An Overview on Disease Prediction for Preventive Care of Health Deterioration

Mohan Kumar K N, S.Sampath, Mohammed Imran

Abstract: Machine learning in health care has recently made headlines. With the wide spread increase of population, the need for reliable mechanism to prevent diseases has increased in manifold. In the recent days there is an increase in health problems in majority of the population across the globe. The reason for health problems is not specific but it has become very uncertain. If we take a sample from the population, it should not be a surprise to see a person suffering from ailments irrespective of age and quality of life. For example chronicle diseases are found in people at a very young age. So this situation poses a serious challenge for clinical experts to find the root cause. It is difficult to accurately predict the future health based on the current health status because the scenario might not be same for all the patients. Providing an affordable, high quality health care service has become a big challenge. In this regard, preventive care of diseases is investigated for decades. It is an area of regular extension of research works and progression day by day and there is sufficient literature available on prediction of diseases. Our work includes a disciplined study to consolidate existing works on prediction and classification of diseases. This paper will provide technical insight and paves way for future developments in the health care field.

Index Terms: Machine Learning, Health, Entropy, Disease Prediction, Diabetes, features, metrics,ICD, HCC.

I. INTRODUCTION

Preventive care of Health Deterioration (PCHD) is a concept to build intelligent application to meet the societal needs. It uses health care data from different sources and predictive human behavior is used for dispensing information that improves the quality of health. Health is said to be the biggest wealth one can have. In the view of biomedical perspective health can be defined as the ability of the human body to function normally [1]. Health can be disrupted by diseases from time to time. The functioning of the human body would become abnormal for various reasons such life style, food habits, poor immunity, smoking, narcotics, alcoholic habits and sometimes by hereditary. Health may also set off due to side effects of long term medication and one disorder may lead many other disorders if not taken care.

An abnormal condition that affects the normal functioning of the human body is called a disease and it should not be due to an external injury. Diseases are learnt by their symptoms, which may be caused due to dysfunction. In humans diseases refer to conditions such as pain, dysfunction, distress, social problems or any similar problems. Diseases may also sometimes include syndromes, deviant behaviors, injuries, disorders, disabilities, infections and variations of function and structure. Diseases not only affect people physically, but also emotionally, as living with a disease can affect the person's perspective on life [2][3][4].

The disease can be categorized into acquired disease, acute disease, chronic disease, congenital disease, genetic disease, hereditary or inherited disease. The majority of the population suffers from chronic disease which is a long term diseases which leads multiple ailments if not taken care. Chronic diseases [4] are diseases which cannot be cured but controlled by vaccines and medication. Diseases usually develop slowly because of behaviors that spoil health. Chronic diseases [4] can be characterized based on factors such as stress, vague beginnings, prolonged affliction and long latency [4] later which leads to impairments. Diseases include both physical and mental disorders which are most common and costly health problems faced by the world. Many chronic diseases are not curable but controllable [3][4].

Chronic diseases are rapidly increasing across the world accounting 45.9% of global burden of disease and are challenging the effectiveness and efficiency of health-care systems without engendering and sparing serious economic and social consequences [4][5]. Chronic diseases are preventable. There is a need to develop intelligent health care system to target health and wellness. This process should includes framing of health policies whenever appropriate and that make healthy choices the easy choice [4][6].

Efforts are needed to build sustainable Preventive Care of Diseases systems (PCD) to solve the problems created by the growing prevalence of health conditions. Decision support systems need knowledge about the type of health problems [7][8][9], the way the services are allocated to resolve these challenges. An interface is needed to access the electronic medical information systems to access the reliable medical information while improving communication with patients by follow-up reminders. The member patients should be provided with access to information about their health to help themselves to take care [4][10].

The technological advancements in electronic medical records have made easy to get information much better than the old paper charts. The advent of technology has improved the power of analytics and machine learning [11]. Machine learning methods develop a model knowing the input and output of the domain and it uses the large amount of computing power which a human brain cannot do. More the domain data fed to

Revised Manuscript Received on April 25, 2019

Mohan Kumar K N, Dept. of CSE., SJB Institute of Technology, Bangalore, India.

S.Sampath, Dept. of ISE., Adichunchangiri Institute of Technology, City Chikkamagaluru, India.

Mohammed Imran, Research and Development, Ejyle Technology, Bangalore, India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

machine learning model more would be the accuracy of the predictions [12]. We have to identify the scenarios in which power of machine learning helps us to prevent diseases [11].

Hence there is a need to analyze and design an efficient system which can predict, classify diseases and detect anomalies from electronic health records. A systematic approach is required to derive knowledge out of electronic health records which can help to prevent diseases. Exploration of existing techniques through literature survey would give better visibility into the context discussed.

