AILMENT IMPLICATION FROM HEALTH-RELEVANT QUERY VIA KNN ALGORITHM

[1] Mrs.N.Deepika.M.Tech.,
Asst.Professor(CSE Department)
DMI College of Engineering
deepika@dmice.ac.in

[2] Padmapriya.T
Student (CSE Department)
DMI College of Engineering
Padma15jeni@gmail.com

[3] Mary Jeeva.M
Student (CSE Department)
DMI College of engineering
Maryjeeva.m@gmail.com

Abstract **- In this Internet Era, the costs of healthcare and burgeoning computer technologies are together driving more customers to spend longer time online to explore health information.To the best of our knowledge, this project is the first work on automatic disease inference in the community-based health services. Data mining is the process of analysing data from different perceptions and summarizing it into useful information. It investigates and categorizes the information needs of health seekers in the community-based health services and mines the related cases of their generated data.It proposes K-Nearest Neighbor learning scheme to infer various kinds of diseases. KNN stores all available cases based on the similarities. This scheme is pre-trained with pseudo-labeled data and further strengthened by fine-tuning with online doctor labeled data. It can be also used for analysis of a diseases cured, non-cured and viral diseases of the time.**

I. INTRODUCTION

Ailment implication from health-relevant query via KNN algorithm-In this method first performed user study to analyze the health seeker needs. This provides the insights of community based health services. It then presented a KNN algorithm that is able to infer the possible disease given the questions of health seekers.This scheme is very simple and easy which work incredibly well in practice. The KNN consists of set of vectors and class label associated with each vector. It works well with arbitrary number of classes.

Data mining is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. It is the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. KNN gives the words related to the case given. All the similar names of the diseases can be found easily from the data that has been stored.

II. MOTIVATION

This paper aims to build a disease inference scheme that is able to automatically infer the possible diseases of the given questions in community-based health services.To identify discriminant features for each specific disease, we propose a K-Nearest Neighbour scheme to infer the possible diseases given the questions of health seekers.

III. EXISTING SYSTEM

The paper aims to build a disease inference scheme that is able to automatically infer the

possible diseases of the given questions in community-based health services. Initially it analyzes and categorize the information needs of health seekers. As a byproduct, it is differentiated as questions of this kind that require disease inference from other kinds. It is worth emphasizing that large-scale data often leads to explosion of feature space in the lights of n-gram representations especially for the community generated inconsistent data. To avoid this problem, they have utilized the medical terminologies to represent our data. The scheme builds a novel deep learning model, comprising two components. The first globally mines the latent medical signatures. They are compact patterns of inter-dependent medical terminologies or raw features, which can infer the incomplete information. The raw features and signatures respectively serve as input nodes in one layer and hidden nodes in the subsequent layer. The second learns the inter-relations between these two layers via pre-training. Following that, the hidden nodes are viewed as raw features for more abstract signature mining.

i. Using deep learning to enhance cancer diagnosis and classification

Three different of sparse encoders have been used to learn features, sparse auto encoder which contains just one hidden layer, two layer stacked auto encoder, and stacked auto encoder with fine-tuning, which is trained based on a greedy layer-wise approach. Therefore gradually diagnosis the symptoms based on the data sets available in DB. To perform the task of cancer detection and cancer type classification, the features learned in the proposed unsupervised feature learning approach are subsequently used with a set of labeled data for specific cancer types to learn a classifier. For the results in this paper we used softmax regression as the learning**.**It is applicable only for the cancer diagnosis and is applicable for limited sizes of data sets.

ii. Bridging the vocabulary gap between health seekers and healthcare knowledge

These two components establishes a tri-stage framework to locally code each medical record. To extract all the noun phrases, we initially assign part-of speech tags to each word in the given medical record by Hospitals. We then pull out sequences that match a fixed pattern as noun phrases. The local mining approach may suffer from information loss and low precision, which are caused by the absence of key medical concepts and the presence of the irrelevant medical concepts.

iii. Learning to Combine Representations for Medical Records Search

This proposed framework that can effectively handle this combination using the Gradient Boosted Regression Trees to learn an effective combination model via retrieval performance predictors such as the clarity score and the query scope. We have shown that our proposed framework is effective for the medical records search, as it could markedly and significantly outperform an effective score combination approach. This is used just for matching up their given problems but yet doesn't give up a proper solution. Like just reading up the patient's effect for data particular symptoms just for data base.

iv. Representation learning: A review and new perspectives

This motivates longer term unanswered questions about the appropriate objectives for learning good representations, for computing representations (i.e., inference), and the geometrical connections between representation learning, density estimation, and manifold learning. Drawing connections between these approaches is currently a very active area of research and is likely to continue to produce models and methods that take advantage of the relative strengths of each paradigm. Although regularization effects can be important on small datasets, the effects that persist on very large

datasets suggest some optimization issues are involved. Are they more due to local minima and the dynamics of the training procedure? Or are they due mostly to ill conditioning and may be handled by approximate second order methods? These basic questions remain unanswered and deserve much more study.

v. A Multi-Task Learning Formulation for Predicting Disease Progression

They formulate the prediction problem as a multi-task regression problem by considering the prediction at each time point as a task. We study the feasibility of predicting AD progression measured by cognitive scores based on baseline measurements.They focus only on linear models. They failed to focus non-linear models.

