

Data Analytics Measure for Reporting Disease Outbreak in a Region by Prediction of Supervised Machine Learning Approach

¹Narendra Yejarla, ²R. Senthil Kumar

¹Student, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India, narendray0406@gmail.com

²Assistant Professor, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India, rsenthilmecse@gmail.com

Article Info

Volume 82

Page Number: 10353 - 10356

Publication Issue:

January-February 2020

Abstract

Generally, healthcare industry has become big business and produces large amounts of health-care data daily that can be used to extract information for predicting disease that can be happen in future. Whereas victimisation the treatment history and health information, this hidden information in the healthcare data will be later used for affective decision making for patient's health. Also, this area needs improvement by using the informative data in healthcare. To prevent this problem in hospital sectors, must predict whether the disease is happened or not by given attributes from given dataset using machine learning techniques. To propose a machine learning-based method to accurately predict the diseases (heart, diabetes, breast etc.) by prediction results in the form of best accuracy from comparing supervise classification machine learning algorithms. The aim is to investigate machine learning based techniques for patient disease forecasting by prediction results in comparing best accuracy with evaluation of GUI application results. Additionally, to compare and discuss the performance of various machine learning algorithms from the given dataset with evaluation classification report, identify the confusion matrix and to categorizing data from priority and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy with precision, Recall, F1 Score, sensitivity and specificity.

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 19 February 2020

Keywords: Dataset, Machine learning -Classification method, python, Prediction of Accuracy result.

1. Introduction

Machine learning is to predict the long run from past data. Machine learning (ML) could also be a method of computing (AI) that has computers with the ability to be told whereas not being expressly programmed. Machine learning focuses on the event of laptop/computer Programs that will modify once exposed to new data and thus the fundamentals of Machine

Learning, implementation of a simple machine learning algorithmic program exploitation python. Technique of training and prediction involves use of specialized algorithms. It feed the training data to associate algorithmic program and thus the algorithmic program uses this training data to supply predictions on a different take a glance at data. Machine learning could also be roughly separated in to a number

of categories. There unit of measurement supervised learning, unsupervised learning and reinforcement learning.

Disease prediction by exploitation of patient treatment history and health data by applying processing and machine learning techniques is current struggle for the past decades. Many works area unit applied processing techniques to pathological data or medical profiles for prediction of specific diseases. These approaches tried to predict the reoccurrence of malady. Also, some approaches try to do prediction on management and progression of malady. The recent success of machine learning in disparate areas of machine learning has driven a shift towards machine learning models that will learn affluent, stratified representations of information with little pre processing and manufacture loads of correct results. Numbers of papers area unit disclosed on several processing techniques for designation of cardiovascular disease like decision tree, naive mathematician and support vector machine showing utterly totally different levels of accuracies in diseases prediction.

2. Literature Survey

Breast cancer is that the most extremely aggressive cancer and a significant pathological state in females, and a number one reason for cancer-related deaths worldwide. To the power of predicting cancer prognosis additional accurately not solely might facilitate carcinoma patients comprehend their life, however conjointly facilitate clinicians build hip choices and any guide acceptable medical aid. Meanwhile, prognostication plays a crucial role in clinical works for all clinicians, significantly those clinicians operating with short term survivor. Once a fairly correct estimation of prognosis is accessible, clinicians usually utilize prognosis prediction information to help with

clinical deciding, establish patient's eligibility for care programmers, style and analysis of clinical trials. To comprehensively assess they planned technique, they tend to use ten-fold cross validation experiment in in keeping with previous existing studies of cancer prognosis prediction. Specifically, the patients in our experiment are randomized into 10 subsets. For each round, 9 of these 10 subsets are divided into training (80%) and validation (20%) sets, whereas the remaining one set is used as testing set. During this approach, they tend to get the prediction immeasurable every testing set once ten rounds and so merge them as associate overall prediction scores.

Besides, in their study, MDNNMD doesn't optimize the model configurations and weight coefficients at the same time. For performance analysis, they tend to plot receiver operational characteristic (ROC) curve that shows the interaction between sensitivity and 1-specificity by variable a call threshold, and computes the United Self-Defence Force of Colombia. The analysis metric, Sensitivity (Sn), Specificity (Sp), Accuracy (Acc), Precision (Pre) and Matthew's correlation (Mcc) are used for performance analysis. They compare the performance of MDNNMD with 3 wide used strategies for prognosis prediction of breast cancer: SVM, RF and LR. Ten-fold cross validation experiment for prognosis prediction of carcinoma is conducted with four totally different strategies. During this study, they have a tendency to use an RF and LR package obtained from sci-kitlearn. carcinoma is that the commonest unwellness and is typically related to poor prognosis. Therefore, there's an pressing ought to develop effective and quick procedure strategies for carcinoma prognosis prediction. In this work, they have a tendency to gift a completely unique multimodal deep neural network by desegregation multi-dimensional

information named MDNNMD to predict the survival time of human carcinoma.

