International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-6, Issue-6)

Research Article | **June 2017**

Prognostication Stereotype of Patients Morbidity and Mortality by Extraction of E-Health Records

[1]Sunitha.T, [2]Shyamala.J, [3]Annie Jesus Suganthi Rani.A
[1, 2]Assistant Professor, Department of CSE, P. B. College of Engineering, Tamilnadu, India
[3]Assistant Professor, Department of CSE, Dr.Sivanthi Aditanar College of Engineering, Tamilnadu, India

Abstract:

Data mining suggest an innovative way of prognostication stereotype of Patients health risks. Large amount of Electronic Health Records (EHRs) collected over the years have provided a rich base for risk analysis and prediction. An EHR contains digitally stored healthcare information about an individual, such as observations, laboratory tests, diagnostic reports, medications, procedures, patient identifying information and allergies. A special type of EHR is the Health Examination Records (HER) from annual general health check-ups. Identifying participants at risk based on their current and past HERs is important for early warning and preventive intervention. By "risk", we mean unwanted outcomes such as mortality and morbidity. This approach is limited due to the classification problem and consequently it is not informative about the specific disease area in which a personal is at risk. Limited amount of data extracted from the health record is not feasible for providing the accurate risk prediction. The main motive of this project is for risk prediction to classify progressively developing situation with the majority of the data unlabeled.

Index Terms—Health examination records, heterogeneous graph extraction.

I. INTRODUCTION

Huge amount of Electronic Health Records (EHRs) collected over the years have provided a rich base for risk analysis and prediction. An EHR contains digitally stored healthcare information about an individual, such as observations, laboratory tests, diagnostic reports, medications, procedures, patient identifying information and allergies. A special type of EHR is the Health Examination Records (HER) from annual general health check-ups. Identifying participants at risk based on their current and past HERs is important for early warning and preventive intervention. By ―risk‖, we mean unwanted outcomes such as mortality and morbidity.

Forecasting Health risks is the important requirement for the patients. In existing system the prior risk prediction from the patient disease severity and mortality is not accurate due to the limited information. The main objective of the project is health risk forecasting by comparing the new patient details with the existing patient records and deriving the results by analyzing the unlabeled data. We present results based on the matching patterns.

The goals, features, functions of the project are defined as follows. The aim of the project is to predict the health risks and mortality in advance by analyzing the existing health records by mining the patterns from unlabeled data. Hence this model is used in the health research field. In this project we identify the prior health risks by mining patient's morbidity and mortality by comparing their conditions with the existing health records. The existing patient records are stored in the health records storage. The new patient stores his health record in the storage. The doctor will check the existing patients who are dead are under risk from the records and he will also pass the new critical and expired patients details to the admin. The admin will update the new records. The patient will enter the query like his disease, age and other attributes like his mental condition. His disease and other attributes will be considered as an unlabeled data and will be analyzed with the existing records. We obtain the results by mining the meaningful pattern from the unlabeled data. The admin will take care of the analysis he will derive the patterns. The patterns consist of matching person records who died or under risk and we mention other important health condition. The admin will view the report and notify to the new patient about the health risks.

II. RELATED WORKS

In this section we review existing related studies, namelythose on mining health examination data and those on classification with unlabeled data in healthcare applications.

2.1 A Bayesian Perspective

As opposed to the standard classification and regression problems where a domain expert can provide labels for the data in a reasonably short period of time, training data in such longitudinal studies must be obtained only by waiting for the occurrence of a sufficient number of events. Survival analysis aims at directly predicting the time to an event of interest using the data collected in the past for a certain duration. However, it cannot give an answer to the open question of ―how to forecast whether a subject will experience an event by end of a longitudinal study using event occurrence information of other subjects at the early stage of the study?[1]. The goal of their work is to predict the event occurrence at

a future time point using only the information about a limited number of events that occurred at the initial stages of a longitudinal study. This problem exhibits two major challenges: (1) absence of complete information about event occurrence (censoring) and (2) availability of only a partial set of events that occurred during the initial phase of the study. They propose a novel Early Stage Prediction (ESP) framework for building event prediction models which are trained at the early stages of longitudinal studies. First, they develop a novel approach to address the first challenge by introducing a new method for handling censored data using Kaplan-Meier estimator. Then extend the Naive Bayes, Tree-Augmented Naive Bayes (TAN) and Bayesian Network methods based on the proposed framework, and develop three algorithms, namely, ESP-NB, ESP-TAN and ESP-BN, to effectively predict event occurrence using training data obtained at an early stage of the study. More specifically, our approach effectively integrates Bayesian methods with an Accelerated Failure Time (AFT) model by adapting the prior probability of the event occurrence for future time points.

