Analysis of various techniques for Knowledge Mining in Electronic Health Records

[1]V.Deepa, [2]P.Mohamed Fathimal

[1]*Research Scholar , [2] Assistant 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   new dimension . "Health Informatics," are the term used to coin application of IT for better healthcare services. Its applications helps 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

I 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 the 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 expectationsThe 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.
9. Patient data for Analysis.

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 proprieter.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

S. No	Model Name	Data Set Source	Data used for Analysis	Approach used	Data Preprocessing Technique	Feature Extraction	Data Mined	Mining Techniques	Analysis of What	Outcome
1	eNRBM[2], Deep Patient[3] Deepr[4],RETAIN [5]	UK Population (Snow Data set)	Image	AI approach	Clinical Data	PCA	Temporal	Clustering	Analyse the preprocessing data.	Temporal nature of the EHR .
2	Dynamic Bayesian Networks Model	US Population Data set	Image	AI approach	Clinical Data	PCA	Prediction	Association	Risk Prediction Model	Casual Analysis was performed
3	Bayesian Recurrence Neural Network	MIMIC-III and ICU clinical datasets	Text	AI approach	Clinical Data	PCA	Temporal	Association	Analysis of the uncertainity of the data	The goal is to maximize the precision until the sensitivity level is reached
4	Genome screening model	Genome Health Data set	Image	AI approach	Clinical Data	PCA	Descriptive	Association analysis	To evaluate the complex disease such as Breast cancer	Genome technique can be used predict the complex disease.
5	PCEHR system Model	Access Control Security Data	Text	AI approach	Clinical Data	PCA	Temporal	Associative Analysis	To minimise the misuse or abuse within healthcare	The outcome is to provide a secure Model for EHR

		set								organizations	
6	Decision Rule Model	Cardiovascular population	Text	Tree Approach	Clinical Data	LDA	Prediction	Prediction – classification of set of Cardio vascular Population.	The Percentage of the UGC Patients were analyzed .	Outcome is negative and positive predictive value.
7	Bayesian Network model is used	US Population	Image	AI approach	Clinical Data	LDA	Prediction	Prediction - Diagnising the Diabetic Patients.	Analysing the Diabetic Associated Complications.	The main advantage of this set of techniques is their capability to handle causation.
8	Data warehousing model	Stanford hospital –uk population	Text	Tree Approach	Clinical data	LDA	Prediction	Association	Analysing the Free text in order to get accurate information	Diagnosing the rare diseases were done To improve the patient .care management
9	Deep learning models	PRIME Data Set for UK Population	Image	AI approach	Clinical data	PCA	Prediction	Association Analysis	To analyse the Blood pressure of the coronary artery Disease .	It can learn the weights from the different prior medical knowledge
10	EHR prediction models	PRIME Dataset for Heart rate	Text	Tree approach	Clinical Data	PCA	Temporal	Association Analysis	Analysing the Heart rate Failure of the different set of Patients	The outcome is to analyse the pattern of the disease.

11	Deep Neural network model	MIMIC III dataset(Hospital database)	Images	AI approach	Clinical data	PCA	Predictive	Association Analysis	Analysing the visual data Tools	Analyse the technique for big data for ehr.
12	Convolution Neural network model is used	Pima Indians Diabetes Dataset (PIDD) for Diabetes	Text	Tree approach	Clinical Data	PCA	Prediction	Association Analysis	Analysing the Diabetic Patients	The different parameters like 'Age', 'Family diabetes', 'Physically active', 'Regular Medicine' were used to predict the Outcome .
13	EHR based Interpretation model	New York medical data	Text	Partition Approach	Clinical Data	PCA	Prediction	Association Analysis	Analysing the architecture of the EHR	Data was analysed using the association analysis.
14	Predictive Distributive model in EHR data	MIMIC-III using the NLP Technique	Structure	AI approach	Clinical Data	PCA	Prediction	Association Analysis	Analysing the Multi task clinical task.	To determine the uncertainty of the Predictive disagreement.
15	Machine learning Model	MIMIC III Dataset using the NLP Tech	Text	AI approach	Clinical Data	PCA	Prediction .	Association Analysis	Analysing the predictive Model	A model for rheumatoid arthritis generalized strategy.
16	PCEHR system Model	Security Data set	Text	Partition Approach	Clinical Data	PCA	Prediction	Association Analysis	Analysing the access control strategies	Four security access control techniques were identified for EHR

