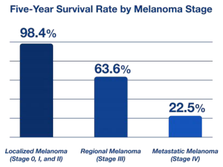
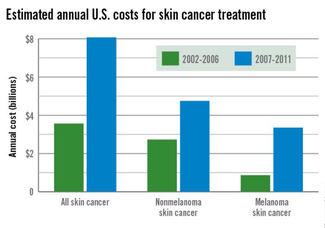
Applying Multiplicative Weights Update To Skin Cancer Classification Applications

Introduction

Skin Cancer Is A Worldwide Health Concern

Skin cancer is a significant global health concern due to its high incidence and mortality rates. According to the World Health Organization, skin cancer is one of the most common cancers worldwide, with nearly **3 million** new cases diagnosed annually. Additionally, with highly growing incidence rates, it is estimated that **one in every three cancers diagnosed globally is skin cancer.** Moreover, skin cancer is responsible for a significant proportion of cancer-related deaths, with an estimated **60,000 deaths globally each year.** The economic burden of skin cancer is also substantial, with the costs of prevention, diagnosis, and treatment adding up to **billions of dollars annually**. There is an urgent need for increased awareness, prevention efforts, and effective treatment strategies to combat the growing global burden of skin cancer.





Machine Learning Optimization

Machine learning algorithms have been instrumental in the field of medical imaging, allowing healthcare professionals to quickly and accurately identify potential medical conditions from scans and images. However, as the complexity of datasets and the number of features to be analyzed continue to grow, it is essential to use novel approaches to optimize these algorithms even further to **improve their accuracy and efficiency**. By using new and advanced techniques, machine learning can significantly enhance the accuracy and speed of medical image classification, leading to better diagnoses and treatments for patients.

Multiplicative Weights Update (MWU)

In the context of machine learning, MWU can be used to **combine the predictions of multiple models** to improve the overall accuracy and reliability of the predictions. The basic idea behind the MWU algorithm is to assign weights to each model based on its performance on the training data. The weights are updated iteratively based on the predictions made by each model on the training and testing data. The updated weights are then used to adjust the predictions made by each model, resulting in a final prediction that is **a weighted average of the individual model predictions.** In this specific study, the following formulas were used:

Initializing the Weights: $w_i = \frac{1}{k} \; for \; all \; i$

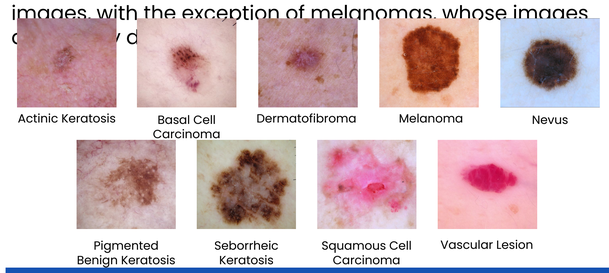
Updating the weights: $w_{i(t+1)} = w_i(t) * e^{(-\eta * Li(t))}$

Combining models: $P = w_1 * p_1 + w_2 * p_2 \cdots + w_k * p_k$

Methods and Procedure

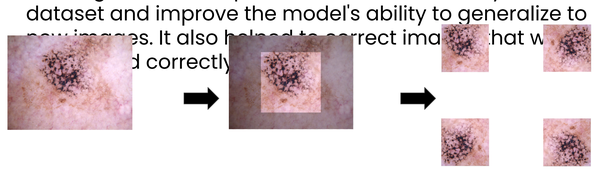
1. Dataset and Preprocessing

The dataset for this project consists of 2357 images of oncological diseases, which were formed by The International Skin Imaging Collaboration (ISIC). All images were sorted with the classification taken from the ISIC, and all subsets were divided into the same number of images, with the exception of melanomas, whose images c[illegible] y c[illegible]



Preprocessing Steps:

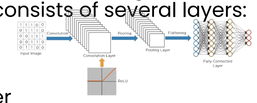
- All Images were **cropped to 150 x 150 pixels** using OpenCV.
- Cropping helps eliminate background noise, focus on relevant parts of the image, and standardize image size.
- Images were **rotated programmatically** by 90, 180, and 270 degrees. This helps to increase the diversity of the dataset and improve the model's ability to generalize to new images. It also helped to correct ima[illegible]hat w[illegible] correctly



2. Models

Convolutional Neural Network (CNN):
A series of layers that process input images to identify features and classify them into categories. The basic architecture of a CNN consists of several layers:

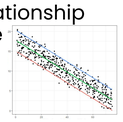
- Convolutional layer
- Activation layer
- Pooling layer
- Fully connected layer



Linear Regression:
Statistical model for estimating the relationship between a dependent variable and one independent variables

$y = \beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon$

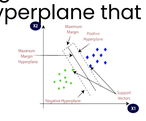
$\beta = (X^T X)^{-1}(X^T y)$



Support Vector Classifier (SVC):
Works by mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into two classes.

