Detecting melanoma using Deep Learning

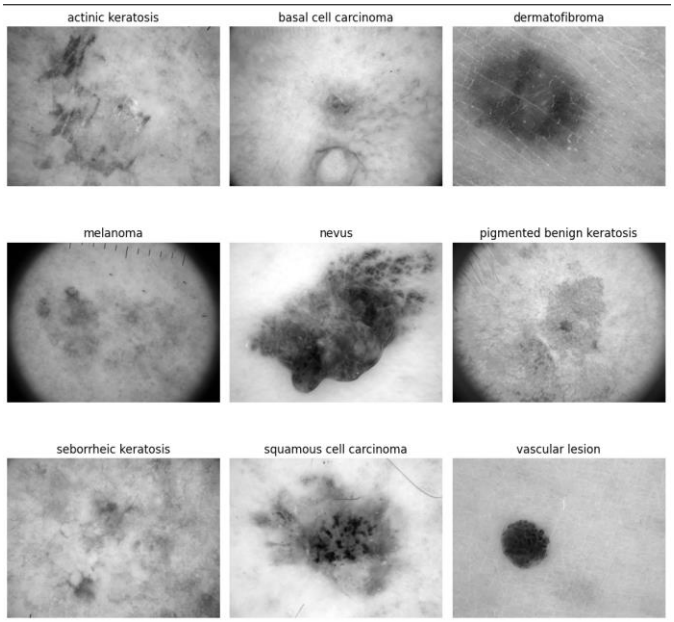
Motivation

There is still a limited application of technology in the process of identifying skin cancer indicators e.g. melanoma. If Deep Learning models can be utilized to detect signs, it would allow patients to seek out medical assistance early. This would allow early-stage skin cancer to be treated more properly. Additionally given the limited resources available in hospitals, more people could be treated more quickly if medical professionals also utilize this technology. Once skin cancer symptoms are verified by professionals, it can be communicated to the patient at a much quicker rate. This modeling could also reduce treatment costs for patients and make healthcare facilities more accessible. However, the main aim is to improve or fasten the process of detecting melanoma. Around 40 and 50% of uveal melanoma remain undetected for longer than a decade and so applications need to be enhanced to screen skin cancer better.

Many studies currently compare AI-assisted diagnoses to dermatologist diagnoses, and they report similar or slightly higher accuracy for AI, with one study showing an 81.1% sensitivity and 86.1% specificity for AI-aided diagnoses compared to 75% sensitivity for dermatologist-only diagnosis. This may be due to human error, but generally professionals can supplement their decisions with AI's diagnosis.

However, these models often can report false positives because of their inability to distinguish signs of melanoma from other similar indicators. This is why our model is built on a dataset consisting of melanoma, keratosis, carcinoma, dermatofibroma, nevus etc. This allows the model to be trained on data that can help it distinguish melanoma from other variables.

For the model to improve in terms of accuracy, the AI models need to be trained on diverse datasets. The existing available data is based on more Caucasian skin types versus darker skin. This makes the model vulnerable towards unseen data and reduces its predictability.



Benchmark

The project is measured in terms of accuracy, precision and recall score. Moreover, the model deploys a rigorous training and validation loop and prioritizes reducing the training loss figure from 0.1831 and validation loss from 0.3802.

Data Description

The dataset used in this project is the Skin Cancer ISIC (The International Skin Imaging Collaboration) dataset, containing images of skin lesions categorized into several classes. Each class represents a specific skin condition, such as melanoma or benign keratosis. The dataset is organized into subfolders, with each folder corresponding to a class, enabling straightforward labeling. However, the class distribution is imbalanced, with some categories having significantly fewer images than others, which can impact model performance.

The images vary in resolution and require preprocessing for standardization. Although the dataset focuses solely on image data without additional metadata like patient demographics, it is well-suited for testing deep learning models for skin lesion classification and holds potential for medical diagnostic applications.

