Diagnosis of Spinal Cord Injury

SCI Spinal Cord Injury

CAIRO UNIVERSITY

BIOINFORMATICS 19-23

COMPUTER SCIENCE
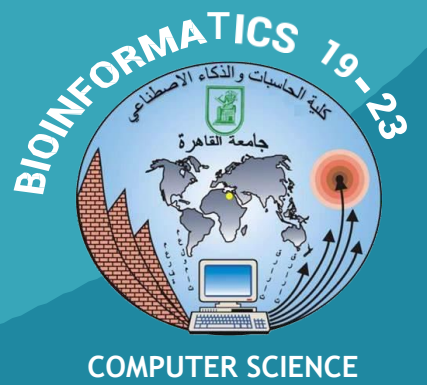
SUPERVISORS: 1. Dr. Ahmed Farouk   2. Dr. Sabah Sayed   3.TA. Sarah Ahmed

TEAM :   1. Habiba Mustafa 2. Maryam Ali 3. Maryam Ghanem 4. Omar Khaled 5. Rania Ghanem

1. Introduction

The spinal cord is a major pathway for motor and sensory signals travelling between the brain and the peripheral-nervous system.When a spinal cord injury (SCI) occurs, the spinal tracts, which convey sensory, motor, and autonomic- signals between the brain and organs, are disrupted.

Spinal cord injuries are very common and usually cause patients to become paralyzed or stop organs from functioning properly. Early detection of SCI is extremely useful which is why several approaches for SCI detection exist. However, these approaches do not make use of the power of artificial intelligence and thus face many challenges.
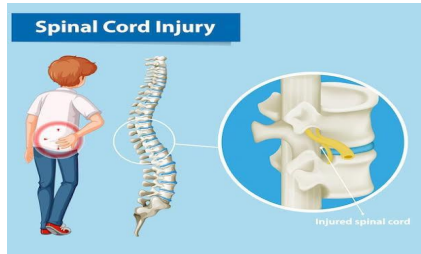
3.Tools & Technologies

The project utilizes a variety of software and hardware tools to develop the application that can predict spinal cord injury using gene expression data.

In terms of software, machine learning libraries such as scikit-learn and TensorFlow are utilized in the project for developing the machine learning models that predict spinal cord injury.

The application is implemented in Python and uses Flask for handling server-side scripting. Furthermore, the project makes use of web technologies such as HTML, CSS, and JavaScript for building the user interface and creating a user-friendly experience.
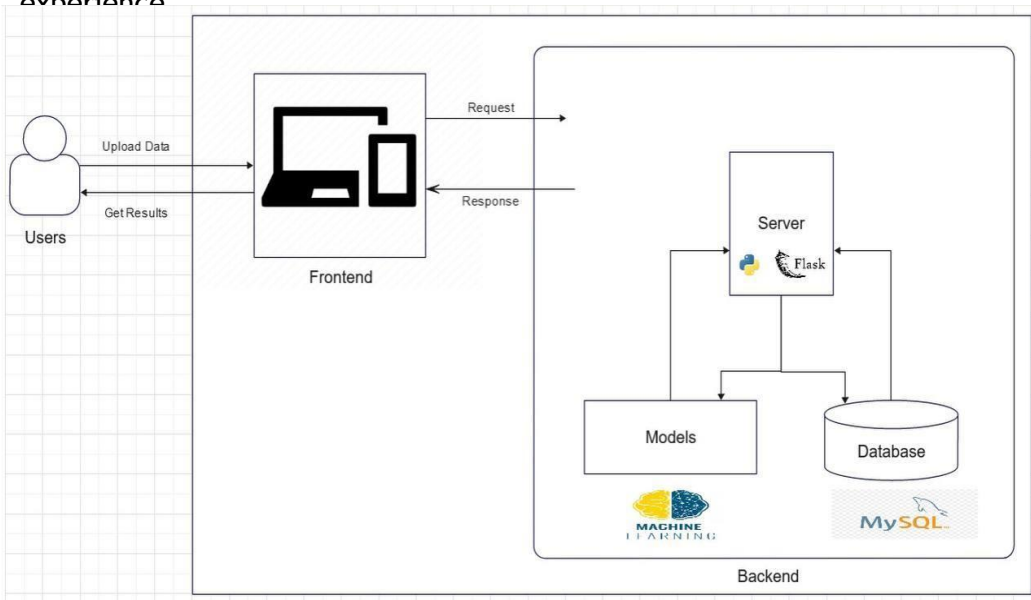
5. Methods

