

Proposed Methodology for Personalized Medicine in Healthcare System using Data Mining Techniques with an Emphasis on Mental Illness Disease

Dr. Bipin Bihari Jayasingh

Professor, CVR College of Engineering/ IT Department., Hyderabad, India.

Email: bbjayasingh9@rediffmail.com

Abstract-Today, the Medicare system has been improved their models to achieve the best care for the patients and customized the delivery of healthcare treatment to the individuals by helping the doctors in the diagnosis of a disease. The systems are useful due to implementation of personalized medicine that is an effective way of doing treatment to a patient by considering omic data and electronic health record data while classifying the patient into subgroups based on common biological basis of diseases. This paper discusses the mental illness disease and its categorization to predict a disease based on its symptoms along with various challenges while processing data sets. There is data mining and machine learning algorithms used for feature selection and classification in order to improve the prediction performance for the specific mental illness disease are discussed to develop a new model. The proposed methodology is to develop a personalized Medicare system considering omic or Electronic Health Record (EHR) data using the best feature selection and classification mining algorithms to help the doctors to diagnose the disease at the early stage and identify the root cause for better treatment.

Index Terms: Personalized Medicine, Mental Illness, Data Mining, Classification, omic and EHR data.

I. INTRODUCTION

Personalized Medicine is an emerging approach for disease treatment and prevention that considers individual variability in genes, environment and lifestyle for each person [7]. It takes personal care about the patient and has a very strong potential to consolidate modern e-health systems called precision medicine. The author [7] has studied about psychiatric disease that has rich information exists in serial measurement of mental health symptom scores.

They basically try to plot a trajectory [1, 2] using multiple symptoms to make personalized predictions about future symptoms and some psychiatric events. This approach deals with an estimation of a population that is average trajectory for symptom scores of other people and estimates the individual deviations from average trajectory. Then, the author tries to estimate the risk of experiencing an event using the symptom trajectory. This paper [7] demonstrates a study of anti psychotic therapy for schizophrenia.

Personalized medicine for mental illness disease takes care of the patient based on personalized measures. Personalized medicine is also treating the patient health appropriately. Clinician can be benefited by using this new approach of precision medicine that is based on individual tailored treatment where patient individual profiles are taken into consideration. Individual patient details are required for tailor made treatment and the details of each step are described. Taking this into account, the author [9] discusses the technology which allows collection of different real time data like electronic medical records and wearable smart device to achieve predictions. This kind of tailor made treatment can also be implemented in psychiatric field [14].

Personalized medicine does not mean that caring to only VIPs' and giving them personalized care instead the drugs which are widely used by many healthcare and media are for more common people also. The drugs that are used by most of the patients and which are 100% effective and safe because of human genome project. Sometimes Personalized medicine can also be defined based on genetic and genomic tests where it treats the disease more effectively than traditional treatments [4]. Personalized medicine can be viewed as short term treatment due to faster recovery of the patient at whatever level the disease resides. Personalized Medicine is very powerful in modern e-health care system that consolidates the personalized drugs to suite to a person. It also takes care of side effects which are very much limited because it applies specific drugs to specific individuals.

II. LITERATURE REVIEW

There is a study about omic data set and Electronic Health Records (EHR) are used as basis for the precision medicine [3] to predict effectively the symptoms of a disease. Omics refers to a field of study in biology ending in omics like genomics, proteomics and metabolomics that are very important features towards personalized health care systems. Genomics is the study of Structure, function, evolution and mapping at editing of genomes. Progeomics is the large-scale study of proteins and Metabolomics is the study of chemical processes that involves metabolites. Omics always differs from person to person that depends upon food habit of a person. Omics data plays a huge role, for example a person who resides near river

banks has more chance of being affected than a person who stays in a cosmopolitan city though both the genomes of the persons are altogether different. When the omic data of a person is collected in form of a catalog, it is used as basis for the precision medicine. Some of the omic structures are responsible for uncovering communication among different omic structures, which form the basis for the medical attention.

