Analysis of various techniques for Knowledge Mining in Electronic Health Records

1V.Deepa, 2P.Mohamed Fathimal

1Research Scholar, 2Assistant Professor, SRM Institute of Science & Technology, Vadapalani, Chennai

Abstract

With the fast development of digital communication all over the world, each domain including healthcare sector has led to a new dimension. “Health Informatics,” are the term used to coin application of IT for better healthcare services. Its applications help to maintain the health record of individuals, in digital form known as the Electronic Health Record. This paper reviews the Electronic health records (EHR) in healthcare organizations. These digital records can help to support clinical activities that have the ability to improve quality of health and to reduce the costs. The aim of this paper is to study the detail analysis of the structure and the components of electronic health record. This paper reviews different techniques for handling the information in the text and different components of the electronic health record include daily charting, physical assessment, medication, discharge history, physical examination and test procedures.

Keywords: Physical Assessment, Survey medication, Semantics, patience, Information Quality, Health care policy planning

1 INTRODUCTION

In healthcare domain, more than 1100 electronic health record vendors are available all over the world. National policy and guidelines have been framed for EHR dealers and medical benefits. EHR collects real time medical data which stores patient’s information in a digital format like a history chart. As all the information are accessible in a single file, extraction of medical data from EMRs (electronic medical records) for analysis is very effective for the analysis.

Alberta Netcare is a public electronic health record system which allows the secured medical practitioners across the state to visualize the health records. The government of Alberta is getting ready to issue the state their electronic health records. “Birth-to-death” electronic health record system was available for every citizen for good health care. Various state privacy laws will continue to stop the efficiency and development of health information exchange (HIE) and in Cyber security.

EHR in India

The evolution of the Health Information Security and Privacy Collaboration (HISPC) is the initial step towards the successful tie-up between states and planning the last stage of operation between different EHR systems with the privacy policy.

1.PTS is an unique cloud based patient management application that performs the careful handling of the health care data. The entire workflow has been taken from a medical practitioner patient point of view. Starting with e-referrals, web based appointment calendar the scheduling was designed with the user interface. The digitalisation was carried out using the cloud based Application.

Some of the vendors currently available are Epic, Allscripts, cureMD, E-clinical works, Cerner, Practice Fusion, Athena Health and Alberta Netcare. Some of the commercial EHR is CHITS, Clear-Health, Cottage med, FFEHR, HosXP, Indivo Health, Medclipse, ‘MED’ in Tux.
Mirrormed, Open EMR, Open MRS, Open VISTA, Oscar McMaster, Patient OS, WorldVISTA, ZEPRS, Mirth, Tolven and GNU Health.

Basic Structure of EHR:

The Electronic health record can be represented in different structures for different purpose: daily charting, physical assessment, nursing care plan, medication monitoring, referrals, nursing information, care plan for diagnosis and prognosis of medical data.

The EHR data consist of following types such as Administrative and billing data, Patient demographics, Progress notes, Vital signs, Medications, Immunization dates, Allergies, Radiology Images, Lab and Test results.

It offers retrieval of the evidence-based tools that helps in making decisions about the patient's health care data. It automates and regularize the health care practitioner workflow. It supports the main market changes in payer needs and the consumer expectations. The EHR data is used in health care domains such as medication and diagnostic procedures.

1. Patient identification information.
2. The health provider's identification information.
3. Care episode data.
4. Health patterns.
5. Nursing minimum data set.

Surgical procedures.

6. Tests and examinations.
7. Information about medication.
8. Preventive measures.

Types of EHR: The different types of the EHR are Custom EHR, Proprietary EHR, Open-source EHR. The custom EHR is the solution that is specifically developed for the specific Health care Organization. The Proprietary EHR is the software that is privately owned by the rules of the proprietor. Open source EHR is computer software that is distributed with its source code available information.

Input to EHR

The Data is fed to the EHR in different formats such as follows

Structured Data

The Healthcare system extracts the data such as lab-results, allergies, social information and other important signs in a structured tabular format.

Unstructured Data:

Clinical notes is an example of the unstructured data which offers maximum flexibility of the data. Clinical data mostly stores free text type of the data.

Process of EHR:

1) Data cleaning.
2) EHR database setup.
3) Generate the raw data in database fields.
4) Construct the Health care Process Model
5) Testing and verification of new data Inputs.
6) Create the workflows of the process.
7) Generate the electronic data for the paper work

Analysis of EHR:

Electronic health records can enhance the growth and output of patient data analysis for medical prediction by performing analysis and computing data within the system. A high quality, reliable system will be able to collect huge amounts of patient data quickly, which is easily sorted and analyzed. Some of the reports are

1. Appointments and scheduling reports:
2. Allergy, medication, and test alerts:
3. Outcome measurement reporting
4. Billing reports

. The medical experts gets the relevant data from the Prediction Model of the EMR Model.