2.2 Extending Association Rule Summarization Techniques

Early detection of patients with elevated risk of developing diabetes mellitus is critical to the improved prevention and overall clinical management of these patients. They aim to apply association rule mining to electronic medical records (EMR)[2] to discover sets of risk factors and their corresponding subpopulations that represent patients at particularly high risk of developing diabetes. Given the high dimensionality of EMRs, association rule mining generates a very large set of rules which need to summarize for easy clinical use. They reviewed four association rule set summarization techniques and conducted a comparative evaluation to provide guidance regarding their applicability, strengths and weaknesses. They proposed extensions to incorporate risk of diabetes into the process of finding an optimal summary. They evaluated these modified techniques on a real-world pre diabetic patient cohort. They found that all four methods produced summaries that described subpopulations at high risk of diabetes with each method having its clear strength. For their purpose, there extension to the Bottom-Up Summarization (BUS) algorithm produced the most suitable summary. The subpopulations identified by this summary covered most high-risk patients, had low overlap and were at very high risk of diabetes.

III. SYSTEM ANALYSIS

Our work focus on the data fusion for the health examination records to be integrated with other types of datasets such as the hospital-based electronic health records and the participants‘ living conditions (e.g., diets and general exercises). By integrating data from multiple available information sources, more effective prediction may be achieved. Although Electronic Health Records (EHRs) have attracted increasing research attention in the data mining and machine learning communities. The approach is limited due to the classification problem and consequently it is not informative about the specific disease area in which a person is at risk. Limited amount of data extracted from the health records is not feasible for providing the accurate risk prediction.The disadvantages are

- The information mined from the EHR is very limited.
- The prior risk prediction from the patient disease severity and mortality is not accurate.
- The classification model based on disease and risk prediction is not efficient.
- Only the labeled data are used for analysis but the majority consists of the unlabeled data which are unused.

IV. SYSTEM DESIGN

We propose a novel risk forecasting model by mining information from the large data sources. Our work focus on the data fusion for the health records to be integrated with other types of datasets such as the hospital-based electronic health records and the participants‘ living conditions (e.g., diets and general exercises). By integrating data from multiple available information sources, more effective prediction may be achieved. The advantages are

- Large amount of information is mined from the large HER data sources.
- The prior risk prediction can be made from the patient disease severity and mortality will be accurate.
- The classification model based on both the labeled and unlabeled data.

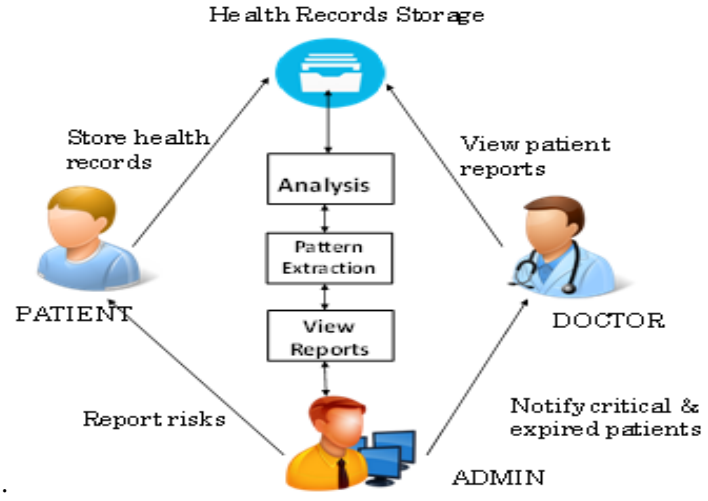

Fig 4.1 System Design

4.1 Health Record Updation

In this module the patients update their health records in the health storage. The patient should register first and then the patient will login with their credentials to update their information. The patient will update their disease along with their risk factors and other mandatory information. The patient can modify their information at any time. Some of the important factors to be filled are age, weight, BMI, BP, addiction and their mood. After the health information updation the corresponding doctor will view the patient records.