Some research has been carried out for genomic interpretation and genetic prediction. To conduct the experiments on psychiatric diseases, get information online at patient health information. The way a person's symptoms change over time and with respect to other symptoms is informative about the course of disease. The author has developed a model for prediction that takes repeated symptom scores as input and makes predictions about future symptoms and a binary event that is an event that are not directly a function of symptoms like suicidal behavior, hospitalization [7, 14]. Due to wide spread use of Electronic Health Records (EHR) in many health care facilities, healthcare data are available for analysis in order to improve the quality of patient care more efficiently [5]. Also, its exploration is challenging due to its heterogeneity incompleteness and unbalanced structure. This paper [11] addresses the issue of imbalanced health care or medical data through a case study using real brain tumor diagnosis problem. Though the Morph metric analysis of histopathological images is rapidly emerging as a valuable tool for a variety of diseases, the author [11] has studied a disease called Oligodendroglioma which is a subset of brain tumors that has high rate of responsiveness to chemotherapy.

The three main goals of personalized medicine is to prevent, diagnosis and treat the patients. It is essential that early detection of problem for mental illness patient is to improve the quality. As a result, the children may not suffer from complicated problems [8]. It may be one of the biggest challenges globally by 2020 [7] to predict in advance the common mental illness problem of children.

III. MENTAL ILLNESS

Mental illness is a chronic illness with recurring persistent episode of schizophrenia, anxiety disorder and depression that can last from one day to months. Mental disorder is one type of disease where the patient may do any inactive thing which may lead to some other problem. This mental illness causes unusual and dramatic shifts in mood, energy, hallucination, lack of motivation, loss of interest, sleep disorder, hyper sexuality and the ability to think clearly. Analyzing an individual mental illness data such as behavioral information, motivational behavior, poorly sleeping information, anxiety

disorder, eating disorder, poor social adjustment etc. are the points to know how he/she will respond to a therapy.

A. Disease Taxonomy

It can be classified the mental disorder disease [7] into four different types 1. schizophrenia 2. Anxiety disorder 3. Depression 4. Alcoholism disorders. ¹Schizophrenia is classified into two types i.e. positive and negative. Positive Schizophrenia is easy to treat and the symptoms for this kind of disease are like hallucination and delusion. It refers to symptoms which make irrational form of thinking or behavior. These symptoms are also known as psychotic symptoms. Negative Schizophrenia symptoms are blunt affect, emotional withdrawal, poor rapport, apathetic social withdrawal, difficult in abstract thinking, lack of spontaneity and lack of flow of conversation, stereotyped thinking.²Anxiety disorder is classified into general anxiety disorder, panic disorder, phobias, OCD (Obsessive Compulsive Disorder), PTSD (Post-Traumatic (socking) Stress Disorder). The symptoms for general anxiety disorder are panic attack, physical symptoms like pain, nausea and headache, fear of leaving the house. ³Depression is also one type of mental disorder where the patient having symptoms like sadness, unhappiness, sleep disturbance, severe lack of energy, loss of interest in hobbies, fogginess, confusion, slower peaking, feeling worthiness or guilty [7]. ⁴Alcoholism disorders are classified into three types mild, moderate and severe. When the people suffering from this kind of disease, these symptoms are observed like being unable to limit the amount of alcohol they drink, using alcohol in situations where it is not safe like driving and swimming. It leads to giving up or reducing social work activities and hobbies.

B. Challenges

With respect to the collection of omic data there are some challenges [3] such as ¹Diverse data collection, ²Data quality, ³High dimensionality and ⁴Different data types. Diverse data collection refers to as omic data differs with time and tissue types at the specified environment. Data quality refers to as Omic data sample may be contaminated at biological factors may take a toll. High dimensionality is refers to as Omic data often has many dimensions which is greater than 104. Different data types refer to as Integrating multi omic data is a challenge because of different variations in biological techniques.

