

Integrated Prediction System for Chronic Disease Diagnosis to Ensure Better Healthcare

Geetha Poornima K.^{1,2} & Krishna Prasad K.³

¹Research Scholar, Srinivas University, Mangaluru, Karnataka, India

²Assistant Professor, Dept of Computer Science, St Philomena College Puttur, Karnataka, India

³College of Computer Science & Information Science, Srinivas University, Mangaluru, Karnataka, India

Email: poornima.sanjay@spcputtur.org

Area/Section: Health Sciences.

Type of the Paper: Research Paper.

Type of Review: Peer Reviewed as per [C|O|P|E|](#) guidance.

Indexed in: OpenAIRE.

DOI: <http://doi.org/10.5281/zenodo.3773844>.

Google Scholar Citation: [IJHSP](#)

How to Cite this Paper:

Geetha Poornma, K. & Krishna Prasad, K. (2020). Integrated Prediction System for Chronic Disease Diagnosis to Ensure Better Healthcare. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 4(1), 25-39. DOI: <http://doi.org/10.5281/zenodo.3773844>.

International Journal of Health Sciences and Pharmacy (IJHSP)

A Refereed International Journal of Srinivas University, India.

© With Authors



This work is licensed under a [Creative Commons Attribution Non-Commercial 4.0 International License](#) subject to proper citation to the publication source of the work.

Disclaimer: The scholarly papers as reviewed and published by the Srinivas Publications (S.P.), India are the views and opinions of their respective authors and are not the views or opinions of the SP. The SP disclaims of any harm or loss caused due to the published content to any party.

Integrated Prediction System for Chronic Disease Diagnosis to Ensure Better Healthcare

Geetha Poornima K.^{1,2} & Krishna Prasad K.³

¹Research Scholar, Srinivas University, Mangaluru, Karnataka, India

²Assistant Professor, Dept of Computer Science, St Philomena College Puttur, Karnataka, India

³College of Computer Science & Information Science, Srinivas University, Mangaluru, Karnataka, India

Email: poornima.sanjay@spcputtur.org

ABSTRACT

Technology innovation has made life easy for human beings. Technology is being used everywhere. This also extends to the healthcare sector. The healthcare sector produces a large amount of data each minute. Because of privacy issues, much of the data generated is not used and is not publicly accessible. Healthcare data comes from diverse sources hence it will be always varied in nature. Keeping track of such data has become much easier these days. Predictive analysis in healthcare is an emerging technology that identifies the person with poor health where the risks of developing chronic conditions are more likely and provide better solutions in the field of healthcare. Statistical methods and algorithms can be used to predict the disease before the actual symptoms are revealed in humans. By using data analytics algorithms one can easily predict chronic diseases such as obesity, high/low Blood Pressure, diabetes, asthma, cardiopulmonary disorders. Because of an unhealthy diet, lack of proper exercise, stress, consumption of tobacco, alcohol, etc. chronic diseases are most common these days. If the symptoms of chronic diseases are detected in the early stages, there will be less risk of hospitalization by cost-effectively maintaining better health. Big data analysis and health care can be mixed to produce accurate results. The application of predictive analytics in healthcare is highlighted in this paper. It provides a broader analysis in the prevention of different chronic diseases by using predictive analytics. The paper also includes various issues that arise when handling health care data. For each chronic disease, diverse models, techniques, and algorithms are used for predicting and analyzing. The paper comprises a conceptual model that integrates the prediction of most common chronic diseases.

Keywords: Predictive Healthcare Analytics, Chronic diseases, EHR, Clinical data, EDW, Wearable devices.

1. INTRODUCTION :

These days, data are in tremendous demand. Each click on the Internet, every heartbeat of a patient on a monitor, every transaction of a customer produces large amount of data that is of some use someone. The voluminous data can now be easily processed for decision-making purposes. Data is turned into information that is used to deliver better performance when used effectively by algorithms. The timely availability of crucial medical data will save people's lives in emergencies and has thus become an important concern for the healthcare community. Predictive Analytics (PA) is a branch of analytics that uses historical data and

different emerging technologies such as Artificial Intelligence, Machine Learning, Data Analytics, etc. for decision making. Healthcare is of vital importance where algorithms can be used to gain insights into patients' diagnoses and identify reliability bottlenecks. Chronic diseases (ChrDs) like obesity, cancer, diabetes, cardiovascular disorders, asthma, etc. are the causes of life-long misery and suffering. With the help of data analytics algorithms, PA provides automated solutions that predict future outcomes for patients in the case of ChrDs. Healthcare costs are becoming highly unaffordable these days. One key value to healthcare institutions is to minimize the use of costly resources and make use of technology to

detects possible risks. Prediction and prevention are closely related to each other. If the risks of getting into chronic situations are predicted early, the long-term sufferings of patients can be avoided [1]. This paper highlights the objectives of the study, General idea of Predictive Analytics, PA and healthcare data, Prevention of chronic diseases using PA, Conceptual model for PA, use cases of PA, Challenges in PA.

2. OBJECTIVES OF THE STUDY :

The objectives of the study are:

- To identify the components of PA systems
- To know the use of PA in the prevention of ChrDs
- To understand the importance of PA in the prevention of ChrDs
- To develop a conceptual model for predicting ChrDs
- To comprehend challenges in PA related to prediction and prevention of ChrDs

3. RESEARCH METHODOLOGY :

The study was conducted by referring to scholarly documents available online and by referring to websites of companies offering healthcare-related services in PA. A conceptual model is developed based on the theoretical perception that incorporates all the techniques needed for PA to predict most common ChrDs [2]. The model is built through the analysis of existing information and it is in its initial stage. It uses only abstract ideas. To test the feasibility of the model, further study is necessary.

4. ABOUT PREDICTIVE ANALYTICS :

PA is a branch of analysis that uses historical data and all the emerging technologies such as IoT, AI, Machine Learning, Deep Learning, etc. of the purpose of decision making. It is used by industries to increase their business. It uses algorithms of data analytics and machine learning techniques to analyze data and determine the possibility of future events. It consists of a wide variety of techniques related to Data Mining, Predictive Modeling, Deep Learning, etc. Predictive models use historical data to identify possible risks and opportunities associated with a decision [3]. The data required for PA come from a wide range of sources. Hence it will be an unstructured one. For efficient decision-making reasons, it must

be processed and analyzed. PA uses historical data to analyze and identify a solution to the problem then it uses real-time data to determine future events. Based on historic data available, a model is built. This model uses algorithms and techniques to analyze the situation. Then the model uses current or real-time data to predict future events exactly. It is used by industries to increase their business. It is also used in healthcare for providing better healthcare services[4].