• **Machine Learning Pipeline**
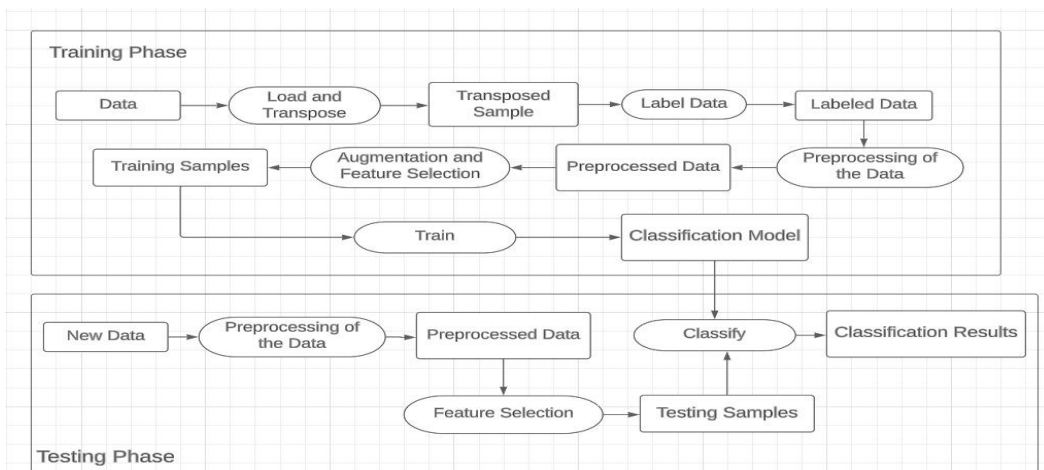
In the case of diagnosing spinal cord injury using gene expression data, there are typically a very large number of genes that need to be considered while providing a few number of samples.This makes the classification task computationally and statistically challenging.

To address this challenge, multiple techniques were tested for feature (gene) selection, including SelectFromModel, VarianceThreshold, and chi-square feature selection. Then, data augmentation techniques such as oversampling and Synthetic Minority Over-sampling Technique (SMOTE) were also tested to increase the number of samples.

The next step is to load and preprocess the data. This involves a series of standard procedures including normalization, shuffling and label encoding.

The final step in the classification process is building, testing, and evaluating different machine learning models. Several ML algorithms were used for this task, including logistic regression, support vector machine (SVM), decision tree (DT), naive Bayes, k-nearest neighbors (KNN), and neural networks.

• **Web Application**

For the website, we used the "Waterfall" development methodology to analyze the system requirements, make the website design, develop the HTML code, and integrate Flask to build the website's functionality.

2. Objectives

The objective of this project is to conduct research in which we build and evaluate several machine learning (ML) models to perform early detection of non-traumatic SCI from gene expression data as well as develop a prognostic tool (web application) that can predict spinal cord injury through identifying and analyzing the genes that affect it.

The proposed application would be designed to take in an individual's genetic data and provide a personalized risk assessment for SCI, based on their genetic profile, to reduce the risk of SCI. Additionally, it can be used by a neurologist to find out if his patient is healthy, with a non- traumatic injury or if he has spinal cord injury just by uploading the patient gene expression file.

4. Materials

• **Dataset**

The GSE151371 ribonucleic acid (RNA) -sequencing dataset, which contains blood RNA biomarkers for spinal cord injury, was downloaded from the Gene Expression Omnibus (GEO) database. The dataset consists of a normalized gene expression matrix for humans, comprising 17,501 genes and 58 sequenced samples.

This dataset provides a valuable resource for the identification of gene expression patterns associated with different types of SCI and the development of diagnostic and therapeutic tools for this debilitating condition.

6. Results

The following graph presents the performance results of different models and data preparation techniques.
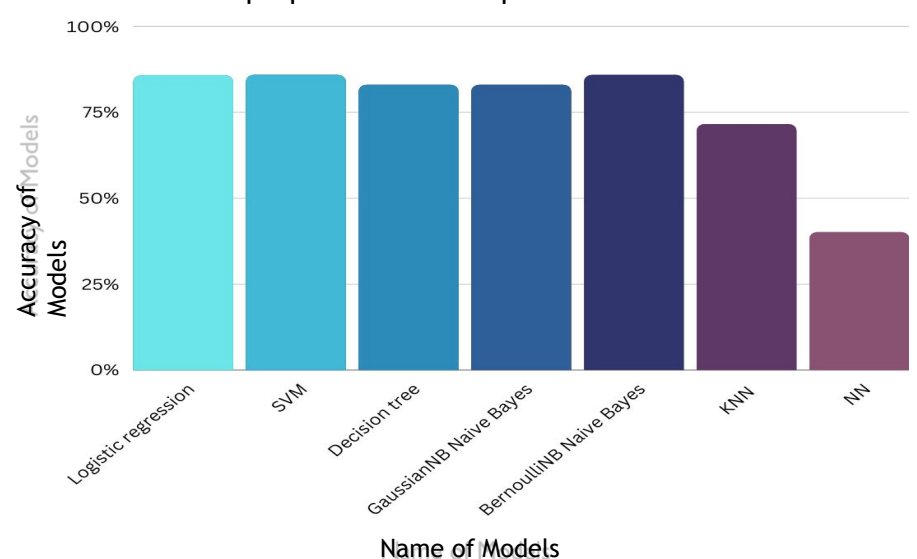
The results show that the highest accuracy is achieved by logistic regression & SVM & BernoulliNB Naive Bayes , with chi-square feature selection and data augmentation, reaching an accuracy of 85.71% ,The Logistic Regression model is deployed in the web app.

7. Future Work

In this project, early detection of SCI is extremely important. so, we built and evaluate several machine learning models to detect SCI from human gene expression data.

Furthermore, with the recent development of technology and availability of data spinal cord injury severity can be obtained easily.

Scan Here