Electronic Health Records are the systemized collection of patient data and population health information in digital formats that are collected during a patient admission into hospital [3]. The EHR Data can be unstructured, structured or semi structured but written clinical notes are the most efficient at human intuitive way of documentation. The Challenges with the EHR data is like diverse data frequency, Data Quality issues, High dimensionality and Different data type [3]. Diverse data frequency refers to some of the data is collected

at high frequency, where as some are taken over days. Data Quality issues refer to Missing data or entry mistakes of data. High dimensionality refers to the large number of sample size available in data set. Different data type refers to the data that is inherently heterogeneous in the EHR.

IV. DATA MINING TECHNIQUES

Every day, the data that are collected from clinician and pharmacological becomes so much huge which require some statistical analysis to improve the prediction performance. In the same way, the machine learning methodology are also used in the psychiatry field to improve the prediction performance. There is a model called Gaussian mixture model [6] is an unsupervised clustering algorithm that fit the data as a number of structures which determine the dimension in which the subgroup are found. The aim is identifying subgroups with in patient rather than classify. Gaussian mixture model with the support vector machine used to segregate the patients where it is explained about the relationship between PANSS (Positive and Negative Syndrome Scale). The PANSS used to identify two statistically distinct subgroups of patients to investigate whether the two identified subgroup are related to treatment or not [7].

A. Classification Mining

Classification of medical data and symptoms of the disease which help in determining the mental disorder based on certain condition [10]. Different classification techniques may be used in order to classify healthcare data such as C4.5 (J48), Random forest (RF) and Random Tree (RT) algorithm. The proposed classification approaches C4.5 (J48) and Random forest (RF) are the efficient techniques which give better classification and disease symptom finding. To improve the accuracy and efficiency of the classification process the filter feature ranking score combined with the wrapper is proved to be good [11]. In the first approach, wrapper process is used for searching better feature for better performance of the model. In this work, the method hybridized a MI (Mutual Information) and MR (Maximum Relevance) with artificial Neural Network (ANN). The unique contribution of this approach is that it amalgamates the potential of wrapper approach to get better features [12, 13].

A classification has the ability to extract useful information from large set of data. This paper [7] discusses a review of classification techniques and analysis on the clinical dataset. However, Bipolar disorder is one type of psychiatry disease where aim is to identify three subgroup of patient according to their age at onset and based on different phenotypical features. To understand the disease and the symptoms of the disease on medical data according to its condition, an approach for predicting the risk [14] of disease by using polymorphism interaction analysis to explore SNP interactions and colon cancer risk. In order to obtain the most likely interaction that the best predict risk of disease uses two unique scoring

functions called Gini index and the percentage wrong. They also have used ten-fold cross validation to analyze the data. The author discusses mental health problem which basically comes at an early stage. There are eight different machine learning mechanism that are used for analyzing and diagnosis the medical health data [8]. The comparative result of classifier on various machine learning technique lead to the conclusion that the Multilayer Perceptron, Multiclass Classifier and LAD Tree produced more accurate results in full attribute set [8].

B. Prediction Models

The author has developed two sub models i.e. ¹sub model for symptom scores and ²sub model for binary events [7]. The sub model for symptom scores needs specifying a form of trajectory that symptoms may take over time with reference to baseline. It has been observed that the way an individual's trajectory for one symptom deviates from mean trajectory is selected or similar to how that person's trajectory or symptom scores for other symptoms deviate from the mean [1, 2]. In some cases, it needs to use sub groups each having a different mean trajectory. This is done because one form of trajectory cannot fit the whole population. So, each sub model, populations have its own mean trajectory. ²sub model for binary events try to compare the individual scores with other and estimate the odds of happening of an event.