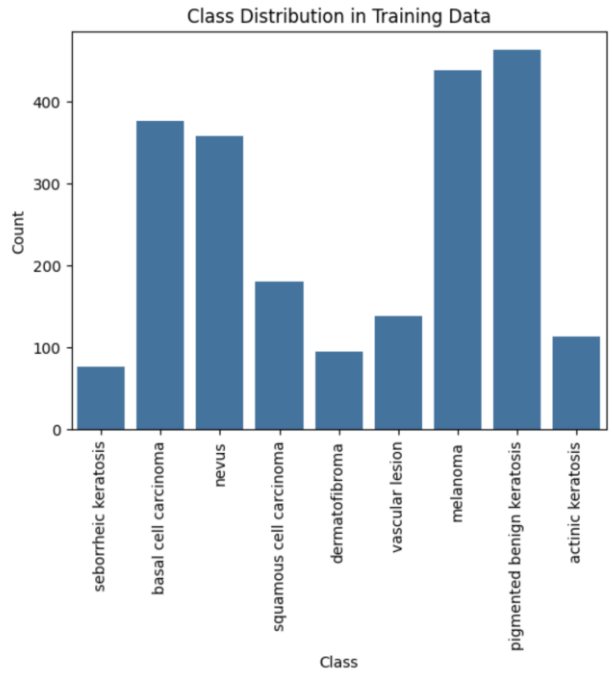
Data Preparation: Normalization and Augmentation

The dataset was preprocessed and prepared for deep learning using several techniques. Images were resized to 224 × 224 pixels for consistency and normalized using ImageNet mean and standard deviation to standardize pixel values.

Data augmentation techniques, including random cropping, horizontal flipping, rotation, and color jittering, were applied dynamically using PyTorch's transforms module. These augmentations helped increase data diversity and reduce overfitting during training.

The existing dataset has a lot of class imbalance in the training data as observed from the image below. This means that the model may be unable to learn of the features from non-majority classes that well. Even though melanoma has a relatively high count of images (438 images), pigmented keratosis appears to have a higher count and the model may then be more biased towards it. To address class imbalance, class weights were calculated based on the number of samples in each class and incorporated into the CrossEntropyLoss function, ensuring that minority classes were not under-penalized during training.

Additionally, a WeightedRandomSampler was used during data loading to sample batches in a balanced manner, ensuring each class was represented equally in each training epoch. This approach dynamically handled imbalance without physically modifying the dataset, resulting in a well-prepared input for training the deep learning model.



Despite class balancing being a bottleneck, the model is being trained on different classes of skin issues because that way the model can learn to distinguish melanoma. As the other images are relatively similar, this can allow the model to not detect false positives.

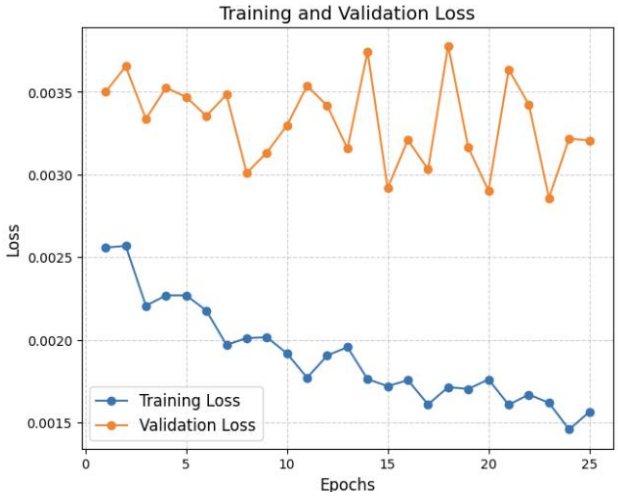
Modeling Implementation: Results and Evaluation

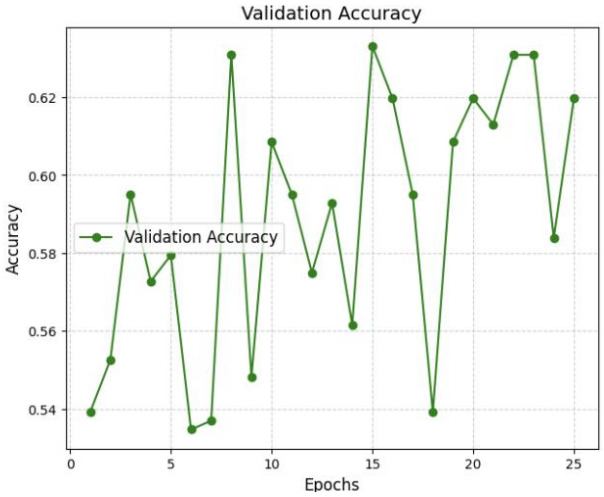
1: ResNET34 Model

A ResNet34 model because was used to extract specific features regarding melanoma from the skin cancer dataset. Using this pre-trained model, we are able to incorporate transfer learning into our model.

