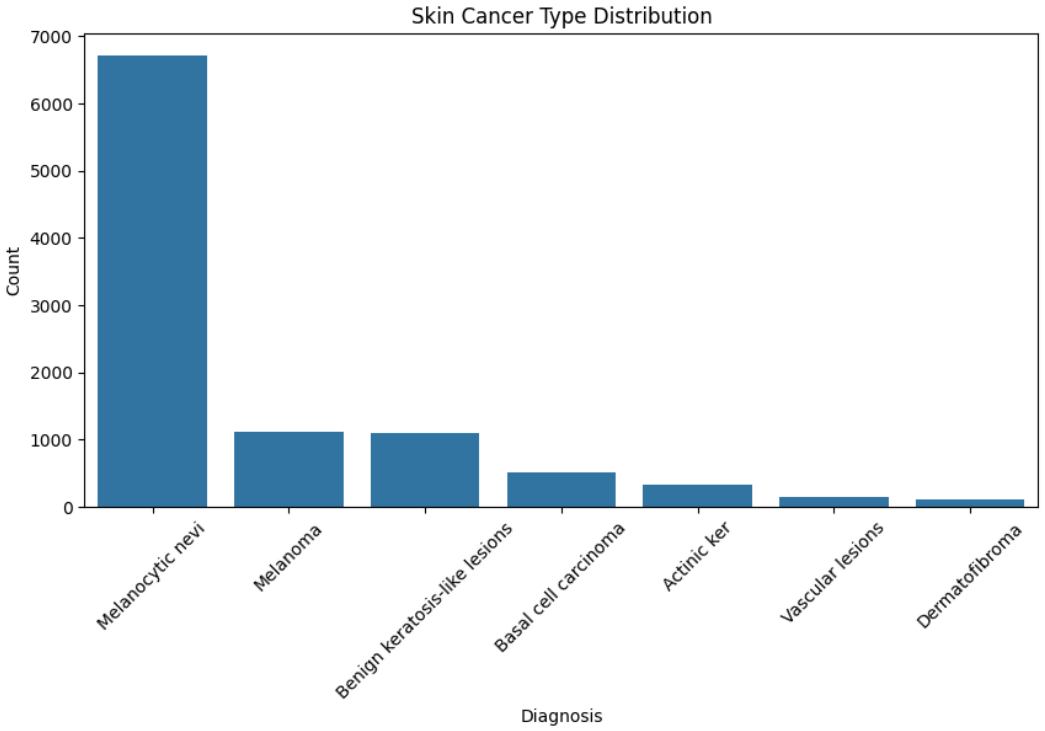


Figure 3: Bar Chart displaying unbalanced distribution of classification types

5 Architecture

We will use a pre-trained convolutional neural network (CNN) with an encoder decoder architecture to classify dermoscopic images into seven distinct skin lesion categories. The encoder consists of multiple convolutional layers, each followed by ReLU activation and max pooling. These layers are

stacked sequentially, allowing the model to build increasingly complex feature representations. This enables it to learn filters (kernels) to extract visual patterns such as texture, asymmetry, and color irregularity through a process known as feature learning. The decoder performs the classification task. It begins by flattening the extracted feature maps into a one dimensional vector, which is passed through fully connected layers. A softmax activation function at the output generates a probability distribution over the lesion classes. The model is trained using the CrossEntropyLoss function, and its parameters are updated during backpropagation using the Adam optimizer to ensure efficient convergence. To improve performance, we will tune key hyperparameters including the learning rate, batch size, and number of training epochs. Dropout regularization will also be applied to reduce overfitting and help the model learn more generalizable features.