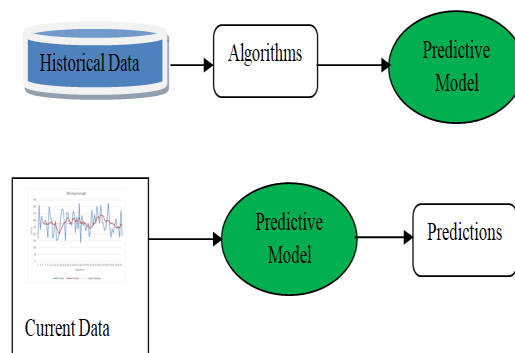


Fig.1: Predictive Analytics

Figure-1 describes the process of PA. A model is designed based on historical data or training set. An algorithm is trained on the training set to perform prediction. Then the real-time data is fed to the model to make an accurate prediction. Based on both historical as well as current data, the model generates predictions. To evaluate stored data, a predictive model uses different techniques. It is possible to classify predictive models as [5] below :

- **Classification model:** This model is used for historical data classification. Once all data related to a particular problem is obtained, the model can be used to categorize it. It uses several classification algorithms to infer from a given training set, such as the Machine Learning algorithm, the Naïve Bayes algorithm, the K-Nearest neighbor algorithm, etc.
- **Clustering Model:** This model splits thousands of records into smaller clusters. If the records are identical, they are put in the same cluster. K-means algorithm, K-medoid algorithm, Expectation-Maximization (EM) clustering, Agglomerative clustering are some of the data classification algorithms used by this model.
- **Forecast Model:** The quantitative value is estimated based on the data

harvested. It is the empirical model that is most commonly used. When historical and numerical data are available, this model can be used. Forecasting is a technique that uses historical data to predict future events. This uses algorithms for machine learning to reliably predict future events.

- **Outlier-detection Model:** this deals with the discovery of anomalous data in historical records. This is done to recognize noise in the data which, if present, will reduce the PA model's effectiveness.

- **Time-series Model:** It describes the data points collected where time is one of the inputs. As data points are captured at different time intervals, it is also called as a temporal model. To determine the future of an event it uses curve-fitting and various forms of smoothing techniques.

- **Logistic Regression Model:** Using statistical techniques, data values are predicted based on the training set and prior observation. It is used when categorical data is to be graded.

5. RELATED WORKS :

The conventional HC is passive. Due to this type, patients can die due to a lack of proper treatment during emergencies such as cardiac arrest. The key to improving HC efficiency is to reduce the mortality rate due to lack of proper treatment and to transform the passive HC program into a continuous one at a reduced cost. The number of people suffering from ChrDs is rising day by day. The biggest downside of the conventional HC program is that the service is only available in hospitals. It is difficult for elderly people suffering from ChrDs to meet their HC requirements in emergencies. While elderly people live on their own (especially during the day), the goal is to provide ubiquitous HC. The use of sensors and wearable devices that effectively track and transmit data such as body temperature, blood pressure, heart rate, or any irregular condition to people such as physicians, clinical staff, or patient caregivers. There are apps that can track sleep patterns, falls, incorrect patient postures, and provide appropriate warning messages.

ChrDs such as diabetes and blood pressure can lead to further diseases such as heart attack, kidney failure, or paralysis. When predicted at early stages, potential damage, as well as

discomfort, can be minimized. Parkinson's is a chronic nervous disease that affects movement. Genetic and environmental elements are the primary causes of this disease. If signs are identified at an early stage, the progression of the disease may be prevented.

Detecting cancer at an early stage increases the survival rate of patients. There will be a chance of long-term remission. Data mining techniques can predict the risk of developing cancers in patients on the basis of genetic and non-genetic factors. If there is a greater chance of developing cancer, then screening, PET testing, and further treatment should be prescribed. Cardiovascular disease is another health problem that affects a large patient population. When previously detected with PA, the survival rate of patients would be increased.

Health data are derived from both internal and external data sources. Internal data sources include patient data, hospital data, diagnostics, and treatment data. External data sources include statistics, publicly accessible data sets, and data from government departments or institutions such as WHO. Various sources contain data of varying quality and use inconsistent representations, code, and labels, which must be reconciled.

The problem is the integration, purification, and standardization of data in the preparation of analytical tasks. Also, data may come from various, diverse systems (e.g., EHRs, clinical decision support systems, etc.) located at different places within the larger health network. Health data can be either structured (quantitative) or unstructured (text). Always such data must be organized, categorized, and standardized in such a way as to prepare them for analysis [6-16].

The application of PA has accelerated a number of industries. Healthcare is a prominent one among them. Healthcare analytics refers to the systematic use of health-related data and computational algorithms, predictive, cognitive, and other models to gain insights. It is capable of improving the quality of life and reducing preventable deaths. By using statistical techniques, machine learning and data mining, etc., it can forecast epidemics, cure illnesses, analyze current and historical data, and make accurate future predictions. For example, it can provide treatment options for patients with chronic illnesses, reducing

unnecessary re-admission to hospital or unnecessary ICU admissions [17].

Healthcare data is very complex. It includes data from different sources such as Clinical reports of patients, doctors, clinical records, Laboratory reports, X-Ray images, Ultra-sound Scans, EHR (Electronic Health Record) readings, PET scans, Blood reports, MRI scans, etc. In addition to this, data can also be gathered from devices such as smart-phones, smart bands, etc. The data collected in this way will be an unstructured one. The data from these distinct resources is captured automatically to understand the present situation of the patient. Then big data analytics algorithms are applied to the selected data to know the future situation of the illness. PA combines several emerging technologies such as Artificial Intelligence, Machine Learning, and Big Data Analytics [18].

