**PREDICTION OF DISEASE USING MACHINE LEARNING.**

**Abstract**

Disease prediction using machine learning is a system that is used to predict the disease from the symptoms which are given by patients or any user. The system processes the symptoms provided by the user as input and gives he output as the probability of the disease. XGBoost classifier is used in the prediction of the disease which is a supervised machine learning algorithm. With an increase in healthcare data, accurate analysis of data benefits early disease

**Introduction**

Machine Learning is the domain that uses past data for predicting. Machine Learning is the understanding of computer system under which the Machine Learning model learn from data and experience. The machine learning algorithm has two phases: 1) Training & 2 Testing. To predict the disease from a patient’s symptoms and from the history of the patient, machine learning technology is struggling from past decades. Healthcare issues can be solved efficiently by using Machine Learning Technology.

I am applying complete machine learning concepts to keep the track of patient’s health. ML model allows us to build models to get quickly cleaned and processed data and deliver results faster. By using this system doctors will make good decisions related to patient diagnoses and according to that, good treatment will be given to the patient, which increases improvement in patient healthcare services. To introduce machine learning in the medical field, healthcare is the prime example.

**Objective**

* To reduce efforts taken by patients to diagnose a disease
* to reduce deaths arising from late diagnosis

The esisting system uses XGBoost model and Logistic regreesion to predict heart failure with an accuracy of 90% and 80% respectively. I would further propose use of CNN for better and more accurate diagnosis by using scanned health images of the heart from either MRI or CT-Scan procedures.

**DATASET AND MODEL DESCRIPTION**

Dataset used in this system is in a structured format. The dataset which is used contains the various features that may lead to heart failure. As the system is based on supervised learning machine algorithms, the dataset is having the label with 0 or 1. Then I divide the dataset into a Training dataset and Testing dataset. The model is trained by a training dataset. Logistic Regression and XGBoost algorithm were applied to this training dataset and then the machine learning model is trained. Then the testing dataset was provided to the trained model to test the accuracy of the model.