| 17 | Decision Tree Model | (CPCSSN) database Canadian Diabetes Association (CDA), | Image | Tree Approach | Clinical Data | PCA | Prediction | Association Analysis | Analysing the Cardio vascular disease of the Canadians . | Ensemble approaches can be applied to other disease like hypertension, coronary heart disease and dementia |
| 18 | Neuro Fuzzy Model | Pima Indian Diabetes (PID) dataset | Image | AI approach | Clinical Data | LDA | Prediction | Classification technique | Analysing the breast cancer data. | Medical Disease classification using the Fuzzification process |

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

1. Jose Roberto Ayalo solares and Yajie Zuhu,"Deep learning for Electronic health records: A comparative review of multiple deep neural architecture" in Elsvier Proceeding Vol 22, pp112-119,2020.
2. T.Tran,T.D.Nguyen,D.Phung and S.Venkatesh,"Learning vector representation via medical objects via EMR-driven non negative restricted Boltzman machines"pp. 96-105,2015.
3. R.Miotto,L.Li,B.A.Kidd J.T.Dudley,"Deep Unsupervised representation to predict the future of patients from the electronic health records",vol.6,pp 34-68,2016.
4. P.Nguyen,T.Tran,N.Wickrama singhe,S.Venkatesh,"Deepr:a convolutional net form medical records",IEEE,pp 22-30,2017
5. E.Choi,M.T.Bahadori,J.Kulas,A.Schuetz,W.Stewart,Reta in "An Interpretable Predictive Model for Healthcare using Reverse Time Attention", pp. 3504–3512,2016
6. Pranjul Yadav, Michael Steibach, Vipun Kumar and Gyorgy Simon, "A Survey on the mining of the Health records"in the proceedings of ACM Magazine, vol. 14, no. 6, pp. 2018.
7. Michael W.Dusenberry,JonasKemp and Edward Choi, "Analysing the Role of Model Uncertainity in Electronic Healt h records"ACM Proceeding pp. 117–122, Oct. 2020.
8. Naura,Abul hassan,Elmear Kenny "Personalised Medicine and the Power of. Naura,Abul hassan,Elmear Kenny , "Personalised Medicine and the Power of the Electronic Health recoed", in the Proceeding of Elsvier,pp. 92-93,2019.
9. Hayirenn,Kaija Sorento and   Pirkko Nainanen , "Definition, content, structure and the Impact of the Electronic Health records "in the Elsvier proceeding ,pp. 90–100, Feb. 2008.
10. Enja Blessie markDe Kroot Titea and Liley Maltz,"Data Mining information from electronic Health records produced high yield and accuracy from current smoking status" in the proceeding Elsvier pp 131-145 2020.
11. Pranjul Yadav,Vipin Kumar and Gyorgy Simo "Mining of Electronic Health Record(EHR) A Survey" in the proceeding of ACM ,vol. 7, pp. 47–49,
12. Nicholas Garcelon Anita Bungun and Remi Solomon,"Electronic health records:A comparative review of multiple deep neural architecture" in Elsvier Proceeding Vol 22,pp112-119,2020.
13. Fenglong ma Jing Gao,Quiling Suo.M Kasermann and D. Vollmer, "Risk   Prediction of the ehr with prior medical Knowledge" vol 7, no.1,pp.47-49,Jan 2018.

14.