Many industries have boosted the deployment of PA. Healthcare is one of the popular among them. In reality, information regarding healthcare is very complicated. This contains data from multiple sources such as Patient Clinical Reports, Laboratory Reports, X-Ray images Ultra Sound Results, EHR readings MRI scans, etc. Other than this, data can also be captured from devices such as smartphones, smart bands, etc. The reports from these distinct tools are automatically collected to determine the patient's present situation. Big data analytics algorithms are then applied to the selected information to know the progression of the disease and future situation. PA can use the readily available patient-related data such as historic data, hereditary records, details of hospital admission, etc. to understand the current situation and predict future risks involved. Using PA can decrease the number of ICU admissions as the current situation is available as real-time data and the software could be used to predict the future condition of the patient accurately. It lowers medical costs and provides efficient healthcare services. The information will be made available to physicians, clinical staff, and patients in real-time. Despite several legal and ethical issues associated, PA is extensively used in countries like UK and the US to provide customized healthcare solutions [19].

6. CHRONIC DISEASES AND PREDICTIVE ANALYTICS :

PA uses historic data of patients stored and analyzes it using algorithms. It uses the patient's real-time data to understand the current situation to foretell the severity of the disease. It can also provide treatment recommendations to the doctor. It results in the best possible healthcare at a lower cost. Predictive technologies focus on identifying patients at high or low risk for severe complications. Situations such as re-admission to hospitals and admission to ICUs can be predicted accurately. Health care (HC) analysis is a systematic approach using clinical data for decision-making purposes. HC tools are decision-support systems that are used to make decisions easier for doctors, nurses, pharmacists, etc. HC models gather data from different sources and create models to perform what-if analysis. Individuals with particular ChrDs can be identified and constantly monitored to prevent certain disease-related medical conditions from progressing further. HC solutions may also suggest alternative plans for treatment and diet plan based on the current situation [1].

7. PROPOSED SYSTEM FOR PREDICTING CHRONIC DISEASES :

Figure-2 shows the conceptual design of the system. The system requires storing a vast amount of data. The sources of input may be Patient's case history, X-ray images, hereditary information, lab reports, clinical records, written documents such as prescriptions, ultrasound or MRI scan images, readings of EHR, or the outputs of smart devices. Raw data is to be stored in the EDW and then the noise in the data can be eliminated using various techniques such as filtering. The validated data is used for analysis where, based on the problem, various algorithms and tools are applied [20]. Since most analytical tools are open source, they can easily be integrated into a single application. Different diseases will have different symptoms and will use different algorithms. It is therefore important to store large amounts of data in the EDW, which uses the cloud platform as storage. The theoretical framework incorporates all the requisite tools and techniques for chronic disease prediction. It's going to be complicated as it integrates many emerging technologies such as AI, ML, Deep Learning, Data Mining, and so on. As the up-to-date information of the patient is

available on the cloud, it can be accessed from anywhere at any time [21].

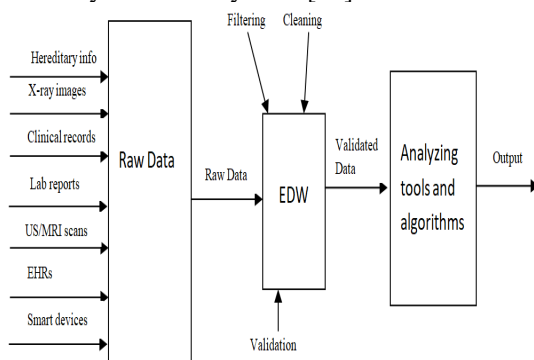


Fig. 2: Conceptual Design of the System

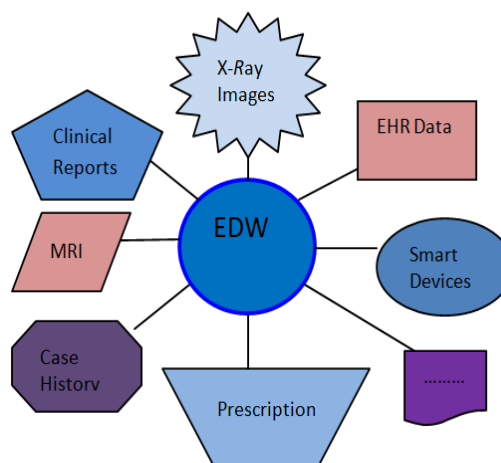


Fig. 4: Data in Enterprise Data Warehouse

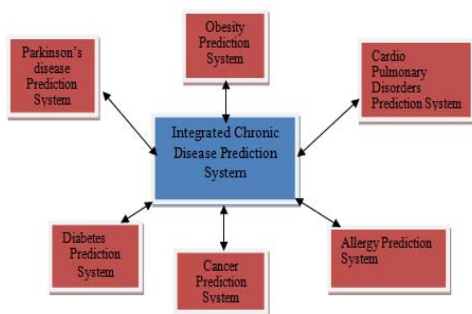


Fig. 3: Integrated Chronic Disease Prediction System

The study of PA for the prediction of ChrDs indicates that various models are used to identify various ChrDs. As shown in Figure-3, the conceptual model suggests a single system for identifying the most common chronic diseases. It is possible to store the raw data required for the study from different resources in a single data warehouse. It will be there in an unstructured form. Once the required data is collected, different techniques can be used to validate it and the noise data can be filtered. The final result is extracted using different analyzing tools and techniques [21].

8. WAYS TO DESIGNING PREDICTIVE MODEL FOR HEALTHCARE

PA is a systematic approach to find the best possible solution to a given problem related to healthcare. It involves two steps [1]. They are

- Creation of Enterprise Data Warehouse (EDW)
- Selection of Model for PA

8.1 EDW:

Figure-4 refers to the diversity of the information to be handled for PA. This is because it is acquired from different resources. Clinical data is structured or unstructured one. It may include the medical history of patients, clinical prescriptions, laboratory reports, Scanning reports, etc. PA requires both historic and real-time data of patients for analysis. A centralized repository is required to store this information. EDW the repository unit to be used to hold different kinds of data for analysis purposes. It becomes easy to analyze when data is stored in a centralized data warehouse. [22]

8.2 Selection of Model for PA:

Predictive modeling is an evolutionary approach as the analysis of healthcare data is very risky. Predictive modeling analyzes the data stored in the EDW using algorithms. There may be more than one solution to a problem, but to get accurate results, the best possible solution is to be identified [20].