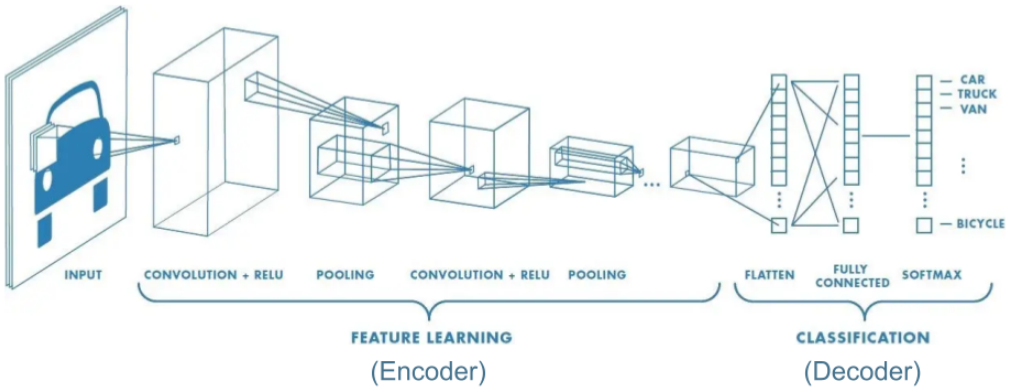


Figure 4: Display of our deep learning model's architecture

6 Baseline Model

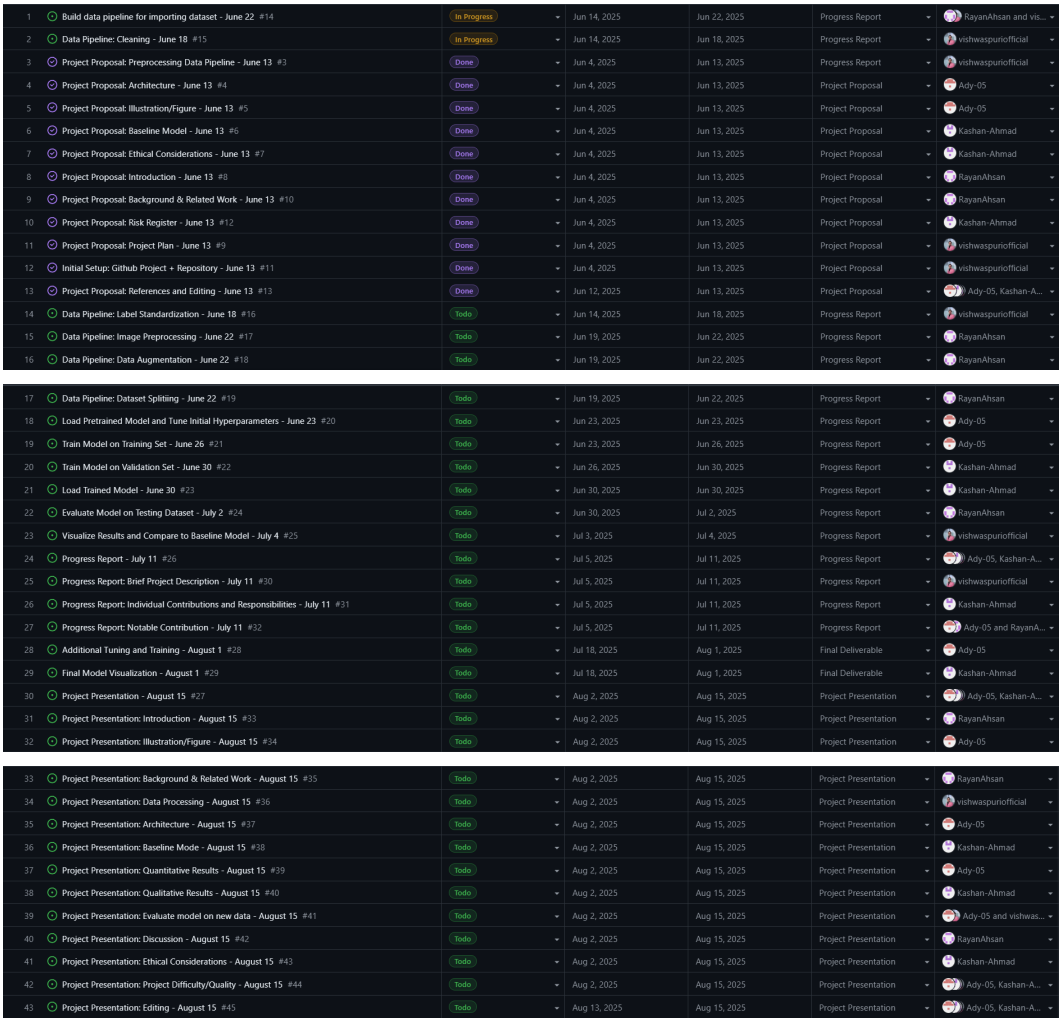
As a baseline for our convolutional neural network, we will implement a Random Forest classifier using flattened grayscale pixel intensities as input features. Each image will be resized to 64×64 pixels to reduce dimensionality and then converted to grayscale. The resulting 2D array will be flattened into a 1D vector of 4096 pixel intensity values, which will serve as the input feature vector for the model. This approach provides a simple and consistent way to represent each image numerically without requiring learned feature extraction. We will use the RandomForestClassifier from the scikit-learn library, with 100 decision trees and all other parameters set to their default values. The model will be trained and evaluated on the same 80-10-10 train-validation-test split used for our CNN. Performance will be measured using accuracy, precision, recall, and F1 score. Random Forests are well-suited for multiclass classification problems like ours—handling seven output classes natively by aggregating predictions across all trees. Their effectiveness in medical image classification has been supported by studies who successfully used Random Forests to classify skin lesions with competitive accuracy. Ali et al. (2023).