To expeditiously incorporate two-dimensional information as well as organic phenomenon profile, CNA and clinical information in carcinoma, 3 freelance DNN models square measure created to get a final multimodal DNN model considering the heterogeneousness of various forms of information.

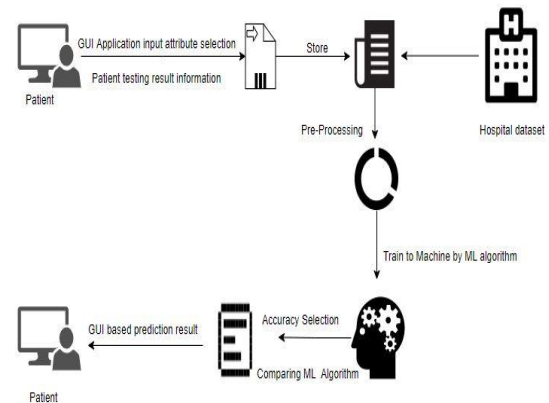
3. Proposed Work

Scope of this strategy is that if disease is expected than early treatment is given to the patients which might scale back the danger of life and save lifetime of patients and value to urge treatment of diseases is reduced up to some extent by early recognition. For this downside, a machine learning approach can train for prediction of disease. The fast adoption of electronic health records has created a wealth of latest knowledge concerning patients, that may be a goldmine for rising the understanding of human health. The above technique is employed to predict diseases using patient treatment history and health knowledge.

In the commencement of accumulating information, data from antecedent patients and online sources are gathered along. These datasets are incorporate to create a standard dataset, on that analysis are going to be done.

In the next step patients are allowed to enter their testing result data through GUI application. This data is going to be compared with the information set provided by hospital. Then the trained machine can predict the results exploitation supervised machine learning algorithmic program with high accuracy. The predicted results are going to be showed to patients through GUI application.

4. System Architecture



5. Conclusion

This analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public data set with higher accuracy score will be find out. This brings some insights about diagnose of the particular disease. Early diagnosis of disease is most important for the patient to reduce its impact. To present a prediction model with the aid of artificial intelligence to improve over human accuracy and provide with the scope of early detection. Early prediction will enhance the health condition of people which helps in decrease of medical expenditure and mortality rate of a patients.

6. Results

It helps the hospitals to automate the detecting the patient disease from eligibility process (real time) supported the given dataset detail, to automate this process by show the prediction end in web application or desktop application and to optimize the work to implement in AI environment.

Variable	Description
Pregnancies	Number of times pregnant.
Glucose	Plasma glucose focus a 2 hours in an oral glucose resilience test.

Blood Pressure	Diastolic circulatory strain (mmHg).
Skin Thickness	Triceps skin overlap thickness (mm).
Insulin	2-Hour serum insulin (mu U/ml).
BMI	Body mass record (weight in kg/(tallness in m) ²).
Diabetes Pedigree Function	Diabetes family work.
Age	Age in years.

with higher risks of diabetic retinopathy and macular edema than poor glycemic control alone”, 12(6):e0180252, Jun 29 2017.

Table shows details of the datasets: (For example: Heart stroke disease)

References

- [1] Breast tumour detection using empirical mode decomposition features, Hongchao Song, Aidong Men and Zhuqing Jiang, 2017.
- [2] Scalable Healthcare Assessment for Diabetic Patients using Deep Learning on Multiple GPUs. Daniel Sierra-Sosa, Begonya Garcia-Zapirain, Member, IEEE, Cristian Castillo, Ibon Oleagordia, Roberto Nuño-Solinis, Maider Urtaran-Laresgoiti, Adel Elmaghraby, Life Senior Member, IEEE.
- [3] Guo-li DU, et al. "Metabolic Risk Factors of Type 2 Diabetes Mellitus and Correlated Glycemic Control/Complications: A Cross-Sectional Study between Rural and Urban Uygur Residents in Xinjiang Uygur Autonomous Region," PloS one, vol. 11, 2016.
- [4] Christian, BOMMER, et al. "The global economic burden of diabetes in adults aged 20–79 years: a cost-of-illness study," The lancet Diabetes & endocrinology, vol. 5, pp. 423-430, 2017.
- [5] L. M., et al. PEÑA-LONGBARDO, "Is quality of life different between diabetic and non-diabetic people? The importance of cardiovascular risks," PloS one, vol. 12, 2017.
- [6] Eva K., FENWICK, et al. "Combined poor diabetes control indicators are associated