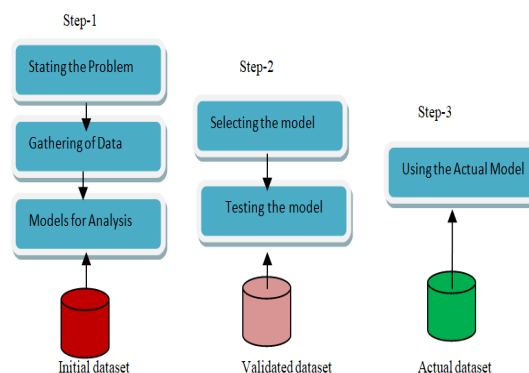


Fig.5: Steps in Predictive Modeling
Steps in predictive modeling are shown in

Figure-5.

- The first step indicates the analyst has to understand the problem clearly. Once the problem is understood, it has to be stated clearly. All the necessary data needed for PA is to be stored in the EDW. Then several algorithms are tested on a trial-and-error basis to find the initial solution. The dataset in the initial step contains unstructured data that may contain noise in it.
- In step-2 the best model is selected and tested, data that contains noise is discarded and the EDW contains only a validated dataset.
- In the final step of PA, the solution is tested on real-time data to arrive at a conclusion.

The predicting method for ChrDs requires three steps: grouping, attribute collection, and evaluation. The data is to be grouped into various clusters in the classification. Data can be classified using techniques such as KNN, Decision Trees, and Random Forest. Attributes that are most relevant to the PA are to be chosen automatically. Redundant data may sometimes be present or there may be noise in the data. The task of selecting attributes is to include the most important attribute, excluding the outliers, and automatically classify the redundant data. After choosing the attributes, various predictive algorithms should be used and the outcome needs to be evaluated, and based on the accuracy one technique is to be chosen for the PA [23].

1. Diabetes Prediction System:

Diabetes is a ChrD where there is no proper cure. The right medications, diet, and exercise will regulate it. The conditions vary from person to person and medication and diet need to be recommended depending on the level of blood sugar. Diabetes mellitus and diabetes insipidus are two types of diabetes. PA can be used to assess diabetes mellitus. Diabetes mellitus risk factors may include family history, environmental factors, dietary habits, excess weight, lack of exercise, high blood pressure, and cholesterol levels. Patients with diabetes mellitus are required to undergo a full clinical examination to test glucose levels, blood pus levels, and kidney function. Values for attributes such as age, gender, blood pressure, RBC count, WBC count, random level of blood

glucose, appetite, level of albumin, level of creatinine, level of potassium, pedal edema information, wounds etc. can be collected from clinical examination. Apart from this, it is also essential to collect and store historical data of patients. If the illness is not treated properly, the working of the kidneys may be disordered. The patient must undergo dialysis or transplantation once the kidneys stop working, which is expensive and painful. Based on inputs such as disease extent, functional kidney abnormalities, and Glomerular Filtration Rate (GFR), disease progression can be predicted. Once all the necessary data is obtained, techniques such as KNN, Decision Tree, SVM, Logical Regression or Random Forest will be applied to the data in order to forecast the disease's future situation. If the disease progresses, hospital admission may be needed over and over again. The disease's future condition can be anticipated by PA. Hadoop platform can be used to evaluate store situation and forecast it [23-26].

2. Allergy Prediction System:

Allergy is a chronic disease that can be found in people of all ages. Low immunity is the main cause of allergy. Allergy causes the body's immune system to react hyper. The number of allergy sufferers is growing day after day. The allergen cannot be identified as there are as many components that cause it.

Allergy symptoms and signs can vary from person to person. Nevertheless, the most common symptoms include blocked nose, itchy and watery eyes, cough, runny nose, flaking or burning skin, skin rashes, swollen lips, facial and throat swelling or shortness of breath, wheezing and rectal bleeding. Risk factors for allergy include family history and poor immune systems.

Allergy factors include ingredients such as chocolates, nuts, egg whites, milk, shellfish, or gluten found in wheat. Saliva, wool or pet hair, penicillin-containing medications, salicylates, or sulfonamides may be the cause of an allergy. Insect bites, vegetables such as eggplants, leaves, stems, or roots of colocasia can also cause allergy.

There are several tests to check for allergy. Popular among them are:

1. **Blood Test:** It measures the amount of Immunoglobulin E (IgE) antibodies produced by the body's immune system.

High IgE rates mean the body is battling against acute allergy. The check cannot figure out which allergen is the cause for the disease whereas it can show how serious the condition is.

2. **Pierce Test:** It is generally known as a puncture or prick test. It is the traditional method of diagnosis. The skin of the patients is impregnated with a small amount of potential allergen. If the skin reacts, the reaction is caused by the resulting allergen. The reaction can be in the form of skin swelling, skin rashes, or itchy skin.
3. **Patch test:** It procedure uses spherical metal disks with minute quantities of potential allergens that are then rubbed onto the back of the patient. After 48 hours the skin is then examined by a doctor. This is replicated in most situations after a couple of days.

Such experiments are time-consuming and expensive. To identify the real allergen, they might need to be repeated many times. Instead, ML, Regression Tree, Statistical Techniques and Clustering can be used to find the causes of allergy and suggest preventive measures [27-28].

3. Cancer Prediction System:

Cancer prediction methods can be used to predict cancer susceptibility or risk, cancer recovery, and cancer recurrence. Instead of carrying out tissue-destructive tests or focusing on genomics, AI algorithms can reap the benefits of image information to identify patients with more severe disease and may require more intensive care. It will also let physicians know which patients have less aggressive cancer, and may be able to avoid chemotherapy side effects. This can be easily predicted by using ML, ANN, SVM. With the aid of ML, physicians can better understand the nature of cancer development. The cancer prediction system can also be used to estimate illness-related mortality. The system can be used to analyze the most common cancers such as cervical cancer, oral cancer, and breast cancer [29].

Nowadays cervical cancer is the most common disease, the detection of which involves time-consuming clinical tests. Cervical Cancer is a condition in which the cervix, is surrounded by thin layers of tissue consisting of normal cells

that transform into cancerous ones, which continue to develop and multiply quite rapidly than normal, and thus transform into a tumor. Medical image processing and tissue texture analysis methods make it possible to screen cervical cancer early and monitor abnormal cells until it reaches the chronic stage. The Intelligent framework technique to analyze malignant cells is cost-effective and has an advantage over conventional detection approaches. AI and ML are considered to be more suitable for the accurate diagnosis of cervical cancer. These techniques can be used to speed up the diagnosis. Prediction requires an examination of the diagnostic images of the pap smear cells. SVMs, DT, CART are the various techniques used in the analysis. In addition to a successful diagnosis, the use of predictive algorithms can reduce the expense of unnecessary procedures for cervical cancer screening [30], [31].