7 Ethical Considerations

Skin cancer classification could present several ethical challenges, particularly around fairness and potential harm from misclassification. A major concern is that commonly used datasets, such as the HAM10000, lack diversity in skin tone representation, with very few examples of dark skinned patients MyHealth Alberta (2025). This may result in a model that is poor in classifying underrepresented groups, thereby inflicting the risk of misclassification. A well-trained model will not always provide accurate results. Identifying the wrong skin cancer type could be harmful for patients, as an incorrect diagnosis could result in unnecessary treatment, or failure to treat a serious condition in time. These challenges highlight the importance of evaluating model performance across diverse groups, and clearly communicating the model's limitations when interpreting results.

8 PROJECT PLAN

To accommodate different schedules, time zones, and personal commitments, we will continue to hold weekly meetings on Discord to discuss progress, address challenges, and plan upcoming tasks. These meetings will help us stay aligned and provide opportunities for real-time problem-solving. For day-to-day communication, we will use our Discord group chat, where team members are encouraged to ask questions, share resources, and post updates. All collaborative writing and documentation will be done through Google Docs, while code development and version control will be managed through GitHub. To prevent conflicts, each new feature or bug fix will be developed on a separate branch, with changes submitted via documented pull requests. We will follow branch naming conventions and PR reviews to ensure clarity and maintain code quality. All project management will be handled through GitHub Projects. We will use assignees, labels (e.g., bug, feature, documentation), milestones, and issues, organized within a Kanban board to track progress for each team member. This system will guide branch creation and streamline the pull request process, while also making it easier to monitor task status and ensure accountability. If any team member falls behind, we will reassign tasks as needed to maintain balanced workloads and ensure the project stays on track. Regular communication and clear deadlines will support effective integration of all contributions and smooth project execution.