The digital record generate a model view by taking the patient population for a particular year Y which allows collecting samples for the specific period. The medical history from Y to (Y+5) and the outcomes are developed. The analysis matrix is constructed to record outcome for D as well as the research related tasks. The matrix with the patients as its rows and the behaviour of the patient characteristics in year Y as columns will be generated

Machine learning techniques in HER

Machine learning methods is used in EHR in order to generate a predictive model in EHR. Machine-learning algorithms is divided into 4 categories.

(i) Supervised learning algorithm is mainly used for the numeric prediction such as linear regression, K-Nearest Neighbors, gradient boosting or classification such as logistic regression, random forest, support vector machine, decision tree or the Neural Network analysis.

(ii) Unsupervised learning algorithm is used for generating a description and association rule learning algorithms, clustering methods such as K-means, vector space model, these algorithms help us to learn the data representation for dimensionality reduction.

(iii) Semi-supervised learning algorithm is used for generating a predictive model which is mainly used for clustering technique. (iv) Reinforcement learning algorithm is used for generating a predictive model which is trained with a set of samples. Clustering technique clustering techniques helps in the observation, analysis of the patient data. Clustering helps in finding the distance between the patient records.

Association Analysis helps in predicting the diseases such as a novel approach to identify the associations between the set of data. The statistical approaches help in detecting the associations between the rare set of diseases.

II RELATED LITERATURE

The following section listed various analysis of EHR techniques used by researchers to improve the design of EHR.
### TABLE III COMPARISON OF DIFFERENT TECHNIQUES USED FOR EHR MODELS