Breast cancer is the main health hazard most frequently diagnosed in women aged 40 years and older. Breast cancer is painless in its early stages. Therefore, it is not identified in most cases during the early stages. In the breast, there may be malignant and non-malignant cysts. When the cyst becomes malignant, it starts to develop exponentially, causing cancer to develop. Physicians need to establish a technique to classify tissues. It can be achieved effectively by using AI, ML techniques. Techniques such as DT, SVM, LR are found to be the most effective in breast cancer PA. To diagnose the disease, estimate the survival rate, and prescribe adequate treatment, genetic and family factors, hormone changes, other drug details, and environmental factors must be analyzed efficiently [32], [33].

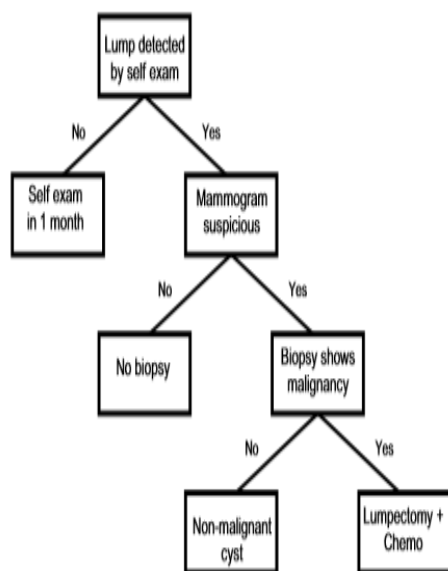


Fig. 6: Decision Tree for Breast Cancer Diagnosis [29]

Figure-6 is an example of a basic decision tree that may be used in the diagnosis and treatment of breast cancer. The early detection of oral cancer leads to an early treatment which reduces the risk of morbidity and mortality. Diagnosis of oral cancer remains a challenge particularly in the identification, assessment, and control of oral cancer in the early phase. The machine learning approach is used to diagnose and identify oral cancer at an earlier level, due to a lack of timely detection using traditional methods. Machine learning algorithms are used to model cancer diagnosis progression and treatment, and can accurately predict future outcomes of cancer. To provide adequate care for the cancer patient, it is important to identify different stages in Oral Cancer. The disease-related features extracted can be categorized using the decision tree and random forest compared to other common classification methods such as SVM, KNN, MLP, and LR [34], [35].

4. Parkinson's Disease Prediction System:

Parkinson's disease is a progressive disease that affects movement. It is a disease that seriously threatens and decreases the quality of life for elderly people. There is no 100% reliable method of testing the condition of Parkinson. Machine learning algorithms can predict it. Inputs such as patients' voice and behavioral changes are analyzed using statistical techniques such as Student's t-test and 2 Sample KS test and decision tree to assess the level of

impairment and disease progression [36], [37].

5. Cardio-Pulmonary Disorder Prediction System:

In developing countries, heart disease is the main cause of death. In the clinical data analysis section, the prediction of cardiovascular disease is considered as one of the most significant topics. Because of many contributory risk factors such as diabetes, high blood pressure, elevated cholesterol, irregular pulse rhythm, and several other factors, it is difficult to detect heart disease. Because of these limitations, scientists have turned for predicting modern methods such as data mining and machine learning. The forecast needs to be tested on many variables. In the case of cardiopulmonary disorders, it is very important to predict events. Machine learning algorithms including Random Survival Forest and Parametric Decision Trees can be used to evaluate the patient's exact situation and predict future events [38], [39].

6. Obesity Prediction System:

Obesity is a worldwide epidemic believed to include many severe medical complications. The number of people suffering from obesity is rising day by day. Obesity is a severe ChrD, which in turn can lead to many other illnesses such as BP, stroke, diabetes, kidney disease, etc. The obese patient has more chances of developing more health issues. The obesity risk factors include the history of the family, pre-existing medical conditions. The main reason for obesity is more calorie intake than is normal. Obesity is difficult to estimate using traditional approaches. It can be predicted using data analytical algorithms that analyze by applying grocery data and data analysis, various dietary intake preferences can be explored to predict the percentage of individuals in a household being obese; this involves calories, micronutrients, food groups, and food intake categories (raw & processed). Using decision tree classification techniques, the clustering algorithms can be used to predict obesity and its future development in patients [40], [41].

9. COMPONENTS OF THE PROPOSED SYSTEM :

Figure-7 shows the components of ICDPS. The system uses training data sets to analyze the problem. Information from different resources is maintained in the hospital's EDW or Cloud

platform of a trusted third party. This data processed is called 'training data'. Such data are used for research purposes. On the training data, Machine Learning (ML), Artificial Neural Network (ANN), or Data Mining techniques are applied and then the model is evaluated on the patient's real-time test data to predict the disease and provide solutions. If everything related to the disease is available, the data can be analyzed using ML or DM techniques. Patients may not like to share all information related to them in certain cases. Thus, some data required will be missing. ANN can be used to analyze the data in such cases [42].

The components of this system are

- 1. Data Acquisition System:** The data about HC is heterogeneous. It includes structured, unstructured, and semi-structured data. To build a state-of-the-art predictive system, data from several sources must be collected efficiently. Quantitative data such as temperature, blood pressure, and heart rate may be controlled by using sensors and wearable devices mounted on the patient's body. Such devices send the data to the access point daily. Also, clinical notes, case history, X-ray images, MRI scans, etc. must be accurately registered.
- 2. Data Store:** It is difficult to store a large volume of heterogeneous data using conventional DBMS. As a result, private or public clouds may be used to store data. When used, confidentiality and privacy must be protected. The model uses EDW for storing data efficiently. When data is stored using the cloud, security, and privacy of data need to be ensured. The system requires both the test set and the training set of data for analysis. Hence a large amount of data needs to be stored to perform an accurate analysis.
- 3. Data Analyzing System:** This system takes care of the decision-making part. Various data analyzing techniques such as clustering, regression tree, ANN, support vector machine, etc are used to perform the critical task of analyzing the current situation of the patient based on the data stored in EDW and real-time data from wearable devices. Choosing the right algorithm for the right disease is a most

challenging task and it is very much essential as the analysis is very critical.