1	Build data pipeline for importing dataset - June 22 #14	In Progress	Jun 14, 2025	Jun 22, 2025	Progress Report	RayanAhsan and vis...
2	Data Pipeline: Cleaning - June 18 #15	In Progress	Jun 14, 2025	Jun 18, 2025	Progress Report	vishwaspuriofficial
3	Project Proposal: Preprocessing Data Pipeline - June 13 #3	Done	Jun 4, 2025	Jun 13, 2025	Progress Report	vishwaspuriofficial
4	Project Proposal: Architecture - June 13 #4	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	Ady-05
5	Project Proposal: Illustration/Figure - June 13 #5	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	Ady-05
6	Project Proposal: Baseline Model - June 13 #6	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	Kashan-Ahmad
7	Project Proposal: Ethical Considerations - June 13 #7	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	Kashan-Ahmad
8	Project Proposal: Introduction - June 13 #8	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	RayanAhsan
9	Project Proposal: Background & Related Work - June 13 #10	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	RayanAhsan
10	Project Proposal: Risk Register - June 13 #12	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	Kashan-Ahmad
11	Project Proposal: Project Plan - June 13 #9	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	vishwaspuriofficial
12	Initial Setup: Github Project + Repository - June 13 #11	Done	Jun 4, 2025	Jun 13, 2025	Project Proposal	vishwaspuriofficial
13	Project Proposal: References and Editing - June 13 #13	Done	Jun 12, 2025	Jun 13, 2025	Project Proposal	Ady-05, Kashan-A...
14	Data Pipeline: Label Standardization - June 18 #16	Todo	Jun 14, 2025	Jun 18, 2025	Progress Report	vishwaspuriofficial
15	Data Pipeline: Image Preprocessing - June 22 #17	Todo	Jun 19, 2025	Jun 22, 2025	Progress Report	RayanAhsan
16	Data Pipeline: Data Augmentation - June 22 #18	Todo	Jun 19, 2025	Jun 22, 2025	Progress Report	RayanAhsan
17	Data Pipeline: Dataset Splitiing - June 22 #19	Todo	Jun 19, 2025	Jun 22, 2025	Progress Report	RayanAhsan
18	Load Pretrained Model and Tune Initial Hyperparameters - June 23 #20	Todo	Jun 23, 2025	Jun 23, 2025	Progress Report	Ady-05
19	Train Model on Training Set - June 26 #21	Todo	Jun 23, 2025	Jun 26, 2025	Progress Report	Ady-05
20	Train Model on Validation Set - June 30 #22	Todo	Jun 26, 2025	Jun 30, 2025	Progress Report	Kashan-Ahmad
21	Load Trained Model - June 30 #23	Todo	Jun 30, 2025	Jun 30, 2025	Progress Report	Kashan-Ahmad
22	Evaluate Model on Testing Dataset - July 2 #24	Todo	Jun 30, 2025	Jul 2, 2025	Progress Report	RayanAhsan
23	Visualize Results and Compare to Baseline Model - July 4 #25	Todo	Jul 3, 2025	Jul 4, 2025	Progress Report	vishwaspuriofficial
24	Progress Report - July 11 #26	Todo	Jul 5, 2025	Jul 11, 2025	Progress Report	Ady-05, Kashan-A...
25	Progress Report: Brief Project Description - July 11 #30	Todo	Jul 5, 2025	Jul 11, 2025	Progress Report	vishwaspuriofficial
26	Progress Report: Individual Contributions and Responsibilities - July 11 #31	Todo	Jul 5, 2025	Jul 11, 2025	Progress Report	Kashan-Ahmad
27	Progress Report: Notable Contribution - July 11 #32	Todo	Jul 5, 2025	Jul 11, 2025	Progress Report	Ady-05 and RayanA...
28	Additional Tuning and Training - August 1 #28	Todo	Jul 18, 2025	Aug 1, 2025	Final Deliverable	Ady-05
29	Final Model Visualization - August 1 #29	Todo	Jul 18, 2025	Aug 1, 2025	Final Deliverable	Kashan-Ahmad
30	Project Presentation - August 15 #27	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Ady-05, Kashan-A...
31	Project Presentation: Introduction - August 15 #33	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	RayanAhsan
32	Project Presentation: Illustration/Figure - August 15 #34	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Ady-05
33	Project Presentation: Background & Related Work - August 15 #35	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	RayanAhsan
34	Project Presentation: Data Processing - August 15 #36	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	vishwaspuriofficial
35	Project Presentation: Architecture - August 15 #37	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Ady-05
36	Project Presentation: Baseline Mode - August 15 #38	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Kashan-Ahmad
37	Project Presentation: Quantitative Results - August 15 #39	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Ady-05
38	Project Presentation: Qualitative Results - August 15 #40	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Kashan-Ahmad
39	Project Presentation: Evaluate model on new data - August 15 #41	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Ady-05 and vishwas...
40	Project Presentation: Discussion - August 15 #42	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	RayanAhsan
41	Project Presentation: Ethical Considerations - August 15 #43	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Kashan-Ahmad
42	Project Presentation: Project Difficulty/Quality - August 15 #44	Todo	Aug 2, 2025	Aug 15, 2025	Project Presentation	Ady-05, Kashan-A...
43	Project Presentation: Editing - August 15 #45	Todo	Aug 13, 2025	Aug 15, 2025	Project Presentation	Ady-05, Kashan-A...