This paper work is organised as, Part-2 detailing literature survey, Part-3 objective of survey, Part 4-6 details the agenda of the survey, Part-7 discusses the scope for future work and ultimately Part -8 concludes the paper by summarising the outcomes.

II. LITERATURE SURVEY

Several models are required to predict the condition of health, which is hard to do and rapidly escalate to next level of analysis. By using the data such as history of patient records, genomic and information collected using questionnaire has been helped the machine learning models in healthcare community to predict and classify diseases [13][14][15]. Notably [13]'s work uses Ontology and support vector machine (SVM) for the prediction of Chronic Kidney Disease. The features such as age, blood pressure, specific gravity, albumin, glucose and red blood cells count were considered for the study. On the other hand, [14] has explored the methods for prediction of diseases associated with genes using SVM and has suggested for Multi Class disease gene classification. In [15], the concept of Diagnosis for Diabetes Mellitus using SVM is developed; for this study they have used patient's historical health records.

A study has also been conducted on medi-claims data and clinical data [16],[17],[18]. In [16] they have applied linear discriminant analysis (LDA) to classify the diseases based on disease codes (ICD) mentioned in large medi-claim data, Text Processing and Natural Language Processing (NLP) is used for health care claim data processing [17] and in[18] LDA along with SVM is used for diagnosis of diabetes. The features such as number of times pregnant, plasma glucose concentration, oral glucose tolerance test and diastolic blood pressure have been considered for the study. Interestingly [16] and [18] have suggested including of drug codes and other features related to diabetics in future study.

In order to achieve insight into diseases classification for patients suffering from Cystic Fibrosis, an algorithm which derives weighted sum scores is introduced in [19]. The study has been conducted on Cystic Fibrosis dataset and features such as antibiotic use, nutritional supplements, and gender and body mass index. The study suggests redesigning the algorithm for the streaming data.

There are lot of studies conducted on chronic diseases using hospital data for decades [20][21][22][23][24]. The records used for the study is either member patient data from hospitals or records from public repositories like Framingham Heart Study dataset [20], CMS [21] etc. UCI repository [22] provides clinical records for every chronic disease individually and CMS repository provides the records for all the chronic diseases in a consolidated CSV file. The ML algorithms such as Association Rule Mining (ARM) [4], Collaborative filtering [21], Clustering [22][24] are used for the prediction of diseases. Data engineering is the first step which is required to be done to standardize the health records before applying the classical ML methods i.e. Data Modeling of the health records makes the data access and usage by avoiding latency. Studies have been done on the several chronicle diseases' dataset available in public repository like CMS and UCI [24].

The commonly used classification algorithms are Logistic Regression, Artificial Neural Network, Decision Tree, C4.5, CART, Clustering [23][24][25][26][27][28][29][30][31][32][33]. The computationally intelligent methods like multilayer perceptron [24][32] are used for heart disease prediction, along with the prediction of diseases. There were studies conducted to reduce the number of features for specific disease, for example heart disease records of UCI repository [24][25] was studied. Decision Tree is used for learning Medical Claim Process to classify disease based on disease code [23][24]. Soft computing techniques are used to predict diabetes [27]. A major challenge of handling of unlabeled data is solved using graph based approach [30].

Certain studies have used AdaBoost, Naïve Bayes, Markov model [34],[35] to study Disease Progression [34] and to evaluate the techniques used for Prediction of Hospital Length of Stay using Medicare Claims data [24][35]. The computational intelligent techniques such as Genetic Algorithm, Fuzzy Rules, Artificial Neural Network, ensemble prediction and Random Forest [36][37][38][39][40][41] are used for the study Comorbid Rare Conditions using medical claims data[42], disease pattern classification[36], heart disease prediction [37], to classify multi chronic condition [38]. The impact was studied by the wellness of pre diabetic patients on self-reported changes in physical activity level and food choices[39], to predict the patients' length of stay, total incurred charges, and mortality rates[24][40] and to predict disease over big data from healthcare communities [41].

The interesting part noticed in our survey is risk prediction in patients suffering from chronic kidney disease patients using Framingham equation [43]. In [43], they have evaluated the ability of prediction by experiencing a cardiovascular event (CVE) through Framingham Risk Score (FRS-CVD). It is also learnt from our survey that Statistical packages such SPSS is used to calculate the regularity of occurrence of chronic diseases and multimorbidity among inpatients [44][45].

In the above survey we could see only one variant analysis but [46][47][48] have conducted a study on multi criterion and multivariate regression analysis to predict and classify the chronic diseases. This approach had led to near optimization of prediction and classification. In [49], a survey has been conducted on advanced technologies for Diabetes Mellitus management; here they have focused on performance of classifiers and risk prediction models.