<table>
<thead>
<tr>
<th>S. No</th>
<th>Model Name</th>
<th>Data Set Source</th>
<th>Approch used</th>
<th>Data Preprocessing Technique</th>
<th>Feature Extraction</th>
<th>Data Mined</th>
<th>Mining Techniques</th>
<th>Analysis of What Outcome</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>eNRBM[2], Deep Patient[3]</td>
<td>UK Population (Snow Data set)</td>
<td>Image AI approach</td>
<td>Clinical Data</td>
<td>PCA</td>
<td>Temporal</td>
<td>Clustering</td>
<td>Analyse the preprocessing data.</td>
<td>Temporal nature of the EHR.</td>
</tr>
<tr>
<td>2</td>
<td>Dynamic Bayesian Networks Model</td>
<td>US Population Data set</td>
<td>Image AI approach</td>
<td>Clinical Data</td>
<td>PCA</td>
<td>Prediction</td>
<td>Association</td>
<td>Risk Prediction Model</td>
<td>Casual Analysis was performed</td>
</tr>
<tr>
<td>3</td>
<td>Bayesian Recurrence Neural Network</td>
<td>MIMIC-III and ICU clinical datasets</td>
<td>Text AI approach</td>
<td>Clinical Data</td>
<td>PCA</td>
<td>Temporal</td>
<td>Association</td>
<td>Analysis of the uncertainty of the data</td>
<td>The goal is to maximize the precision until the sensitivity level is reached</td>
</tr>
<tr>
<td>4</td>
<td>Genome screening model</td>
<td>Genome Health Data set</td>
<td>Image AI approach</td>
<td>Clinical Data</td>
<td>PCA</td>
<td>Descriptive</td>
<td>Association analysis</td>
<td>To evaluate the complex disease such as Breast cancer</td>
<td>Genome technique can be used predict the complex disease.</td>
</tr>
<tr>
<td>5</td>
<td>PCEHR system Model</td>
<td>Access Control Security Data</td>
<td>Text AI approach</td>
<td>Clinical Data</td>
<td>PCA</td>
<td>Temporal</td>
<td>Associative Analysis</td>
<td>To minimise the misuse or abuse within healthcare</td>
<td>The outcome is to provide a secure Model for EHR</td>
</tr>
<tr>
<td>No.</td>
<td>Methodology</td>
<td>Population</td>
<td>Data Approach</td>
<td>Data Type</td>
<td>Prediction</td>
<td>Analysis</td>
<td>Outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------------</td>
<td>------------</td>
<td>--------------</td>
<td>-----------</td>
<td>------------</td>
<td>----------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Decision Rule Model</td>
<td>Cardiovascular population</td>
<td>Text Tree Approach</td>
<td>Clinical Data</td>
<td>LDA Predictio n</td>
<td>Classification of set of Cardiovascular Population.</td>
<td>The Percentage of the UGC Patients were analyzed.</td>
<td>Outcome is negative and positive predictive value.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Bayesian Network model is used</td>
<td>US Population Image AI approach</td>
<td>Clinical Data</td>
<td>LDA Predictio n</td>
<td>Prediction - Diagnosing the Diabetic Patients.</td>
<td>Analysing the Diabetic Associated Complications.</td>
<td>The main advantage of this set of techniques is their capability to handle causation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Data warehousing model</td>
<td>Stanford hospital –uk population Text Tree Approach</td>
<td>Clinical data</td>
<td>LDA Predictio n</td>
<td>Association Analysis</td>
<td>Analysing the Free text in order to get accurate information</td>
<td>Diagnosing the rare diseases were done To improve the patient care management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Deep learning models</td>
<td>PRIME Data Set for UK Population Image AI approach</td>
<td>Clinical data</td>
<td>PCA Predictio n</td>
<td>Association Analysis</td>
<td>To analyse the Blood pressure of the coronary artery Disease.</td>
<td>It can learn the weights from the different prior medical knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>EHR prediction models</td>
<td>PRIME Dataset for Heart rate Text Tree Approach</td>
<td>Clinical Data</td>
<td>PCA Temporal</td>
<td>Association Analysis</td>
<td>Analysing the Heart rate Failure of the different set of Patients</td>
<td>The outcome is to analyse the pattern of the disease.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Method</td>
<td>Dataset/Approach</td>
<td>Clinical Data</td>
<td>PCA</td>
<td>Predictive Analysis</td>
<td>Association Analysis</td>
<td>Analysis</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>-------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>----------</td>
<td>-----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Deep Neural network model</td>
<td>MIMIC III dataset (Hospital database)</td>
<td>AI approach</td>
<td>Clinical data</td>
<td>Predictive Analysis</td>
<td>Association Analysis</td>
<td>Analysing the visual data tools</td>
<td>Analyse the technique for big data for ehr.</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Convolution Neural network model is used</td>
<td>Pima Indians Diabetes Dataset (PIDD) for Diabetes</td>
<td>Tree approach</td>
<td>Clinical Data</td>
<td>Prediction</td>
<td>Association Analysis</td>
<td>Analysing the Diabetic Patients</td>
<td>The different parameters like ‘Age’, ‘Family diabetes’, ‘Physically active’, ‘Regular Medicine’ were used to predict the Outcome.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>EHR based Interpretation model</td>
<td>New York medical data</td>
<td>Text approach</td>
<td>Clinical Data</td>
<td>Prediction</td>
<td>Association Analysis</td>
<td>Analysing the architecture of the EHR</td>
<td>Data was analysed using the association analysis.</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Predictive Distributive model in EHR data</td>
<td>MIMIC-III using the NLP Technique</td>
<td>Structure</td>
<td>Clinical Data</td>
<td>Prediction</td>
<td>Association Analysis</td>
<td>Analysing the Multi task clinical task.</td>
<td>To determine the uncertainty of the Predictive disagreement.</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Machine learning Model</td>
<td>MIMIC III Dataset using the NLP Tech</td>
<td>AI approach</td>
<td>Clinical Data</td>
<td>Prediction</td>
<td>Association Analysis</td>
<td>Analysing the predictive Model</td>
<td>A model for rheumatoid arthritis generalized strategy.</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>PCEHR system Model</td>
<td>Security Dataset</td>
<td>Text approach</td>
<td>Clinical Data</td>
<td>Prediction</td>
<td>Association Analysis</td>
<td>Analysing the access control strategies</td>
<td>Four security access control techniques were identified for EHR.</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Decision Tree Model (CPCSSN) database Canadian Diabetes Association (CDA), Image</td>
<td>Tree Approach</td>
<td>Clinical Data</td>
<td>PCA Prediction</td>
<td>Association Analysis</td>
<td>Analysing the Cardiovascular disease of the Canadians.</td>
<td>Ensemble approaches can be applied to other diseases like hypertension, coronary heart disease and dementia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Neuro Fuzzy Model Pima Indian Diabetes (PID) dataset, Image AI approach</td>
<td>Clinical Data</td>
<td>LDA Prediction</td>
<td>Classification technique</td>
<td>Analysing the breast cancer data.</td>
<td>Medical Disease classification using the Fuzzification process</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IV CONCLUSIONS AND FUTURE WORK

This paper specifies the various stage of the electronic health record and their data format has been studied. Literature review of different algorithms used for analysis of data in electronic health record was studied. The various algorithms listed here will help to find the

REFERENCES


17. Myra Spiliopoulou and panagiotis,"Mining and Model Understanding on te Medical Data “in the proceeding of the ACM,pp 112-119,2020


28. Himansu Dasa, Bighnaraj Naik, H.S. Behera “Medical Disease Analysis using the Neur Fuzzy with the Feature Extraction Model for classification” in the Proceeding of Elsvier, pp. 131-145, 2019

