ENGR-E516 Engineering Cloud Computing

PROJECT PROPOSAL

SKIN LESION CLASSIFICATION FOR MELANOMA DETECTION

Sakshi Rathi

Sreshta Reddy N

Shruti Houji

September 20th, 2023

1. Project Overview:

Skin lesions, especially melanoma, can be life-threatening if not detected early. The goal of this project is to utilize machine learning and image processing techniques to analyze dermoscopic images of skin lesions and classify them as benign or malignant (melanoma). By deploying this solution on AWS, we aim to provide a scalable and efficient system that can assist dermatologists in early diagnosis and potentially save lives. AWS, as a leading cloud platform, provides several advantages over traditional hosting methods, including scalability, affordability, and adaptability. By harnessing AWS's robust suite of services, we envision a system that not only efficiently classifies uploaded skin images but also scales dynamically based on demand, ensuring uninterrupted, swift, and reliable service for end-users.

The project's life cycle will encompass several phases, including data acquisition, preprocessing, feature extraction, model training, and hyperparameter optimization. AWS's cloud computing capabilities will be instrumental at each stage, from storing large datasets in S3 to employing powerful EC2 instances for training complex neural networks. Once the model reaches satisfactory accuracy, it will be deployed as a web service using AWS Lambda and API Gateway. Continuous monitoring, facilitated by AWS CloudWatch, will be set in place to oversee the system's performance, promptly identify potential anomalies, and ensure optimal service uptime and reliability.

2. Project Introduction:

Skin cancer, particularly melanoma, is a lethal disease with a rising global incidence and early detection is paramount for successful treatment and improving patient outcomes. The "Skin Lesion Classification for Melanoma Detection" project addresses this critical need by leveraging the power of machine learning and deep neural networks. The project aims to develop an advanced model capable of accurately classifying skin lesions into 9 different classes like Melanocytic nevus, Basal cell carcinoma, Actinic keratosis, Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis), Dermatofibroma, Vascular lesion, Squamous cell carcinoma, and None of the above. The dataset used is ISIC 2019 which contains 25,331 images used for classification of dermoscopic images.

3. Related work and gap analysis:

We have done research on Skin Lesion Images for melanoma classification based on the following papers:

1. “Deep Semantic Segmentation and Multi-Class Skin Lesion Classification Based on Convolutional Neural Network” by Muhammad Almas Anjum, Javaria Amin, Muhammad Sharif , (Senior Member, Ieee), Habib Ullah Khan, (Member, Ieee), Muhammad Sheraz Arshad Malik, And Seifedine Kadry et al. (2020) focused on localization, segmentation, and classification of the skin lesion in early stages (3 phases) which is evaluated on the top MICCAI ISIC challenging 2017, 2018 and 2019 datasets based on CNN model. Phase I utilizes a tinyYOLOv2 model with ONNX and squeeze Net as the backbone, effectively localizing various types of skin lesions. Phase II employs a 13-layer 3D-semantic segmentation model to segment the lesions, with a pixel classification layer ensuring accuracy. Finally, in Phase III, deep features extracted using a ResNet-18 model are

optimized using the ant colony optimization (ACO) method and passed to classifiers like optimized (O)-SVM and O-NB.
2. "An approach for multiclass skin lesion classification based on ensemble learning" by Zillur Rahman, Md. Sabir Hossain, Md. Rabiul Islam, Md. Mynul Hasan, Rubaiyat Alim Hridhee et al. (2021) proposed a weighted ensemble model using 5 deep neural networks like ResNeXt, SeResNeXt, ResNet, Xception, and DenseNet based on the dataset ISIC 2019 and 18730 dermoscopy images from official Human Against Machine using grid search method with recall score of 94%.
3. "Skin Lesions Classification into Eight Classes" for ISIC 2019 Using Deep Convolutional Neural Network and Transfer learning" by Mohamed A. Kassem, Khalid M. Hosny, and Mohamed M. Fouad et al. (2019) presented pre trained model with GoogleNet and transfer learning to classify into 8 types and one class as unknown images which achieved an accuracy of 94.92%.
4. "WonDerM: Skin Lesion Classification with Fine-tuned Neural Networks" by Yeong Chan Lee, Sang-Hyuk Jung, and Hong-Hee Won et al. (2018) described a neural network fine-tuned with segmentation task data and classified into 7 different types and achieved an accuracy of 0.899 and 0.785 in the validation and test sets respectively.
5. "Skin Lesion Classification Using Hybrid Deep Neural Networks" by Amirreza Mahbod, Gerald Schaefer, Chunliang Wang, Rupert Ecker, Isabella Ellinger et al. (2019) stated that CNN is the best among all the classical methods, which uses optimized deep features from a number of well established CNN's along with pre trained deep models are used with 83.83% for melanoma classification and of 97.55% for seborrheic keratosis classification.