4.2 Report View

In this module the doctor will view the patient health records. The doctor must register their information before viewing the records. During the registration phase the doctor will mention their corresponding department along with other mandatory information. During login they will mention their department along with their login credentials. After successful login the doctor will view the patient details based on their department. They will compare the patient records along with the existing patient records and check for risks and deaths and notify admin.

4.3 Risk Analysis

In this module the risk is analyzed based on the health risks and mortality. Prior to risk analysis the admin will receive the notification from the doctor in case of health risks and death. The admin will update the health risks and deaths of the corresponding patients in the database. Based on the risks the patient records are classified. The risks are analyzed based on the their age, gender and type of the disease.

4.4 Pattern Extraction

In this module the corresponding patterns are extracted from the health storage. The admin will update the health records before the pattern extraction. The patterns are extracted based on the unlabeled data .Here we extract the patterns based on the health risks. We find all the patients details with matching patterns. Here we implement a classification based on the patterns. The patient details are extracted as patterns based on their disease and age.

4.5 Report Generation

In this module the report is generated based on the health risks. After pattern extraction we will match the new patients with the existing patients who are already died or having health risks. We will generate the report based on that. The report consists of the patient who is under risk and who are dead we will also mention the risks factors along with that. So that new patients can check their status along with the existing patient under risks. We also plot a graph based and notify the new patient. We plot a Pie Chart for the graphical representation of the age vs. disease by considering the risk factors.

4.6 Patient Alert

After report generation the corresponding patients will be alerted who might have health risks in the future. They will view the report of the existing users who are already died and under health risks. They can compare their information with the existing users and correct their health risks. They can protect themselves by taking care of their blood pressure level, monitoring their moods, controlling their BMI and getting rid of their addiction.

V. EXPERIMENTAL RESULTS

5.1 Algorithm For Heterogeneous Graph Extraction.

- Start by Obtained the Patient Records
- Store the Patient Information in the Database during Patient Registration
- Check the Patient Credentials from the Database during Patient Login
- Navigation to the Patient Account page after successful verification
- Store the Disease information of the patient in database
- Allow Patient to update the disease information anytime
- Show the corresponding Doctor details to the Patient
- Allow Patient to report his Age, Gender and Disease to the Admin
- Store the Doctor Information in the Database during the Registration
- Check the Doctor Credentials from the Database during Doctor Login
- Navigation to the Doctor Account page after successful verification
- Show the Patient records to the Doctor
- Verification of the Patient records under Risk and Death
- Allow Doctor to report the Risk and Death status to the Admin
- Authentication of Admin credentials during Login.
- Navigation to the Admin Account page after successful verification
- Show the Patient information under Risk and Death
- Allow Admin to update the status of the patient
- Show the Patient requests to the Admin

- Allow Admin to analyze the Patient status based on Age, Gender and Disease
- Pattern matching from the Database based on Age and Disease
- Pattern matching from the Database based on Gender and Disease
- Pattern Extraction from the Database based on Age and Disease
- Pattern Extraction from the Database based on Gender and Disease
- Notify the Admin if the new Pattern matches the Existing Patterns under Risk and Death
- Allow Admin to alert the Patient through Mail when Patient is in Danger
- Report will be viewed by the Patient
- Show the Existing unlabeled data to the Patient by pattern matching
- Comparison of the Patient Health condition with the existing records
- Check the parameters of Disease, BMI, Mood, BP and Addiction
- Patient correct his health conditions to be safe Graphical representation of the Age and Disease under Risk and Death
- Graphical representation of the Mood and Disease under Risk and Death
- Let X1,X2,X3…Xn be the Patients
- Xd is the Dead Patients
- Xs is the Patients with severe conditions
- Ad =Disease Analysis
- Ad = (X1,X2,X3…Xn) -> (Xd, Xs)
- If (X1,X2,X3…Xn) € (Xd, Xs)
- then Alert (X1,X2,X3…Xn)