After exploring the existing work [24][50][51], it is evident that majority of the studies is done for the prediction of cost for the insurance premium using claim data. The methods of clustering and classification are applied on unstructured and structured data.

Not many chronic diseases have been dealt in detail considering large volume of data and features. There is a need for a competent model to extract knowledge from health care records including medi-claim data. The relation between a chronicle disease and Hierarchical Chronicle disease Code (HCC) is yet to be explored to find ways to prevent health deterioration.

III. SYSTEMATIC REVIEW PROCESS

An overview of prediction of diseases for Preventive Care of Health Deterioration (PCHD) using ML approaches. Fundamentally, the following objectives are defined and then certain criteria's have been identified to address the issues on PCHD.

- Machine Learning methods (techniques), which are used for prediction and classification of diseases.
- The features that are commonly used for disease prediction and classification.
- The feature selection techniques used.
- The data sets on which the studies are conducted
- Evaluation of performance metrics that are adopted for Disease prediction.

IV. DISEASE PREDICTION TECHNIQUES

The techniques used in disease prediction till now can be categorized into parametric and nonparametric. The parametric techniques are Logistic Regression [23],[24],[30],[36],[46],[48],[34],[22] Linear Discriminant Analysis [18],[37], they have proved better performance and accuracy in prediction. In Table-1, the algorithms review results are clearly consolidated from all the existing works to find out that are majorly adopted. Logistical Regression techniques for disease prediction dominated, with 27.5%, this is because of the nature of health data.

Methods like clustering [22] and artificial neural network [24] could perform well with non-linear health records. The accuracy of the prediction of diseases is better compared to linear methods hence these methods work out better in real world situation. The plot in Figure-1 describes the application of the algorithms in prediction of diseases. It seems like logistical regression (27.5%) along with other methods is used in majority of the studies [32][33]. Notably Support Vector Machine [32][34] dominates after logistical regression and is extensively (25%) used for the prediction and classification. Other techniques such as fuzzy [45] have proved better contextually. Decision Tree and Random Forest are of choice too at 22.5% and 5%. Non parametric models are expensive computationally. This makes it a bad choice when the size of data is big and optimization of these techniques is also a far away shot. This should be one reason for a combination of Machine Learning with big data technique and statistical methods [43]. Ensemble [45] machine learning model provide better accuracy on prediction of diseases with the benefits of combining multiple methods, which gives the average value of the prediction of multiple methods. Neural Network (NN) [45] and its variants have become increasingly popular with other techniques applied contextually and their usage is detailed in Table 1. Big data and Internet of Things (IOT) techniques have pitched in advancements in handling health care management. These technologies will play a crucial role in the near future.

Table I. List of Techniques Used for Predicting Disease

Sl. No.	Techniques	Percentage of Studies
1.	Logistic Regression	27.5
2.	SVM	25
3.	Decision Tree	22.5
4.	Clustering	20
5.	Artificial neural network	17.5
6.	C4.5, CART	12.5
7.	LDA	7.5
8.	Linear Regression	7.5
9.	Statistical Methods(SPSS)	5
10.	Multilevel Regression Analysis	5
11.	Genetic Algorithm	5
12.	Random Forest	5
13.	Convolutional Neural Network (CNN)	2.5
14.	Greedy Optimization Methodology	2.5
15.	Framingham Equation	2.5
16.	Fuzzy Rules	2.5
17.	Ada Boost	2.5
18.	Markov model	2.5
19.	NLP/Text Mining	2.5
20.	Naïve Bayes	2.5
21.	Ensemble Prediction System	2.5
22.	Collaborative filtering	2.5
23.	Automated Weighted Sum	2.5

