**Introduction**

This report analyses the diagnosis of skin lesions using machine learning techniques. We basically analyse the background of the project, its objectives, the methods used in this project and what are findings are. It also reflects on the process used on course and the tools that are used to make this process happen.

**Project Background and Justification**

The importance of diagnosis of pigmented skin lesions using Machine learning and image processing techniques is growing in the field of medicine by helping in early detection of skin conditions like skin cancer. The project aim is to develop a machine learning model using Deep learning technique like CNN for the diagnosis of pigmented skin lesions. The aim of the project is to develop and train a model with a large dataset and using these models with test data to evaluate their performance to develop a final model.

**Project Criteria/Objectives**

This project objective is to develop machine learning model for the diagnosis of pigmented skin lesions. The diagnosis method must be able to classify each image to a set of given conditions. The models will also be evaluated using various model evaluation techniques as part of the project. The process is mentioned as below:

* Develop a CNN based models
* Train the models using a diverse and large sample sized dataset
* Classify the diagnosis based on a selection of skin conditions (stated above)
* Find and analyse model confusion matrices and measure model’s accuracy
* Optimize the model and find the best fit model.

**Tools, Methodology and Evaluation Design**

The project will follow the CRISP-DM process model. Git and Github is used for version control to manage the progress of project template. The project was developed on a cloud-based IDE with access to GPU to run the neural networks which helped in reducing the running time to train the models.

The project will be divided into these stages:

1. Business Understanding:

Identifying the purpose of the data and determining the best method to approach the problem.

1. Data Preparation:
   * Cleaning the data to be used to build the model with necessary changes.
2. Data Understanding:
   * Performing the basic EDA of the dataset provided after cleaning the data.
3. Modelling and Evaluation:
   * Building a neural network using CNN technique and then optimizing the model to get better results.

**Exploratory Findings and Results**

The EDA has shown that skin lesions are more likely to occur at the back and trunk. Also, lesions are overwhelmingly of the Melanocytic Nevi type.

There were some missing values with age which was replaced with their mean values. Most people affected by lesions fall under the age bracket between 35-60 years. There was less number lesion that occurred on the back and genital area.

.

**Evaluation and Future work**

When it comes to models, the CNN technique seems to work better than the traditional machine learning models. There are 3 different activation functions used in the model of CNN. The 3 activation functions used were SoftMax, Relu and tanh. Of all the 3 activation functions, Relu had a better accuracy and performance over SoftMax and tanh. The entire project was built using a cloud-based IDE with GPU to decrease the computation time due to limited infrastructure availability.

Due to account limitations on the cloud-based IDE, the model was developed using the SoftMax as an activation function. But the model was also verified with tanh and relu functions.

Using SoftMax function the accuracy was around 67% and for tanh function the accuracy was also around 68%. Using relu function the accuracy was close to 74% which by far performed best amongst the three.

In future projects, when solving problems with high dimensions it is better to use GPU to decrease the computation time to process. In this project, the difficulty occurred when the model was trained on CPU/ local system and due to infrastructure limitations, the need to use cloud-based IDE seemed more fruitful. Not only it reduced the time in training the models but also gave more accuracy.

This work can be further developed by putting these models into production which can help people in making much more accurate decisions with the help of machine learning. The entire project can be developed and hosted on a cloud to improve the performance and not worry about infrastructure.