There are large volumes of data that are taken from more features of a particular disease. Also, there is an importance to select the feature set for the solution of a particular disease. There is a work for novel feature selection called backdrop neural networks that propose a weight analysis based heuristics to choose only effective features of a disease. The weight analysis based method is called ANNIGMA [11, 12] where the weight is associated with feature basically used in training stage to remove the noisy data.

There is a proposed classification mining algorithm for brain tumor disease which provides a simple diagnosis rule. In order to generate the rules, the author has used decision tree along with bagging algorithm. The author has claimed the work as a globally optimized hybrid feature selection algorithm derived from artificial neural network [11]. The GANNIGMA (Globally Optimized Artificial Neural Network Input Gain Measurement Approximation) is a hybrid feature selection in the proposed approach finds the significant features which help to generate a simplified rule. The GANNIGMA Hybrid feature selection [11] consists of three independent approaches called ¹Filter approach, ²Wrapper approach and ³Hybrid approach. ¹filter approach is cost effective but not so accurate. ²Wrapper approach is not cost effective but accurate. Therefore a ³Hybrid approach is involved in this feature selection process by combining filter approach with wrapper approach. The proposed hybrid feature selection with ensemble classification technique in this paper [11] combines a maximum relevance and minimum redundancy filter heuristic with globally optimized wrapper heuristic

GANNIGMA. Hence, the results indicate that GANNIGMA + MRMR + Bagging + Decision Tree feature selection based proposed ensemble approach provides a more simplified decision rule with higher accuracies.

V. PROPOSED METHODOLOGY

There are proposed methods [3] that deal with the idea of classifying people into subgroups who share the same biological disorder to provide precision medicine using algorithms like Bayesian classification, KNN neural networks.

A. Objective

- The main objective of the proposed system is to develop a Personalized Medicare system in such a way that it treats the individual requirement of a patient according to environment and space.
- The system also allows the doctors to diagnose at the early stage and immediately starts the treatment procedure.
- This system considers the nature, genes of an individual and environmental factor for the preparation of a particular drug.
- The system presents the disease by facilitating disease prediction and treatment along with minimizing the risks of development of disease.
- Personalized treatment helps doctor to identify the root cause of the problem for better treatment and soon recovery of the patients.
- As a suggestion, the doctor can tell the patient what the treatments are necessary for the patient instead of trial various types of treatment.
- Thus, major goal of personalized medicine is to predict an individual vulnerability to develop illness, achieve accurate diagnosis and optimized the most effective and efficient treatment.

B. The Process

- The EHR or omic data set are collected from hospital and try to precise the data set to select only those features that are related to that disease.
- The proposed Medicare system collects and processes all the EHR or omic data set in order to select best features using feature selection algorithms and categorize the disease using classification mining algorithms.
- The feature selection which extracts the original set of features into new subspace to select the best feature set that gives more accuracy and eliminates the unwanted feature.

- The purpose of the classification mining is to organize the data into categories of the disease for its most effective and efficient usage.
- It is planned to classify the mental illness disease data based on the symptoms of the disease for the better diagnosis of the patient.
- The users of the proposed Medicare System may be a patient or a nurse to send the query and gets the report of the patient.
- The doctor is also a peer user to the proposed Medicare System who executes the query and receives the suggestions for the drug in order to select a specific drug for the patient.

VI. CONCLUSIONS

The focus of this research is to propose a model for today's healthcare system using data mining techniques that can guide a doctor to provide right drug to right patient. There is various literature for personalized medicine approach that are studied and explained in this paper. The personalized medicine for mental illness disease is found to be taken care of because it treats the patient health by considering the individual profiles. Though mental illness is a chronic illness lasting from one day to months, patient may do inactive things which may lead to some other problem, its challenges and symptoms are also analyzed. In this paper, a proposed process model is developed considering the omic and Electronic Health Record data of a patient for mental illness disease. In future, the results will be compared by implementing various data mining algorithms for accuracy and proving suitability of the algorithm.