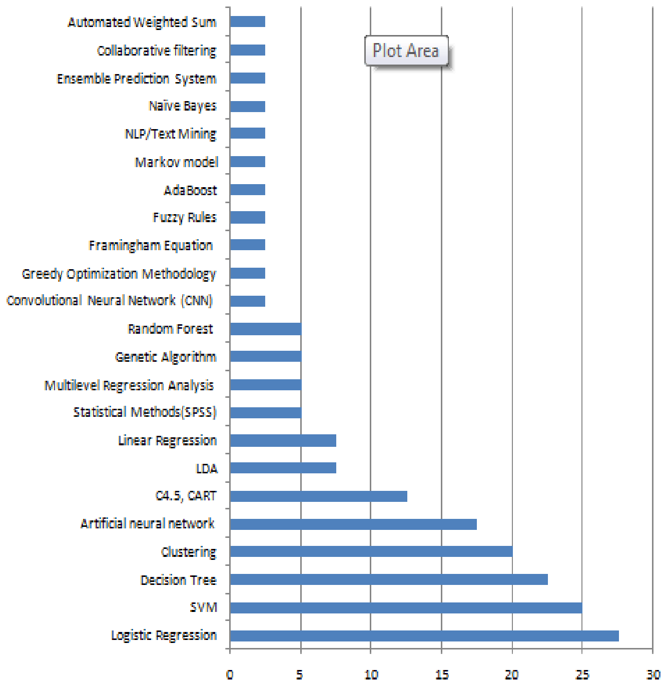

Fig 1. Plot of Prediction Techniques

V. DATA SETS USED FOR DISEASES PREDICTION

Table 2 is a list of data sets which are studied for disease prediction. Many of studies have been conducted on UCI Machine Learning Repository [14][27] and Medi-claim Coverage Database [15]. Other major source of data is Hospital records [39][40][49] which is extensively used (32.5%); authors have gathered this data from different hospital and clinics [40][41]. Studies are also conducted on Electronic Medical Record [24], Medi-claim record data [15][24][45] and diabetes data set (10% - 15%) [26][28][30][36]. To achieve optimization studies have been conducted on the above mentioned data sets as well as on others data sets such as heart disease data [16][26], behavioral and social data [25], local health station data [28], call center logs [20] and National Health Nutrition data (2.5 %- 5%) [33]. It is evident from the studies that certain data set has been used to study the specific diseases. Interesting part is that Medi-claim data [15] is used for the prediction and classification of diseases. Figure 2 depicts the count of different data set usage in disease prediction.

TABLE II. List of Data Sets Used for Predicting Disease

Sl.No	Data Sets	Percentage of Usage
1	Hospital Records	32.5
2	Diabetic	15
3	Electronic Medical Record (EMR)	12.5
4	Medi-Claim Data	10
5	Manual medical data	7.5
6	Behaviour and Social	5
7	Heart	5
8	Genomic and Proteomic	5
9	Local health station	5
10	Hospital Service Appointments	5
11	Cystic Fibrosis	2.5
12	Call centre Logs	2.5
13	Cancer	2.5
14	National Health Nutrition	2.5

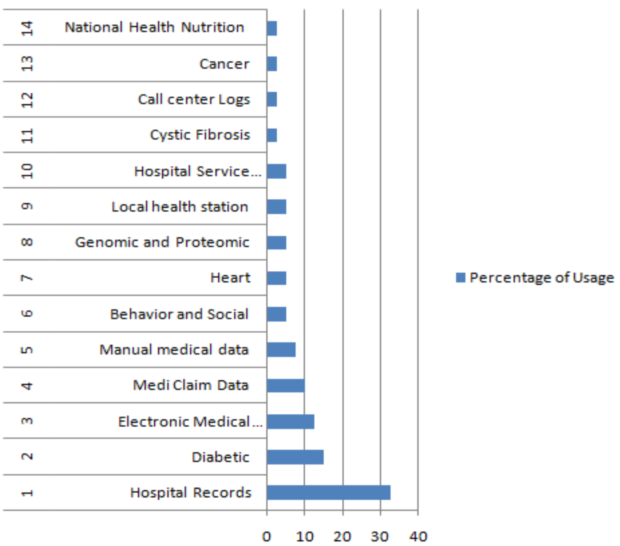

Fig 2. Plot of Data Sets used for Disease Prediction

VI. FEATURES USED FOR DISEASE PREDICTION

In Table-3, it is clear that age, gender and race are the most used feature in disease prediction (35% - 45%). Age, gender, race [13][14][16][17][19] are commonly associated features along with other features. Age and gender provides a description of the health state at different phases of human life. From this study it is deciding factor in disease prediction (17.5%). To achieve precision, features such as habits [23], education [38] and total cholesterol [23] are also studied (10% to 12.5%). Interestingly some of the studies have proved that features such as marital status [19], number of times pregnant [32], current employment [38] and also helps in disease prediction (5% - 7.5%). The latest trend is that disease codes (ICD) [15][24] described in medi-claim data have been used for prediction analysis. The other commonly studied features that are used in process of disease prediction and prevention are Length of Stay in hospital [22], RBC [31], Antibiotic use [15], Clinical features [21], Bacterial Pneumonia [25], Chest pain [27], Family history [31] and albumin [13] for specific disease study. There are certain studies which incorporated non conventional study are house-hold income [51], length of Professional Employment [51], has a Family Doctor [51], family history, previous occupation [40] and Hospital admission type and season (2.5% - 5%) [22]. In [31] it is clear that features such as Body mass index (BMI), blood pressure, sugar level are mainly considered for prediction of dieses. Figure-3, summarizes the features used for prediction of diseases.