5.2 Execution Steps for Heterogeneous Graph Extraction.

1. This is the first stage which is known as patient registration
2. New Patient has to register in the web application
3. By filling all the details that are needed
4. Existing patients or old patients can login the application by giving their username and password.
5. The general details that has to be given by the patients like name, mobile number ,email id,gender,weight,body mass index,address,type of mood and disease.
6. Here we consider only the main deadly disease from which most of the people suffer.
7. The Department of disease that we consider are diabetology,oncology,neurology,
8. After the patient registration, patient can view the doctor details,then all the reports will be send to the appropriate doctor.
9. The patient can update their health details by themselves the updated details will be directly send to the doctor.
10. A large number of datasets will be maintained by the admin
11. Datasets consist of labeled and unlabelled data.
12. Labeled data has already classified and taken into consideration but unlabelled data was not taken.
13. The unlabelled data attributes are a general group of datas that has not been divided on the basis of their category
14. some of the examples of unlabelled data attributes are name,age,gender,mood and hobbies.
15. The second stage is the Doctor login
16. Doctor login using their username and password which is a authenticated one.
17. Doctor view the reports of the patient based on their department
18. By examining their reports doctor will send a message that the concern person is under risk or death or safe to the admin.
19. Admin is the one who updates the patients health status.
20. After receiving the doctor report,admin will also check the patients report on the basis of 2 attributes age and mood.

Table 5.1 shows the various diseases along with the status of the patient with their age and mood.

Disease	**Age**	**Mood**	**Status**
Ischemic Heart Disease	39	Stress	Risk
Stroke	40	Aggression	Risk
Diabetes Mellitus	30	Stress	Risk
Malignant Tumors	75	Anxiety	Dead
Lung Cancer	35	Anxiety	Dead

Graphical representation of the disease and age who are under health risk.