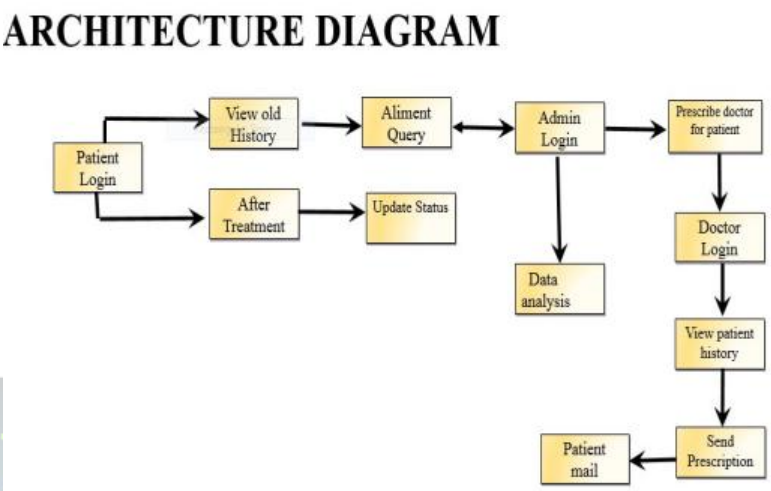
IV. ISSUES IN EXISTIN SYSTEM

A. The existing model are unable to identify discriminant features for each specific disease.
B. It uses deep Sparsing method that produces problem in understanding the exact condition of the patient.
C. It does not perform online doctor consultant. It produces output based on the database records.

V. PROPOSED SYSTEM

Our proposed system focuses mainly on online health seekers. We have introduced an online doctor support to enhance the efficiency of the application. In addition to that we have added an additional feature as auditing. In auditing we have used the K-Nearest neighbor algorithm to improve the performance. Using this new feature we would be able to identify the small and rapid changes in a particular area. We would be able to identify and resolve diseases affecting in an area at an earlier stage.

VI. SYSTEM ARCHITECHURE



A. QUERY MODULE

a. The user does registration by specifying their details like His/her Name, Contact Number, Address, and Mail Id etc. After validation the user will receive a message regarding his/her membership.
b. After registration the user can Login to this site with his/her Unique Id and password. The Unique Id is provided by Admin.
c. If the Unique Id and password is valid one time password (OTP) is sent to the patient's mail id. The Patient received that OTP in his/her mail Id.
d. After Enter that OTP the Patient Login into the Web Site.
e. The Patient enters the details about the problem, Duration of Symptom, Frequency, affected place etc.
f. All Prescription are maintains by the admin given by the doctor for future use. Patient can view their prescription at any time.
g. The Patient are also view all the medical details about Him.
h. The patient also update the status about the diseases it was cured or not.

B.ALLOCATION MODULE

The admin is the sole control of the application/project. The admin do some operation like view the patient and doctor personal details like name, address, phone no, etc., and the admin prefer the doctor for the patient.

Moreover Admin analyses all the database and perform addicting based on the entry of the patient records.

C.RECOVERY MODULE

a. Every Doctor will have their own unique Id and Password with which, they will login to this site.
b. In this module displays the doctors information.
c. After they logged into this site. They will have their main form. From there by choosing the link, they can see their appointments. He/she can see their new appointments and they can also see the previous appointments.
d. After that, the Doctor will send the details about the Patient disease, Medicine and diet through his/her mail id and the patient will receive the mail about the Medicine preferred by the Doctor for his/her treatments**.**

D.DATA ANALYSIS MODULE

In this module we have to analysis the following information. We have to use KNN (**K-nearest neighbor**) algorithm for clustering process. In this process the following questions are analyzed.

1. How many patients have come in with the same symptom?
2. Find the Best Doctor.

E.REPORT MODULE

In this module we can directly download the details of the patients, doctors information and auditing reports in the formats like Excel, pdf and word document. Admin maintain the all reports. It helps the consultants to get their own medical reports on hand that can be kept for their own references.

VII. CONCLUSION

This Online health care management system is a web-based application that assists in management of staffs, doctors and patients in easy, comfortable and effective service. Online Healthcare application for maintaining patient's information and doctor's information in a single application. This system will reduce manual work for maintaining records in files. The proposed application aims to create a friendly working environment for any health care centers and to overcome the drawbacks in existing system of health care management. This system is very reliable and flexible from all aspects, so new features and modules can be easily integrated into the system in future.