15. Pooja H.Prabudev Jagadeesh,"A collective dataMining Techniques for the Electronic Health Records "in the proceeding of the IEEE vol 7,pp47-49,2019.

16. John woong,Kim,Beakon and jachel,"Deep learning for the Electronic Health Record mining in Analysis "in the processing of IEEE,pp-90-100,Feb2019.

17. Neha Prema Tiggaa and Shruti Garg Prediction   of the Type-2 Diabetes using the Machine learning classification Methods" in the ACM proceeding ,pp 131-145,2019.

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

19. Michael W.Dusenberry JonasKemp and EdwardChoi "Analyzing the Role of Model Uncertainity for Electronic Health Records" in   the proceeding of the ACM pp:156 167,Mar2020.

20. Jen J Gong, Tristann Naumann and pete Zovoltis,"Predicting Clinical Outcome across Changing Electronic Health Record Systems   ",in the proceeding of ACM pp112-119,2017.

21. Pasupathy Vimalachandran ,Hua Wang and Yanchunn Zhang,"Securing Electronic Medical Record Systems through an Improved Access Control" in the proceeding of the Springer pp131-145 ,2018.

22. Sajida Perveena,Muhammed Shahbaza,Azi Guerga chib and Karim data through Keshavjeec,"Performance analysis of the Data Mining classification techniques to predict Diabetes "in the proceedings of the Elsvier,pp131-145,2016.

23. Himansu Dasa,Bighnaraj Naikb,H.S.Behera "Medical Disease Analysis using the Neur Fuzzy with the Feature Extraction Model for classification "in the Proceeding of Elsvier,pp131-145,2019

24. S. K. Nataraj, F. Al-Turjman, A. H. Adom, R. Sitharthan, M. Rajesh and R. Kumar, "Intelligent Robotic Chair with Thought Control and Communication Aid Using Higher Order Spectra Band Features," in IEEE Sensors Journal, doi: 10.1109/JSEN.2020.3020971.

25.   B. Natarajan, M. S. Obaidat, B. Sadoun, R. Manoharan, S. Ramachandran and N. Velusamy, "New Clustering-Based Semantic Service Selection and User Preferential Model," in IEEE Systems Journal, doi: 10.1109/JSYST.2020.3025407.

26.   Ganesh Babu, R.; Obaidat, Mohammad S.; Amudha, V.; Manoharan, Rajesh; Sitharthan, R.: 'Comparative analysis of distributive linear and non-linear optimised spectrum sensing clustering techniques in cognitive radio network systems', IET Networks, 2020, DOI: 10.1049/iet-net.2020.0122

27.   Rajalingam, B., Al-Turjman, F., Santhoshkumar, R. et al. Intelligent multimodal medical image fusion with deep guided filtering. Multimedia Systems (2020).

28. Himansu Dasa,Bighnaraj Naikb,H.S.Behera "Medical Disease Analysis using the Neur Fuzzy with the Feature Extraction Model for classification "in the Proceeding of Elsvier,pp131-145,2019

29. S. K. Nataraj, F. Al-Turjman, A. H. Adom, R. Sitharthan, M. Rajesh and R. Kumar, "Intelligent Robotic Chair with Thought Control and Communication Aid Using Higher Order Spectra Band Features," in IEEE Sensors Journal, doi: 10.1109/JSEN.2020.3020971.

30. B. Natarajan, M. S. Obaidat, B. Sadoun, R. Manoharan, S. Ramachandran and N. Velusamy, "New Clustering-Based Semantic Service Selection and User Preferential Model," in IEEE Systems Journal, doi: 10.1109/JSYST.2020.3025407.

31. Ganesh Babu, R.; Obaidat, Mohammad S.; Amudha, V.; Manoharan, Rajesh; Sitharthan, R.: 'Comparative analysis of distributive linear and non-linear optimised spectrum sensing clustering techniques in cognitive radio network systems', IET Networks, 2020, DOI: 10.1049/iet-net.2020.0122