$f(x) = sign((w^T * x) + b)$

$minimize \; \frac{1}{2}||w||^2 + C * \sum_{i=1}^{n}\left(\sum_{j=1}^{d} max(0, 1 - y_i((w^T * x_i) + b))\right)$



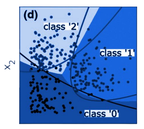
Logistic Regression:
Based on the sigmoid function, which maps any input value to a probability between 0 and 1. The output of the model is the predicted probability of certain skin diseases.

$log(odds) = w^T x + b$

$p = \frac{odds}{1 + odds} = \frac{e^{(-w^T x - b)}}{1 + e^{(-w^T x - b)}}$

$L(w, b) = \Pi_{\{i=1\}}^{n} p_i^{y_i}(1 - p_i)^{1 - y_i}, p_i$

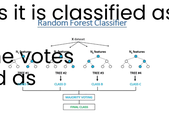
$l(w, b) = \sum_{i=1}^{n}[y_i * log(p_i) + (1 - y_i) log(1 - p_i)]$



Random Forest Classification:

- Randomly selects a subset of the training data set
- Grows a decision tree using the subset as input
- Repeats many times to create a "forest" of decision trees
- Passes new examples through each decision tree and counts the number of times it is classified as each class
- The class with the majority of the votes is what the example is classified as

$f = \frac{1}{K}\sum_{k=1}^{K} f_k$

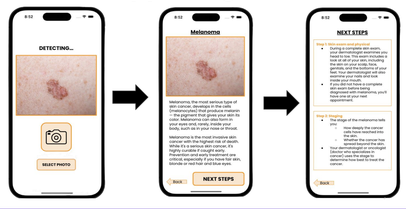


$H = -\sum_{k=1}^{n}(p_k \log_2(p_k))$

Training and Testing:
All models were trained for 30 epochs with a batch size of 50 images. The Multiplicative Weights Update algorithm was trained for 115 epochs with a batch size of 25 images to allow it to have more chances to iterate and update the weights of each model.

3. Interfacing Technology

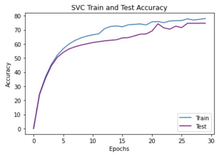
IOS App Design Sketch

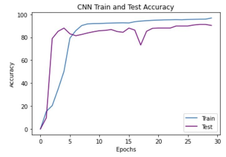


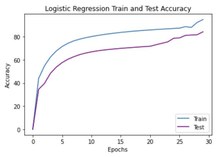
This widespread use of mobile devices presents an opportunity for healthcare professionals to provide accessible and convenient healthcare services to people **all over the world**.

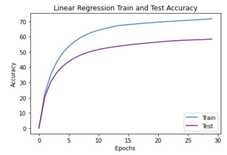
As smartphones become more ubiquitous, the availability of healthcare services on these devices has the potential to reach individuals who may not have access to traditional healthcare services, particularly individuals who live in rural or remote areas, where access to healthcare may be limited. Additionally, having a mobile app that allows for the early detection of skin cancer can lead to **improved health outcomes** and **reduced healthcare costs**.

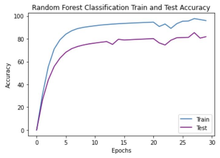
Results





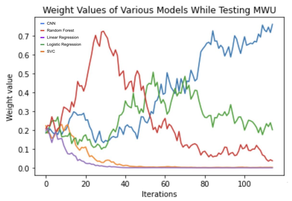






Various Classification Models And Accuracies

Model	Training Accuracy	Testing Accuracy
SVC	78.05%	74.75%
CNN	96.78%	90.03%
Logistic Regression	94.97%	84.31%
Linear Regression	71.67%	58.44%
Random Forest Classification	95.97%	81.85%
MWU	98.76%	97.20%



Final Model Weights After MWU

Model	Weight Value
SVC	**0.00**
CNN	**0.71**
Logistic Regression	**0.23**
Linear Regression	**0.00**
Random Forest Classification	**0.06**

Discussion & Impact

The results show that the best models during the training and testing phases were the CNN, Logistic Regression, and Random Forest Classification. In the future, these models should be emphasized more than the Linear Regression and SVC models. The approach with MWU achieved an accuracy of **97.2%** in skin cancer classification, **outperforming each individual model**. Furthermore, this study demonstrates the effectiveness of the MWU algorithm in optimizing skin cancer classification models. MWU has the potential to improve the accuracy of skin cancer diagnosis and facilitate early detection, ultimately improving patient outcomes globally.

The MWU algorithm can also be applied to other fields of study as it is a powerful tool that can be used to optimize a wide range of algorithms. The technique has the potential to revolutionize how we approach some of the world's most pressing problems.

Future Steps

1. Testing on real skin cancer patients
   - Clinical trials for validation
   - Talking with medical professionals and industry experts
   - Adding more classifications of skin cancer
2. Further development of mobile app
   - Leveraging GPS and public health data to provide better healthcare options and recommendations
   - Cleaner and more user-friendly UI
3. Enhancing camera and zoom capabilities
   - Super-resolution
     - Takes low-resolution images and generates a high-resolution version of it using deep learning and interpolation