TABLE III. List of Features Used for Predicting Disease

Sl.No	Features	Percentage of Usage
1	Age	45
2	Race	40
3	Sex	35
4	BMI	17.5
5	Blood pressure	17.5
6	Sugar level	17.5
7	Habits	12.5
8	Education	10
9	Total Cholesterol	10
10	Marital Status	7.5
11	Employment	7.5
12	ICD	7.5
13	Demographic data	7.5
14	Length of Stay in hospital	5
15	Number of times pregnant	5
16	RBC	5
17	Antibiotic use	5
18	Clinical features	5
19	Bacterial Pneumonia	5
20	Chest pain	2.5
21	Length of Professional Employment	2.5
22	House hold Income	2.5
23	Has a Family Doctor	2.5
24	Previous Occupation	2.5
25	Family history	2.5
26	Albumin	2.5
27	Hospital admission type and season	2.5

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

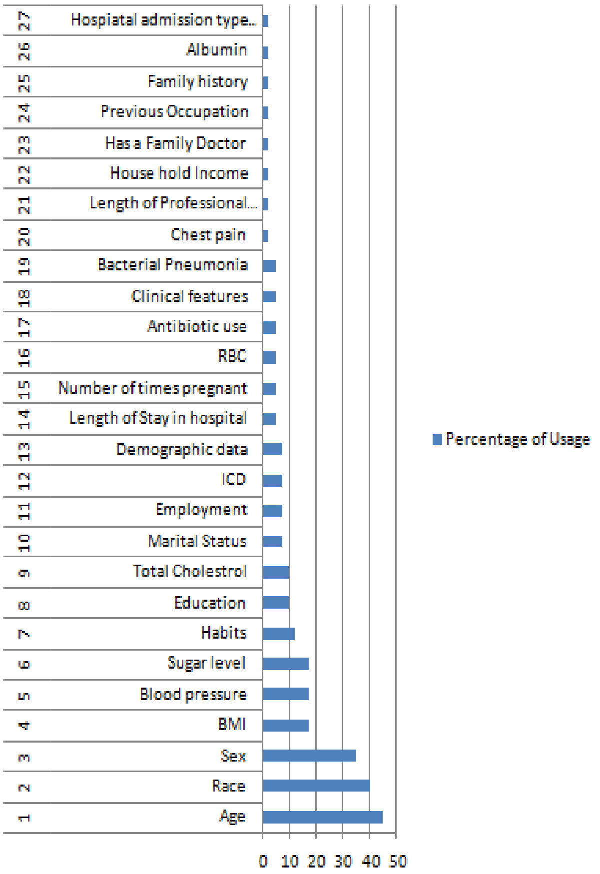

Fig-3. Plot of Features used for Disease Prediction

VII. FEATURE SELECTION TECHNIQUES.

The techniques that have been used for feature selection are detailed based on context of the data and requirement of the applications. The linear kernel is used to train SVM classifier to find the maximum-margin hyper-plane from the context properties data [13]. Construction of several sparse matrices from a temporal aggregate of claims data and the matrices thus created can be viewed as a document-term matrix which can be used for training machine learning models [14]. The commonly used feature selection techniques are Re-sampling and Attribute Selection Techniques on manually collected data [24]. Rule Learners are used to extract rules that describe a relation between input features and output class labels directly from diabetic data set [30]. AI-based models are used to develop term document matrix for diagnosis data [44]. Lastly some statistically significant predictor factors are used commonly for feature extraction [50].

VIII. PERFORMANCE EVALUATION MEASURES

The most widely used evaluation metric for classification is confusion matrix and accuracy [24][25][27][33]. Notably quantitative performance evaluation is used for contextual properties [13]. The other performance evaluation techniques are baseline ranking [17], Gibbs sampling [21] and Ten Fold Cross Validation on the training set and measure of accuracy [18]. The evaluation metric for prediction methods are root mean squared error and Area under Region Of Convergence (ROC) [33].

IX. ESTIMATION OF WEIGHTS OF ALGORITHMS, DATA AND FEATURES USING ENTROPY

Entropy methods estimate the weights of various criteria for the three variables Algorithms, Data and Features are calculated. Degree of diversification of the information D1, D2 and D3 for each criteria is calculated using entropy value Different weights of ratings for each criteria is calculated using Di, is given below. The order of the priorities of criteria is found using Ci.

Component	Entropy Ei	D = 1-Ei	Weights of Criteria Wc = (Di / ∑Di)
Algorithms	0.8625	0.1375	0.4877
Data	0.8706	0.1294	0.459
Features	0.985	0.015	0.015

The order of the priorities of criteria is **Algorithms** > **Data** > **Features**

X. FUTURE WORKS

The following are the research gaps found in the thorough review on prediction of diseases for preventive care of health deterioration.