- 4. Data Visualization System:** Once data is analyzed, it can be presented in the form of a report or a graph to the patients. By incorporating visualization tools, the system makes creates reports which are to be given to patients as well as physicians. [17], [43].

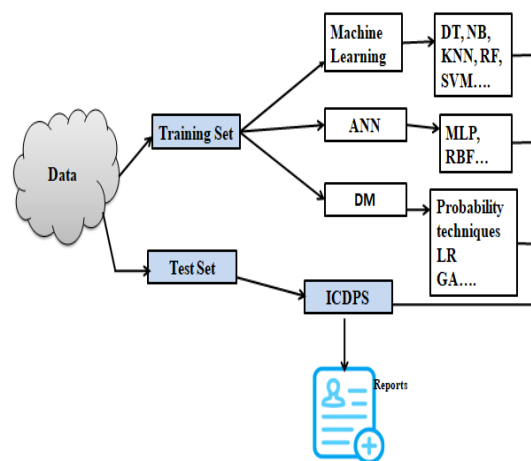


Fig. 7: Components of Integrated Chronic Disease Prediction System

Figure-7 shows the components of ICDPS. The system uses training data sets to analyze the problem. Information from different resources is maintained in the hospital's EDW or Cloud platform of a trusted third party. This data processed is called 'training data'. Such data are used for research purposes. On the training data, Machine Learning (ML), Artificial Neural Network (ANN), or Data Mining techniques are applied and then the model is evaluated on the patient's real-time test data to predict the disease and provide solutions. If everything related to the disease is available, the data can be analyzed using ML or DM techniques. Patients may not like to share all information related to them in certain cases. Thus, some data required will be missing. ANN can be used to analyze the data in such cases [42].

The components of this system are

- 5. Data Acquisition System:** The data about HC is heterogeneous. It includes structured, unstructured, and semi-structured data. To build a state-of-the-art predictive system, data from several sources must be collected efficiently. Quantitative data such as temperature,

blood pressure, and heart rate may be controlled by using sensors and wearable devices mounted on the patient's body. Such devices send the data to the access point daily. Also, clinical notes, case history, X-ray images, MRI scans, etc. must be accurately registered.

6. **Data Store:** It is difficult to store a large volume of heterogeneous data using conventional DBMS. As a result, private or public clouds may be used to store data. When used, confidentiality and privacy must be protected. The model uses EDW for storing data efficiently. When data is stored using the cloud, security, and privacy of data need to be ensured. The system requires both the test set and the training set of data for analysis. Hence a large amount of data needs to be stored to perform an accurate analysis.
7. **Data Analyzing System:** This system takes care of the decision-making part. Various data analyzing techniques such as clustering, regression tree, ANN, support vector machine, etc are used to perform the critical task of analyzing the current situation of the patient based on the data stored in EDW and real-time data from wearable devices. Choosing the right algorithm for the right disease is a most challenging task and it is very much essential as the analysis is very critical.
8. **Data Visualization System:** Once data is analyzed, it can be presented in the form of a report or a graph to the patients. By incorporating visualization tools, the system makes creates reports which are to be given to patients as well as physicians.
[17], [43].

10. IDENTIFIED BENEFITS :

The proposed system will ensure continuous care for the patients suffering from any ChrDs mentioned in the block diagram during hospital admission as well as after discharge [44].

- **Critical care during hospital admissions:** Higher death rates in patients with ChrDs are due to persistent or over-use of medicines such as antibiotics. The

actual situation of patients can be evaluated exactly through the use of PA. Based on the data of the patient stored in EDW and their current situation, clustering algorithms can be used to reliably categorize data as low, medium, and high-risk patients and then provide adequate healthcare when they are hospitalized.

- **Post-discharge Treatment:** Hospital readmission is an important cause of expenditure for patients with ChrDs. In most cases, after the discharge from hospitals, there will be no continuous communication between the patients and the doctors. Patients are given wearable devices to ensure quality healthcare during the post-discharge period. Such wearable devices will monitor the heart rate, blood pressure, blood glucose levels, and other information and regularly send the data to the EDW. Doctors may use PA to assess patient condition and recommend the treatment and diet plans required. After the discharge from hospitals usually, the follow-up activity will not be there or it requires the process of fixing appointments, carrying various reports related to the patient, etc. Patients can be given some application/software where they can be asked to enter their food and medication details every day so that diet and exercise plan can be suggested by the doctors based on the information provided by patients. This ensures two-way communication between doctors and patients even after discharge from the hospital. The EDW maintained by hospitals will store all information of patients. This can be centralized in such a way that the data can be shared by other doctors or specialists. This will relieve the patients from keeping multiple hard copies of reports. It guarantees continued treatment, even after discharge from hospitals.

The model proposed can generate a prediction for six different ChrDs. Only three most commonly found cancers namely Oral cancer, Breast cancer and Cervical cancer are included in current model. In future, other ChrDs such as Kidney diseases, Liver diseases, Alzheimer's disease, some cancers such as Prostrate cancer, Gall bladder cancer, Lungs cancer can also be

included. ChrDs such as asthma, blood pressure and hyper tension which are considered as 'silent killers'. Their prediction may also be included. When treatment is provided for multiple diseases. Inherent complexity and life-critical nature of HC systems make the task of designing the solution challenging one [44], [45].

12. DISCUSSION & FUTURE WORK :

The application PA in HC is discussed in this work especially in the prediction of ChrDs. To form a conceptual model, the contributions of different authors were incorporated. The key to every prospective predictive model is its ability to make better predictions. To make the traditional HC system simpler and efficient use of PA is very much essential. The endless stream of data, multiple techniques, and widespread tool availability certainly triggered PA in HC. Because of this, it is likely to generate many outcomes when the symptoms are fed to the model. In HC, it depends entirely upon patient feedback and the physician's decision-making capabilities. The outcome of a predictive model in HC is very crucial since it is not just in terms of graphs or reports, but also in terms of something that is actionable and life-critical.

The model comprises six major processes which provide only theoretical analysis. Future work to be developed into the PA will combine these processes with new technologies and will use predictive control models to test the ChrDs and provide accurate results. By making productive use of emerging technologies to analyze large datasets and reveal actionable insights that effectively address HC challenges. Usage of machine learning in data processing will improve productivity. The analyst can easily identify data with serious complications by training an algorithm to recognize severe health-care issues in the data related to six major ChrDs addressed in the model.