The model uses loss function hyperparameters to account for the class imbalance in the dataset. Melanoma is just 1 out of the 9 classes and focal loss is being used here to adapt the model to learn rare classes. Focal loss is a beneficial function that is appropriate for imbalanced class datasets. The model additionally uses augmentation such as. Rotations, crops etc to adjust the images. This is useful as the existing dataset images may be insufficient in understanding melanoma's features.

To further optimize the model, we used Adam so that the model retains complexity. Additionally, we use oneCycleLR Scheduler to make learning rates dynamic so that it allows the model to generalize more quickly.

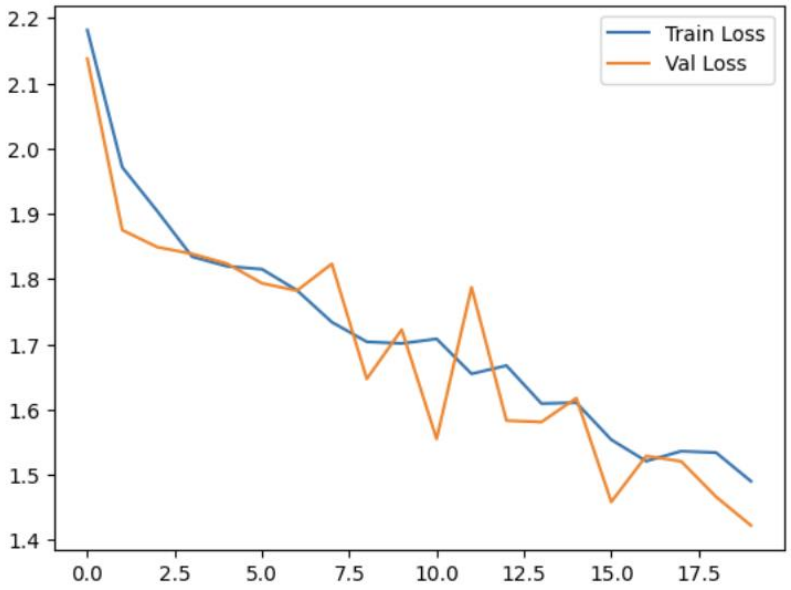




Results:
Epoch: 25
Train Loss: 0.0016, Val Loss: 0.0032, Val Accuracy: 0.6197
Weighted Precision: 0.5772
Weighted Recall: 0.5424
Weighted F1-Score: 0.5205

While the validation loss remains higher than training loss, the difference is very marginal. The low loss rate implies that the model is close to true predictions and is minimizing errors. However, its validation accuracy has improved over more epochs.

Model2: CNN model (With Augmentor)



Model 2 is a CNN model that was adapted with loss functions (CrossEntropyLoss) to improve its predictive ability. It has 5 convolutional layers with ReLU and max pooling. This allows the model to learn from different levels of images. Additionally, it uses a dropout layer of 30% to avoid overfitting. Additionally, we augmented the data to improve its generalization ability.

As we also analyzed that the data is class imbalanced, we used class weights for the model to allow the model to better learn from the representation. Taking the best practices from model 1, we also implemented Adam to adapt the model's learning rates.

Results:
Epoch - 25
Train Loss: 1.3956, Val Loss: 1.4863, Val Accuracy: 0.4519

The validation accuracy of 45% means that the model is only accurate in 45% of cases. This suggests that it's still not effective in detecting melanoma. While validation loss is lower than training loss, it's still relatively higher than 1 and also higher than model 1's loss figures.

Rather model 1 is a ResNet34 pre-trained model that can more strongly grasp image features. Additionally, ResNet has 34 layers, and this allows them to learn more quickly without experiencing issues with its gradients. However, we selected model 1 over model 2 due to its higher accuracy score which indicates that it has a better ability to predict and distinguish melanoma.