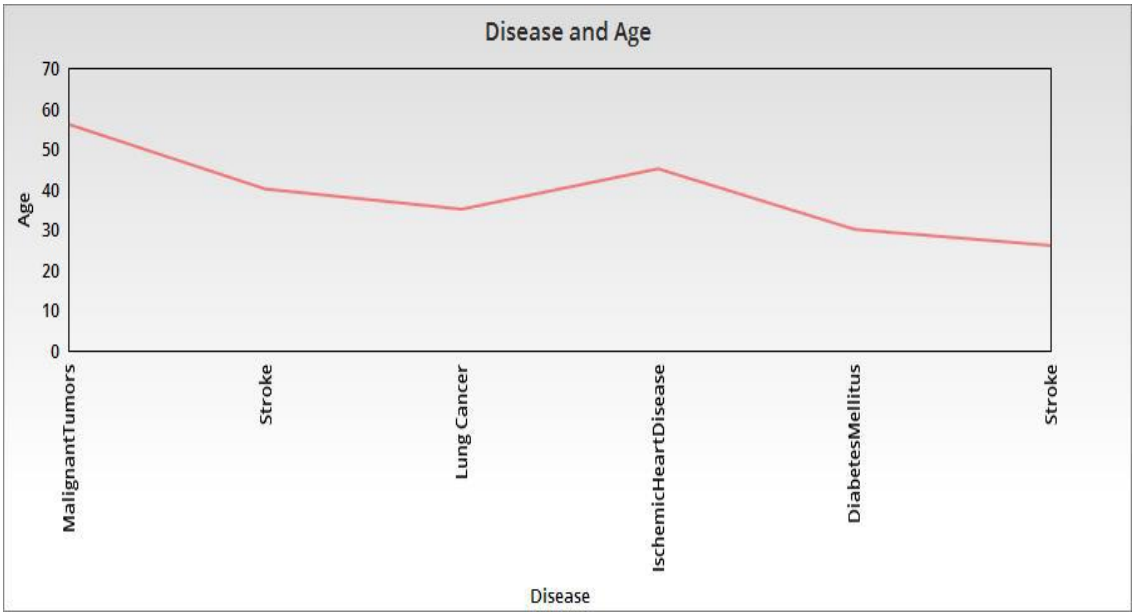

Fig 5.1 Age and Disease Graph

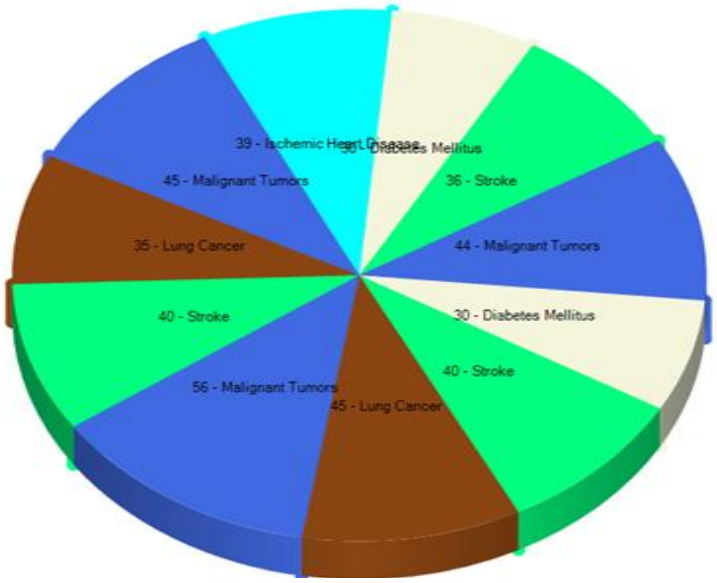

Fig 5.2 pie chart for disease and age

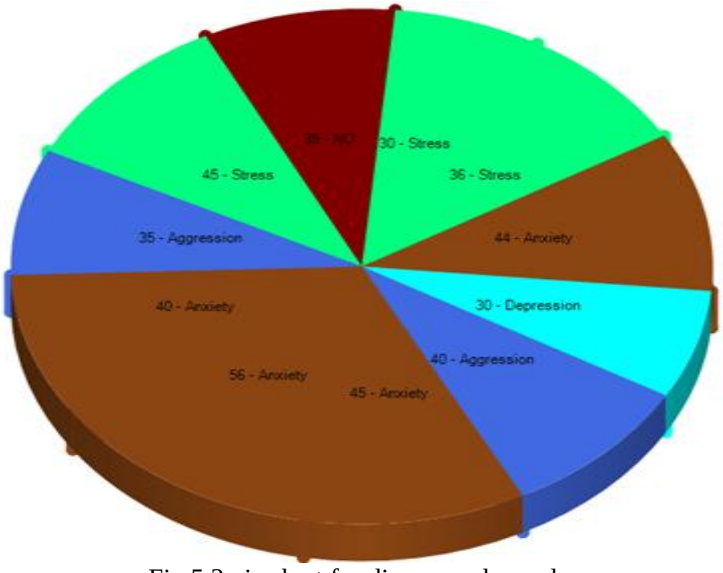

Fig 5.3 pie chart for disease and mood

VI. CONCLUSION

We forecast the health risks in advance and alert the patient .We will compare the new patient health condition with the existing records and notify the patient in case of danger. The patient will take precaution and correct his health condition in advance.As the future enhancement we will classify the health condition based on all the diseases and include more parameters for the comparison. We further improve security by implementing cryptographic technique for securely storing the data in health storage.

REFERENCES

[1] A Bayesian perspective on early stage event prediction in longitudinal data Mahtab J. Fard, Ping Wang, Sanjay Chawla, and Chandan K. Reddy, Senior Member, IEEE,2016

[2] Cog Boost- Boosting for fast cost sensitive graph classifications S. Pan, J. Wu, and X. Zhu, 2015

[3] A Relative similarity based method for interactive patient risk prediction B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang., 2015

[4] Mining personal health index from annual geriatric medical examinations L. Chen, X. Li, S. Wang, H.-Y. Hu, N. Huang, Q. Z. , and M. Sharaf., 2014

[5] Extending association rule summarization techniques to assess risk of diabetes mellitus , G. J. Simon, P. J. Caraballo, T. M. Therneau, S. S. Cha, M. R.Castro and P. W. Li., 2013

[6] Varun Chandola and Vipin Kumar. Summarization – compressing data into an informative representation. *Knowledge and Information Systems*, 2006.

[7] Gary S Collins, Susan Mallett, Omar Omar, and Ly-Mee Yu. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Medicine*, 2011.

[8] Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England Journal of Medicine*, 346(6), 2002.

[9] Gang Fang, Majda Haznadar, Wen Wang, Haoyu Yu, Michael Steinbach, Timothy R Church, William S Oetting, Brian Van Ness, and Vipin Kumar. High-order snp combinations associated with complex diseases: efficient discovery, statistical power and functional interactions. *PLoS One*, 7(4):e33531, 2012.

[10] Mohammad Al Hasan. Summarization in pattern mining. In *Encyclopedia of Data Warehousing and Mining, (2nd Ed)*. Information Science Reference, 2008.