13. CONCLUSION :

PA has brought major improvements to the healthcare sector. With the aid of PA, the difficult and life-critical tasks such as chronic disease diagnosis are made easy and reliable. It has brought about revolutionary changes in hospital, clinic, and laboratory procedures. By analyzing historical and real-time data, doctors can predict the future situation of patients.

Wearable devices and EHRs have made the patient-related data omnipresent. Daily jobs such as scheduling appointments, waiting for health reports, bringing clinical records to hospitals for every visit, routine check-ups such as BP, Blood-Sugar test, checking the temperature, and other tests are removed as real-time data are available. It would be simpler for both patients and doctors if a single model is used to examine the most common chronic diseases. Once the data is stored in the centralized EDW, anywhere it can be accessed. PA-based HC will increase the efficiency in human operations, reduce medical errors, storage, and HC costs, and improve the lives of patients who are suffering from ChrDs. PA can be used to provide improved accuracy, transparency, and cost-effective treatment. As healthcare institutions enter into a partnership with information technology companies to develop more advanced models, challenges still exist related to the inherent complexity of patient data and the need to integrate the results of predictive models within existing physician workflows. The model designed in this paper contains six major components and it provides only the conceptual view. It is designed by integrating the ideas of different contributors in the related field.

REFERENCES :

- [1] Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., & Lo, B. (2014). The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Affairs*, 33(7), 1139–1147. DOI: <https://doi.org/10.1377/hlthaff.2014.0048>
- [2] Retrieved from <https://simplyeducate.me/2015/01/05/conceptual-framework-guide/> on 25-12-2019.
- [3] Wikipedia, Predictive Analytics. Wikipedia. Retrieved on 25-12-2019 from https://en.wikipedia.org/wiki/Predictive_analytics.
- [4] What is Predictive Analytics? 3 Real-World Examples of Predictive Analytics in Business Intelligence. *Logi Analytics*. Retrieved on 25-12-2019 from <https://www.logianalytics.com/>

- [predictive-analytics/what-is-predictive-analytics/](https://www.ijhsp.com/predictive-analytics/what-is-predictive-analytics/)
- [5] Top 5 Predictive Analytics Models and Algorithms | Logi Analytics Blog. *Logi Analytics*. Retrieved on 26-12-2019 from <https://www.logianalytics.com/predictive-analytics/predictive-algorithms-and-models/>
- [6] Li, C., Hu, X., & Zhang, L. (2017). The IoT-based heart disease monitoring system for pervasive healthcare service. *Procedia Computer Science*, 112, 2328–2334. DOI: <https://doi.org/10.1016/j.procs.2017.08.265>
- [7] Chen, Y., Hao, H. W., Hu, Y. X., & Li, L. M. (2013). Wireless wrist-wearable wake/sleep identification device for closed-loop deep brain stimulation. *Electronics Letters*, 49(7), 497–499. DOI: <https://doi.org/10.1049/el.2013.0382>.
- [8] Viju Raghupathi, W. R. (2013). An Overview of Health Analytics. *Journal of Health & Medical Informatics*, 04(03), 1–11. DOI: <https://doi.org/10.4172/2157-7420.1000132>.
- [9] Uniyal, D., & Raychoudhury, V. (2014). *Pervasive Healthcare-A Comprehensive Survey of Tools and Techniques*. Retrieved from <http://arxiv.org/abs/1411.1821>.
- [10] Dhar, V. (2014). Big Data and Predictive Analytics in Health Care. *Big Data*, 2(3), 113–116. DOI: <https://doi.org/10.1089/big.2014.1525>.
- [11] Peck, J. S., Benneyan, J. C., Nightingale, D. J., & Gaehde, S. A. (2014). Characterizing the value of predictive analytics in facilitating hospital patient flow. *IIE Transactions on Healthcare Systems Engineering*, 4(3), 135–143. DOI: <https://doi.org/10.1080/19488300.2014.930765>.
- [12] Harisekaran, S. S. D. G. (2015). Big Data Analytics Predicting Risk of Readmissions of Diabetic Patients. *International Journal of Science and Research (IJSR)*, 4(4), 534–538. Retrieved on 25-12-2019 from <https://www.ijsr.net/archive/v4i4/SUB152923.pdf>.
- [13] Bardhan, I., Oh, J. H., Zheng, Z., & Kirksey, K. (2015). Predictive analytics for readmission of patients with congestive heart failure. *Information Systems Research*, 26(1), 19–39. DOI: <https://doi.org/10.1287/isre.2014.0553>.
- [14] Parikh, R. B., Kakad, M., & Bates, D. W. (2016). Integrating predictive analytics into high-value care the dawn of precision delivery. *JAMA - Journal of the American Medical Association*, 315(7), 651–652. DOI: <https://doi.org/10.1001/jama.2015.19417>.
- [15] Malik, M. M., Abdallah, S., & Ala'raj, M. (2018). Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review. *Annals of Operations Research*, 270(1–2), 287–312. DOI: <https://doi.org/10.1007/s10479-016-2393-z>.
- [16] Shamli, N., & Sathiyabhama, B. (2016). Parkinson's Brain Disease Prediction Using Big Data Analytics. *International Journal of Information Technology and Computer Science*, 8(6), 73–84. DOI: <https://doi.org/10.5815/ijitcs.2016.06.10>.
- [17] Kankanhalli, A., Hahn, J., Tan, S., & Gao, G. (2016). Big data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers*, 18(2), 233–235. DOI: <https://doi.org/10.1007/s10796-016-9641-2>.
- [18] Ng, K., Ghoting, A., Steinhubl, S. R., Stewart, W. F., Malin, B., & Sun, J. (2014). PARAMO: A PARAllel predictive Modeling platform for healthcare analytic research using electronic health records. *Journal of Biomedical Informatics*, 48, 160–170. DOI: <https://doi.org/10.1016/j.jbi.2013.12.012>.
- [19] Sun, J., Mcnaughton, C. D., Zhang, P., Perer, A., Gkoulalas-Divanis, A., Denny, J. C., ... Malin, B. A. (2014). Predicting changes in hypertension control using electronic health records from a chronic