- More number of studies has been conducted using SVM and Logistical Regression considering only certain features. We could hardly see models using majority features in their study. There is a scope to include more features contextually to achieve higher accuracy and optimization. There is a need for more comparative study of Machine Learning algorithms in health care domain to decide an optimistic approach [13][20][26]
- There are certain studies which have used medi-claim data for prediction of diseases using ICD but ICD itself does not provide the clarity on hierarchy of chronicle diseases. Therefore there is a scope to include HCC code for the prediction of chronicle diseases [14][20][24].
- The role of IoT and Big data in health care has become increasingly important and to meet the needs of growing infrastructure of health care data collection.
- The existing models consider local hospital records of small population and features like age, gender and habits are studied. They do not consider the history of patients with respect to hierarchy of diseases [17].
- The existing models do not predict multiple chronicle diseases correlation and their hierarchy of occurrences [24].
- The existing models consider ICD9 for prediction but considering ICD10 and HCC would provide better results [14][20][24]. The existing models fail on big data records. Hence there is a need for adoption of big data technology to study medical records of large population [36].

XI. CONCLUSION

This paper records and consolidates an overview of existing concepts on disease prediction for preventive care of health deterioration.. Firstly the needs of the overall survey were defined. Secondly, the objectives were framed with the perspective of the problem we are dealing with, so that these objectives should pave way for inferences of our investigation. Several methods and models (techniques), were explored for preventive care by predicting diseases. Browsed the Website to explore more on the title and succeeded to get information of interest and consolidated the findings to present a review on the research area from the year 2004 to 2018. The finding of the study has uncovered some issues which are yet to be dealt; this opens up the opportunity for the researchers for further study. The information popped out from this work would encourage us to carry out further studies in the field of preventive care of diseases for health deterioration.