From all the above papers, we can conclude that the ensemble model excels in recall, transfer learning in accuracy, and CNNs in overall effectiveness. However, there remains a potential gap in addressing the interpretability and explainability of these models, which is crucial for gaining trust among medical professionals and ensuring the models' clinical relevance. Hence, scalability and ease of deployment are factors that need further consideration in real-world applications.

4. Proposed tasks (with detailed methodology)

The project aims to develop a robust skin lesion classification system that has 9 distinguishing categories. The model will use deep learning/machine learning techniques for classification and be deployed on Amazon Web Services (AWS) to enable real-time classification and accessibility.

The image classification part of the project would follow this:

1. Data pre-processing: Preprocess the dataset by resizing, normalizing, and augmenting the images to ensure uniformity and increase model robustness.
2. Model Development and evaluation : A suitable deep learning architecture for image classification will be chosen and a model will be trained using the preprocessed dataset. We can also use the SageMaker's GPU-powered instances for efficiency. The optimization of the model will be done by hyperparameters tuning through automatic or manual tuning. The trained model will be then deployed to a web server for monitoring and inference. The model's performance will be assessed on the trained model's performance on a validation dataset.

The AWS aspect involves leveraging Amazon Web Services for various stages of the project.

- Amazon S3: It'll serve as a primary data storage solution for the project. This dataset stored will include both the training and validation images. The model checkpoints and logs are also stored in S3 for easy access and versioning.
- AWS Lambda: The model can be deployed as a Lambda function. It can also be used for predicting the image using the model and returns the classification results in real-time. For the web application, the various backend tasks will be handled by lambda.
- Amazon SageMaker: It'll be used to train the machine learning/deep learning model on GPU-powered instances, enabling faster training times. The built-in algorithms or custom scripts can be used for model development. The automatic model tuning can be used for hyperparameter optimization. In addition, sagemaker will be used for deployment of the trained model and creating a scalable and secure API for real-time inference.
- AWS QuickSight: For data visualization and monitoring purpose, to analyze the performance of the models and to create an interactive dashboard. This will be useful for the user to monitor and to keep a track of any anomalies.

5. Team members and workload allocation

Sakshi will be doing the machine learning and the model development part starting from data pre-processing, classification model and evaluation by improving model's performance. Shruti will be doing the AWS Infrastructure and development, by setting the AWS account, setting up s3 buckets for storing data and model checkpoints and model deployment using Sagemaker. Sreshta will be integrating the lambda functions and working with AWS quicksight for visualization dashboard and with the other user related interface. Overall, it'll be a combined team effort with these being the main tasks.

6. Planned timeline

Week 1: System environment configuration and Dataset upload in AWS S3.

Week 2: Preprocessing and EDA on the Data using AWS SageMaker.

Week 3: Implement Prediction Models on the data in AWS SageMaker.

Week 4: Building UI, Back-end and Integration of AWS Lambda.

Week 5: Cloud Infrastructure Setup and Creation of API Gateway for our model. Development of a website to utilize the Model.

Week 6-7: Deployment

Week 8: Monitoring, Maintenance and submit the project report

The specific timeline and tasks may vary depending on the scope and complexity of the project, but this should give you a general idea of the milestones that need to be completed.

References:

1. https://www.kaggle.com/datasets/andrewmvd/isic-2019
2. https://www.sciencedirect.com/science/article/pii/S2352914821001465
3. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9139937
4. AWS SageMaker Inbuilt Model Tuning
5. AWS API Gateway Integration with AWS Lambda and AWS SageMaker