- disease management program. *Journal of the American Medical Informatics Association*, 21(2), 337–344. DOI: <https://doi.org/10.1136/amiajnl-2013-002033>.
- [20] Big Data Analytics in Healthcare: Promise and Potential. *Health Information Science and Systems*. Retrieved 1/1/2020, from <https://link.springer.com/article/10.1186/2047-2501-2-3>.
- [21] Engineering, C. (2018). *A Machine Learning Approach for Prediction of Diseases Using Unstructured Datasets*. 6265–6270. DOI: <https://doi.org/10.15680/IJIRCCE.2018.0606024>.
- [22] Predictive Analytics Solutions in Healthcare. *Health Catalyst*. Retrieved on 25-12-2019 from <https://www.healthcatalyst.com/predictive-analytics>.
- [23] Salekin, A., & Stankovic, J. (n.d.). *Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes*. Retrieved on 25-12-2019 from <https://www.ijsr.net/archive/v4i4/SUB152923.pdf>.
- [24] Harisekaran, S. S. D. G. (2015). Big Data Analytics Predicting Risk of Readmissions of Diabetic Patients. *International Journal of Science and Research (IJSR)*, 4(4), 534–538.
- [25] Yang, Y., Luo, M., Xiao, L., Zhu, X., Wang, C., Fu, X., & Yuan, S. (2016). Exploration of pathological prediction of chronic kidney diseases by a novel theory of bi-directional probability. *Nature Publishing Group*, (May), 1–5. DOI: <https://doi.org/10.1038/srep32151>.
- [26] Farouq Halawa, Samar Al-Hihi, Wenfei Shen, Daehan Won. A Model-Based Approach of Data Analysis and Prediction in Chronic Kidney Diseases (CKD). *Proceedings of the 2017 Industrial and Systems Engineering Conferenc*. Norcross (2017), 1018-1023.
- [27] Platts-Mills, T. A. E., & Perzanowski, M. (2018). *The use of machine learning to understand the relationship between IgE to specific allergens and asthma*. *PLOS Medicine*, 15(11), e1002696, 1-3. DOI: <https://doi.org/10.1371/journal.pmed.1002696>
- [28] Allergies: Causes, Diagnosis, and Treatment. Retrieved on 02/02/2020 from <https://www.medicalnewstoday.com/articles/264419.php#diagnosis>
- [29] Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, 2, 59–77. DOI: <https://doi.org/10.1177/11769351060020030>.
- [30] Singh, Y., Srivastava, D., Chandranand, P. S., & Singh, D. S. (2018). *Algorithms for screening of Cervical Cancer: A chronological review*. Retrieved on 02/02/2020 from <https://arxiv.org/ftp/arxiv/papers/1811/1811.00849.pdf>.
- [31] Alam, T. M., Milhan, M., Khan, A., Iqbal, M. A., & Wahab, A. (2019). *Cervical Cancer Prediction through Different Screening Methods using Data Mining*. 10(2), 388–396.
- [32] On, S., Analytics, P., To, M., & Breast, D. (2016). *Analysis and survey on predictive analytics model to diagnose breast cancer*, 3(6), 19-25.
- [33] Rajbharath, R., Sankari, L. & Tholkappia Arasu, G. (2017). Predicting Breast Cancer using Novel Approach in Data Analytics. *International Journal of Engineering Research*, V6(05), 72–76. DOI: <https://doi.org/10.17577/ijertv6is050018>.
- [34] Venkata Prasad, C. & Siva Kumar, A.P. (2016). ORAL CANCER DETECTION USING DATA ANALYSIS. *International Journal of Advanced Research in Computer Engineering & Technology*, 5(11), 2577–2580.
- [35] Lavanya, L., & Chandra, J. (2019). Oral cancer analysis using machine learning techniques. *International Journal of Engineering Research and Technology*, 12(5), 596–601.
- [36] Shamli, N., & Sathiyabhama, B. (2016). Parkinson's Brain Disease Prediction

- Using Big Data Analytics. *International Journal of Information Technology and Computer Science*, 8(6), 73–84. DOI: <https://doi.org/10.5815/ijitcs.2016.06.10>.
- [37] Ray, S. (2019). A Predictive Diagnosis for Parkinson's Disease Through Machine Learning. *The Canadian Science Fair Journal*, 2(1), 5–16.
- [38] Bardhan, I., Oh, J. H., Zheng, Z., & Kirksey, K. (2015). Predictive analytics for readmission of patients with congestive heart failure. *Information Systems Research*, 26(1), 19–39. DOI: <https://doi.org/10.1287/isre.2014.0553>.
- [39] May, H. T., Lappé, D. L., Knowlton, K. U., Muhlestein, J. B., Anderson, J. L., & Horne, B. D. (2019). Prediction of Long-Term Incidence of Chronic Cardiovascular and Cardiopulmonary Diseases in Primary Care Patients for Population Health Monitoring: The Intermountain Chronic Disease Model (ICHRON). *Mayo Clinic Proceedings*, 94(7), 1221–1230. DOI: <https://doi.org/10.1016/j.mayocp.2018.06.029>.
- [40] Bilal, A., Vellido, A., & Ribas, V. (2018). Big Data Analytics for Obesity Prediction. *Frontiers in Artificial Intelligence and Applications*, 308, 141–145. DOI: <https://doi.org/10.3233/978-1-61499-918-8-141>.
- [41] Hammond, R., Athanasiadou, R., Curado, S., Aphinyanaphongs, Y., Abrams, C., Jo Messito, M., Elbel, B. (2019). Correction: Predicting childhood obesity using electronic health records and publicly available data. *PLoSone*, 14(4), 1–18. DOI: <https://doi.org/10.1371/journal.pone.0223796>.
- [42] Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe. *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*, 1–303. DOI: <https://doi.org/10.1007/978-3-319-21569-3>.
- [43] Salekin, A., & Stankovic, J. (2016). Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. *Proceedings - 2016 IEEE International Conference on Healthcare Informatics, ICHI 2016*, 262–270. DOI: <https://doi.org/10.1109/ICHI.2016.36>.
- [44] Yach, D., Hawkes, C., Linn Gould, C., & Haman, K. J. (2015). *I= SPECIAL COMMUNICATION. The Global Burden of Chronic Diseases Overcoming Impediments to Prevention and Control*. 29I(21). Retrieved on 12/02/2020 from www.iamacom.
- [45] Parikh, R. B., Kakad, M., & Bates, D. W. (2016). Integrating predictive analytics into high-value care the dawn of precision delivery. *JAMA - Journal of the American Medical Association*, 315(7), 651–652. DOI: <https://doi.org/10.1001/jama.2015.19417>.