REFERENCES

- [1] McLaughlin, K.A., King, K., Developmental trajectories of anxiety and depression in early adolescence. *J. Abnorm Child. Psychol.* 43 (2), 2015, pp. 311-323.
- [2] Musliner, K.L., Munk-Olsen, T., Eaton, W.W., Zandi, P.P., Heterogeneity in longterm trajectories of depressive symptoms: patterns, predictors and outcomes. *J. Affect Disord.* 192, 2016, pp. 199-211.
- [3] Po-Yen Wu, Chih-Wen Cheng, Chanchala D. Kaddi, Janani Venugopalan, Ryan Hoffman and May D. Wang, "Omic and Electronic Health Record Big Data Analytics for Precision Medicine", *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, VOL. 64, NO. 2, FEBRUARY 2017, pp. 263-273.
- [4] G. H. Fernald *et al.*, "Bioinformatics challenges for personalized medicine," *Bioinformatics*, vol. 27, Jul. 2011, pp. 1741–1748.
- [5] L. Hood and S. H. Friend, "Predictive, personalized, preventive, participatory (P4) cancer medicine," *Nature Rev. Clin. Oncol.*, vol. 8, Mar. 2011, pp. 184–187.

- [6] Azorin JM, Bellivier F, Kaladjian A, Adida M, Belzeaux R, Fakra E, Hantouche E, Lancrenon S, Golmard JL. “Characteristics and profiles of bipolar I patients according to age-at-onset: findings from an admixture analysis”, *Journal of Affective Disorders* 150, 2013, pp. 993–1000.
- [7] Anthony T. Fojo, Katherine L. Musliner, Peter P. Zandi, Scott L. Zeger, “A precision medicine approach for psychiatric disease based on repeated symptom scores”, *VALUE IN HEALTH* 16(2013)S4 – S 9, <http://dx.doi.org/10.1016/j.jval.2013.06.005>, PP.147-155.
- [8] Ms. Sumathi M.R., Research Scholar , Dr. B. Poorna, Principal, “Prediction of Mental Health Problems Among Children Using Machine Learning Techniques”, (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 1, 2016, pp.552 – 557.
- [9] Mariantonia D Sanzo, “Clinical applications of personalized medicine: a new paradigm and challenge”, Feb 2017.
- [10] Shivangi Jain, Dr. Mohit Gangwar, “A Data Mining Analysis Over Psychiatric Database for Mental health Classification”, *International Journal on Future Revolution in Computer Science & Communication Engineering* ISSN: 2454-4248, Volume: 4 Issue: 1, pp. 241 – 246.
- [11] Shamsul huda¹, john yearwood¹, herbert f. Jelinek², mohammad mehedi hassan³, giancarlofortino⁴, and michael buckland⁵, “A Hybrid Feature Selection With Ensemble Classification for Imbalanced Healthcare Data: A Case Study for Brain Tumor Diagnosis”, 2169-3536 2017 IEEE. *Translations*, Volume4, 2016, PP. 9145-9154.
- [12] Shamsul Huda, John Yearwood, Andrew StranieriCIAO, “Hybrid wrapper-filter approaches for input feature selection using Maximum relevance-Minimum redundancy and Artificial Neural Network Input Gain Measurement Approximation (ANNIGMA)”, *GSITMS*, University of Ballarat, Victoria, Australia.
- [13] Subhas C. Misra, Sandip Bisui , “Modelling vital success factors in adopting personalized medicine system in healthcare technology and management”, *Science Direct, Engineering Science and Technology, an International Journal* 21 (2018), pp. 532–545.
- [14] Basma Boukenze, Hajar Mousannifand AbdelkrimHaqiq, “predictive analytics inHealthcare system using data Mining techniques”, DOI : 10.5121/csit.2016.60501, *CSE-2016* , © CS & IT-CSCP 2016 , pp. 01–09.