Figure 5: Github Projects task board displaying our project plan

9 RISK REGISTER

Risk	Likelihood	Solution
Course dropped by a team member.	$< 5\%$	• Reassign responsibilities among team members. • Encourage open communication within the team so members can ask for help when needed. • Communicate with the teaching team for any additional support.
Busy schedules due to work, family commitments and different time zones.	$> 80\%$	• Schedule regular check-ins at mutually agreeable times. • Utilize asynchronous collaboration tools such as Google Docs and Discord to keep communication flowing. • Assign tasks with deadlines so progress can continue independently when needed.
Disagreement on technical decisions, task division, or design choice.	$> 80\%$	• Document decisions to avoid confusion. • Listen to all decision explanations. • Call vote to solve the disagreement.
The model takes too long to train.	50%	• Edit and update hyperparameters. • Purchase Google Colab Pro if needed.
Low model accuracy.	30%	• Extensive tuning of hyperparameters. • If time permits, experiment with different models.

Table 1: Risk register for the project, including likelihood and mitigation strategies.

10 LINK TO GITHUB REPOSITORY

The following is an open source repository for the project: https://github.com/vishwaspuriofficial/APS360-Project

REFERENCES

Mohammad A. Al-Antari et al. Explainable deep learning model for melanoma classification using dermoscopy images. *BMC Medical Imaging*, 24(1), 2024. doi: 10.1186/s12880-024-01356-8. URL https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-024-01356-8.

Shoaib Ali et al. Explainable ai for medical imaging: A survey. *Patterns*, 4(1), 2023. doi: 10.1016/j.patter.2022.100520. URL https://www.sciencedirect.com/science/article/pii/S2588914123000205.

Titus Josef Brinker et al. Skin cancer classification using convolutional neural networks: systematic review, 2019. URL https://arxiv.org/pdf/1901.10802.

Canadian Skin Cancer Foundation. Skin cancer facts, 2025. URL https://www.canadianskincancerfoundation.com/skin-cancer/.

Noel C. Codella et al. Skin lesion analysis towards melanoma detection, 2021. URL https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T. Harvard Dataverse.

Andre Esteva et al. Dermatologist-level classification of skin cancer with deep neural networks, 2017. URL https://arxiv.org/pdf/1702.08434.

Balazs Harangi. Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of Biomedical and Health Informatics*, 2019. doi: 10.1109/JBHI.2019.2899327. URL https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8669763.

R. Marks. An overview of skin cancers: incidence and causation. *The Lancet Oncology*, 3(4):165–169, 2002. doi: 10.1016/S1470-2045(02)00679-4. URL https://www.thelancet.com/journals/lanonc/article/PIIS1470204502006794/abstract.

MyHealth Alberta. Signs of melanoma, 2025. URL https://myhealth.alberta.ca/Health/pages/conditions.aspx?hwid=hw206547.

Shakir Pathan et al. Melanoma skin cancer detection using deep learning and cnn. *Journal of Healthcare Engineering*, 2022. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9693628/.

Skin Cancer Foundation. Skin cancer facts, 2025. URL https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/.

Nima Tajbakhsh et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? *Journal of Visual Communication and Image Representation*, 2024. doi: 10.1002/jvc2.477. URL https://onlinelibrary.wiley.com/doi/full/10.1002/jvc2.477.