MACHINE LEARNING BASED DETECTION OF CANCEROUS TUMORS

**1. Introduction**

Cancer remains a leading cause of death worldwide, emphasizing the urgent need for advancements in diagnostic methodologies. Traditional diagnostic techniques, often involving manual examination of histopathological slides by pathologists, are time-consuming and prone to human error. To address these limitations and improve patient outcomes, this study explores the potential of machine learning (ML) and deep learning (DL) techniques for automated cancer cell detection.

Machine learning algorithms offer a promising approach to analyze complex patterns within histopathological images, enabling the identification of cancerous cells with greater accuracy and efficiency. By automating this process, pathologists can be assisted in making more informed and timely diagnoses, leading to improved treatment outcomes.The objective of this research is to develop a robust and accurate cancer cell detection system using a variety of machine learning algorithms. This includes comparing the performance of traditional ML algorithms like support vector machines, logistic regression, linear discriminant analysis, quadratic discriminant analysis, decision trees, Gaussian naive Bayes, random forest, Gaussian process classifier, AdaBoost, and XGBoost with deep learning models, particularly convolutional neural networks (CNNs).

By evaluating the performance of these algorithms on a comprehensive dataset of histopathological images, this study aims to identify the most effective approach for cancer cell detection. Furthermore, the research will explore the factors influencing the performance of different algorithms, such as dataset size, image quality, and feature engineering techniques.

The findings of this study will contribute to the advancement of medical image analysis and provide valuable insights for the development of automated cancer diagnostic tools. By leveraging the power of machine learning, this research has the potential to improve patient care and outcomes.

**2. Product Goal**

The primary goal of this sprint is to create a new ensembled model based on different classifiers and achieve higher accuracy than single classifier models.

**3. Demography (Users, Location)**

**Users**

* **Target Users:** Pharmaceutical Industry
* **User Characteristics:** Any company which is involved in making cancer detection systems.

**Location**

* **Target Location:** Global

**4. Business Processes**

The key business processes include:

* **Deployment:**
  + Deployment and integration of proposed model with the already deployed models
* **Scaling:**
  + Scaling the model to handle even larger datasets.

**5. Features**

This sprint will focus on implementing the following key features:

* **Ensembling Method:**
  + Develop an ensemble model to combine various classifiers into a single model.
* **Improving Accuracy:**
  + The final goal is to create a model with the highest achievable accuracy.

**6. Authorization Matrix**

Define the roles and their corresponding access levels:

| **Role** | **Access Level** |
| --- | --- |
| Developers | Full access to the model |
| End user | Full access to the model |
|  |  |
|  |  |

**7. Assumptions**

* The development environment and infrastructure will remain stable throughout the sprint.
* The team possesses the necessary skills and resources to develop and integrate AI-based fraud detection models effectively.