REFERENCES

1. World Health Organization. (2006). Constitution of the World Health Organization – Basic Documents, Forty-fifth edition, Supplement, October 2006.
2. White, Tim (19 December 2014). "What is the Difference Between an 'Injury' and 'Disease' for Commonwealth Injury Claims?". Tindall Gask Bentley. Archived from the original on 27 October 2017. Retrieved 6 November 2017.
3. http://www.phac-aspc.gc.ca/ccdpc-cpcmc/topics/chronic-disease_e.htm
4. https://www.gnb.ca/0051/pub/pdf/2010/6960e-final.pdf
5. WHO (2008) Global Strategy on Diet, Physical Activity and Health http://www.who.int/dietphysicalactivity/publications/facts/riskfactors/en/index.html.
6. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=Abstract&list_uids=16046654&query_hl=4&itool=pubmed_DocSum Pham H.H., Schrag D., Hargraves J.L., Bach P.B. (2005) Delivery of preventive services to older adults by primary care physicians. JAMA. July 27; 294(4):473-81.
7. Institute of Medicine. (2001) Crossing the Quality Chasm: A new health system for the 21st century. Washington, DC: National Academy Press.
8. Schoen C., Osborn R., Huynh P.T., et al. taking the pulse of health care
9. Systems: Experiences of patients with health problems in six countries.
10. Health Affairs. Nov 2005;
11. Leatherman S., Berwick D., Iles D., et al. The business case for quality:
12. Case studies And ananalysis.HealthAffirs.2003;22(2):17-30
13. http://www.natpact.nhs.uk/eventmanager/uploads/dm eventrafaelbengoa1 .ppt
14. https://www.healthcatalyst.com/clinical-applications-of-machine-learning-in-Healthcare.
15. http://web.orionhealth.com/rs/981-HEV-035/images/Introduction_to_ Machine_ Learning.pdf
16. Hatim Guermah, Tarik Fissaa, Bassma Guermah, "Using Context Ontology and Linear SVM for Chronic Kidney Disease Prediction" , ACM LOPAL'18, May 2018.
17. Nahla H. Barakat, Andrew P. Bradley, Mohamed Nabil H. Barakat,
18. "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus",
19. IEEE Transactions on Information Technology in Biomedicine, vol. 14,
20. no. 4, July 2010.
21. Duc-Hau Le, Manh-Hien Nguyen, "Towards more realistic machine learning techniques for prediction of disease-associated genes", ACM 9th International Symposium on Information and Communication Technology, Dec 15.
22. Varun Chandola, Sreenivas R. Sukumar, Jack Schryver, "Knowledge Discovery from Massive Healthcare Claims Data", 19thACM conference on knowledge discovery and data mining, 2013.
23. Fred popowich, "Using Text Processing and Natural Language Processing for health care claim processing", ACM transactions, SIGKDD Exlorations, Volume 7, pp-59.
24. Duygu Çalis , Esin Dog antekin , "An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier", *ELSEVIER Expert Systems with Applications 38 (2011) 8311–8315.*
25. Anthony Quinn, Andrew Stranieri, John L. Yearwood, Gaudenz Hafen, *"*A Classification Algorithm that Derives Weighted Sum Scores for Insight into Disease*", ACM HIKM 2009.*
26. Shen Song, Jim Warren, Patricia Riddle, "Developing High Risk Clusters for Chronic Disease Events with Classification Association Rule Mining", *in the Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management,* vol. 153, 2014.
27. Darcy A. Davis, Nitesh V. Chawla, Nicholas Blumm, "Predicting Individual Disease Risk Based on Medical History", in the *Proceedings of the 17th ACM conference on Information and knowledge management*, 2008.
28. Darcy A. Davis, Nitesh V. Chawla, Nicholas Blumm, "Predicting Individual Disease Risk Based on Medical History", in the *Proceedings of the 17th ACM conference on Information and knowledge management,* 2008.
29. F. H. D. Araújo, A. M. Santana and P. Santos Neto, "Evaluation of Classifiers Based on Decision Tree for Learning Medical Claim Process", *IEEE Latin America Transactions*, vol. 13, no. 1, Jan. 2015.
30. Deepa Gupta, Sangita Khare , Ashish Aggarwal, "A Method to Predict Diagnostic Codes for Chronic Diseases using Machine Learning Techniques", *In the Proceedings of IEEE International Conference on Computing, Communication and Automation* ,2016.
31. Huzaifa Shabbir Dhorajiwala, Er. Asadullah Shaikh, "Survey on Decision Support System for Medical Diagnosis Using Data Mining",*International Journal of Engineering Research and General Science*, vol. 3, issue 6, Dec 2015.
32. Jing Xie, Dongsheng Hu, Dahai Yu, Chung-Shiuan Chen, Jiang He, Dongfeng G, "A quick self-assessment tool to identify individuals at high risk of type 2 diabetes in the Chinese general population", *in the proceedings of J Epidemiol Community Health*, 2010
33. M. Durairaj, G. Kalaiselvi, "Prediction Of Diabetes Using Soft Computing Techniques- A Survey", I*nternational journal of scientific & technology research*,vol. 4, issue 3, March 2015.
34. D. Ashok Kumar and R. Govindasamy, "Performance and Evaluation of Classification Data Mining Techniques in Diabetes", International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015.
35. Costas Sideris, Nabil Alshurafa, Mohammad Pourhomayoun, Farhad Shahmohammadi, Lauren Samy, Majid Sarrafzadeh, "A Data-Driven Feature Extraction Framework for Predicting the Severity of Condition of Congestive Heart Failure Patients", *IEEE Transaction*, 2015.
36. Asha Gowda Karegowda, Punya V, M.A.Jayaram, A.S .Manjunath,"Rule based Classification for Diabetic Patients using Cascaded K-Means and Decision Tree C4.5", *International Journal of Computer Applications* , vol. 45, no.12, May 2012.
37. Chih-Hung Jen , Chien-Chih Wangb, Bernard C. Jiang , Yan-Hua Chu , Ming-Shu Chen,"Application of classification techniques on development an early-warning system for chronic illnesses", Transaction of Expert Systems with Applications, ELSEVIER, 2012.
38. Djordje Gligorijevic , Jelena Stojanovic , Zoran Obradovic, "Disease types discovery from a large database of inpatient records: A sepsis study", *ELSEVIER Expert Systems with Applications*, vol 111, pg. 45-55, Dec 2016.
39. Xiang Wang, David Sontag, Fei Wang, "Unsupervised Learning of Disease Progression Models", *20th ACM conference on knowledge* discovery and data mining, vol. 1, Aug 2014.
40. Dimitrios Zikos, Konstantinos Tsiakas, Fadiah Qudah, "Evaluation of Classification Methods for the Prediction of Hospital Length of Stay using Medicare Claims data", ACM PETRA MAY-2014.
41. Chirag Nagpal, Kyle Miller, Tiffany Pellathy, Marilyn Hravnak, Gilles Clermont, Michael Pinsky, Artur Dubrawski, "Semi-Supervised Prediction of Comorbid Rare Conditions Using Medical Claims Data", in the Proceedings of IEEE International Conference on Data Mining Workshops, 2017.
42. S. M. Kamruzzaman, Ahmed Ryadh Hasan, Abu Bakar Siddiquee and Md. Ehsanul Hoque, Mazumder, "Medical Diagnosis using Neural Network", in the proceedings of 3 rd International Conference on Electrical & Computer Engineering-2004.
43. Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg, "Data Mining in Clinical Decision Support Systems for Diagnosis, Prediction and Treatment of Heart Disease", *International Journal of Advanced Research in Computer* Engineering & Technology (IJARCET), vol. 2, issue 1, Jan 2013.
44. Ravi S. Behra, Pranitha Pulumati, Ankur Agarwal, Ritesh Jain, Vinaya Rao,"Predictive Modeling for Wellness and Chronic Conditions", In the Proc. 14th International Conference on Bioinformatics and Bioengineering, 2014.
45. Ramona S. DeJesusa, Matthew M. Clark, Lila J. Finney Ruttena, Robert M. Jacobsona, Ivana T. Croghanb, Patrick M. Wilsong, Debra J. Jacobsong, Sara M. Linkh, Chun Fang, Jennifer L. St. Sauver, "Impact of a 12-week wellness coaching on self-care behaviors among primary care adult patients with prediabetes", ELSEVIER, Preventive Medicine Reports, pp. 100–105, 2018.