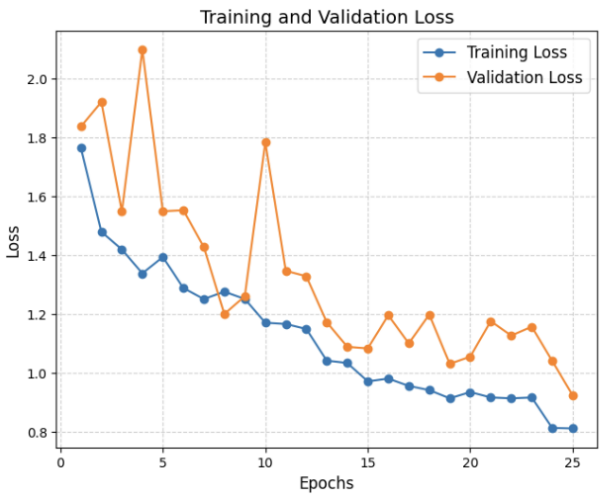
However, model 2 should still be investigated simultaneously as it can be customized for skin cancer detection. This may mean that it's better for deployment and can be replicated with less resources. If this model is fine-tuned more extensively with more data, it could address class imbalance more effectively and ultimately improve its accuracy

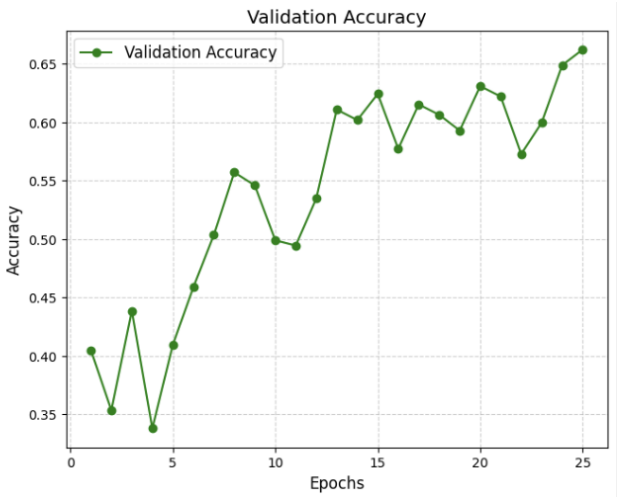
Model 3: ResNet 18 pretrained model

In model 3 and 4, we use ResNet 18 and adapt it more in model 4. This model uses 18 layers and requires less computational resources than ResNet34. This model also uses early stopping to avoid overfitting the model if the validation score deteriorates.

We modified the model by replacing its final layer to match the number of classes in our dataset, allowing it to focus on our specific problem. The learning rate was set to 0.001 to ensure the training process was stable and smooth, and a batch size of 32 was chosen to balance memory use and performance. We used the Adam optimizer because it adjusts the learning rate automatically, helping the model learn faster. To prevent overfitting, we applied a small weight penalty (0.0001) and used a loss function (CrossEntropyLoss) that takes class imbalances into account. Additionally, we used a learning rate scheduler (ReduceLROnPlateau) to slow down the learning rate when the model's improvement started to plateau.

After being trained with 25 epochs, the model has a train Loss of 0.8119, a Validation Loss of 0.9240 and a Validation Accuracy: 0.6622. Additionally, the model's Precision score is 0.4950, Recall score is 0.4746 and F1-Score is 0.4526. While the model is improving in validation accuracy, it still has a higher loss value than model 1. As the model's aim is to improve its predictive accuracy, a higher loss values means that the model is less effective in minimizing the error between its predictions and labels.





However, model 1 provides a training loss of 0.0016 and validation loss of 0.0032. While model 3 has a validation loss of 0.9240 and a training loss of 0.8119. This indicates that model 1 has a smaller gap and is more effective in predicting unseen data. Based on these metrics, model 1 can reduce overfitting better.

Model implementation challenges

The main challenge was addressing class imbalance in the dataset when implementing the models. It seems that despite using class weights and augmentor, the model isn't fully accurate yet. Melanoma also specifically has lower precision, recall and F1 scores. Despite adopting the model to focus on melanoma, the accuracy doesn't improve significantly. This indicates that the model needs to fine-tune more to focus on detecting melanoma more effectively. However, even if this is done, this may mean that the model is unable to distinguish melanoma from other similar images (e.g. like keratosis).

Business impact

This model could be used for early detection of skin cancer. Users can more easily scan skin lesions using the algorithm and then consult professionals. This also reduces healthcare costs dedicated towards biopsies. Moreover, the model would be useful in differentiating between melanoma and other variables e.g. keratosis. This could be useful in the long term to detect any skin deficiencies. The model can also be advanced to identify melanoma in various skin types.

Ethical considerations

As these models require huge sets of data to improve its accuracy and precision, more patient data is required. This means that patient data protections are crucial in implementing this model as we would need sensitive images of melanoma to train the model. Additionally, disclosure is an important ethical issue as the tool needs to disclose how user health information will be utilized.