46. Jelena Stojanovic, Djordje Gligorijevic, Vladan Radosavljevic, Nemanja Djuric, Mihajlo Grbovic, and Zoran Obradovic, "Modeling Healthcare Quality via Compact Representations of Electronic Health Records", ACM Transactions on computational biology and bioinformatics, vol. 14, no. 3, 2017.
47. Min Chen, Yixue Hao, Kai hwang, Lu wang, Lin wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities" , IEEE. Translations and content mining, vol. 5, 2017.
48. Santiago Cede, Marian Goicoechea, Esther Torres, Úrsula Verdalles, Ana Pérez de José, Eduardo Verde, Soledad García de Vinuesa, José Lu, "Cardiovascular risk prediction in chronic kidney disease patients", nefrolgia, pp. 7293–300. 2017
49. Paul Osemeke Nwani, Ambrose Ohumagho Isah, "Chronic diseases and multimorbidity among elderly patients admitted in the medical wards of a Nigerian tertiary hospital", ScienceDirect, Journal of Clinical Gerontology & Geriatrics, pp. 83-86, 2016.
50. Kumari Deepika, S. Seema, "Predictive Analytics to Prevent and Control Chronic Diseases", IEEE Transaction, 2016.
51. Divya Jain , Vijendra Singh, "Feature selection and classification systems for chronic disease prediction: A review", ELSEVIER, Egyptian Informatics Journal 2018.
52. Danika Tremain, LukeWolfenden, Kate Bartlem, Emma Doherty, "Provision of Chronic Disease Preventive Care in Community Substance Use Services: Client and Clinician Report", ELSEVIER, Journal of Substance Abuse Treatment 68 (2016) 24–30.
53. Jane Murray Cramm, Anna Petra Nieboer, "Short and long term improvements in quality of chronic care delivery predict program sustainability", ELSEVIER, Social Science & Medicine, 2014.
54. Konstantia Zarkogianni, Chih-Wen Cheng, Konstantina S. Nikita, Eleni Litsa, Konstantinos Mitsis, Po-Yen Wu, Chanchala D. Kaddi, "A Review of Emerging Technologies for the Management of Diabetes Mellitus", IEEE transactions on Biomedical Engineering, vol. 62, no. 12, Dec 2015.
55. Gustavo Glusman, "A Data-Rich Longitudinal Wellness Study for the Digital Age", IEEE pulse , 2017.
56. Kate Bartlem, Jenny Bowman, Kate Ross, Megan Freund, Paula Wye, Kathleen McElwaine, Karen Gillham, Emma Doherty, Luke Wolfenden, and John Wiggers, "Mental health clinician attitudes to the provision of preventive care for chronic disease risk behaviours and association with care provision", BMC Psychiatry ,2016.
57. Denise Y. S. Tam , Yvonne Y. C. Lo and Wendy Tsui, "Knowledge, practices and expectations of preventive care: a qualitative study of patients attending government general outpatient clinics in Hong Kong.", BMC Family Practice, 2018.

AUTHORS PROFILE

Mohan Kumar K N, research scholar in HOD in Dept. of Information Science & Engineering at Adichunchanagiri Institute of Technology, Chikkamagaluru, Karnataka. Presently working as Assistant Professor at SJB Institute of Technology. His area of research includes Machine learning, Deep learning, Pattern Recognition, Game Theory, and Predictive analysis. Authored few International publications which include Journals and Peer-reviewed Conferences.

Dr. S.Sampath presently working as Professor & HOD in Dept. Of Information Science & Engineering at Adichunchanagiri Institute of Technology, Chikkamagaluru, Karnataka. His fields of interest are High Performance computing, Data Mining & Machine Learning. He has been recognized as reviewer for many reputed International Journals in the field of Computer Science including Journal of Computational Science from Elsevier. He has published many research papers in good international journals and conferences.

.

Dr. Mohammad Imran received Ph.D. in computer Science in year 2013 from university of msyore. Presently working as Senior Data Scientist, His area of research includes Machine learning, Deep learning, Pattern Recognition, Computer Vision, Biometrics, Image Processing, Predictive analysis. Authored 35 International publications which include Journals and Peer-reviewed Conferences.