Journal of
Biomedical Systems & Emerging Technologies

Review Article Open Access

Emergence of Automated Computing Technologies in Biomedical Disease and Drug Discovery

Md. Golam Sarowar*

Department of Computer Science and Engineering, East West University, Bangladesh

Abstract

Present era is estimated as the era of technological evolution. Every aspect of human life is greatly influenced through it. Just similar like all the sectors, emergence of information computing technologies in Biomedical drug, disease, big data is worthy of remark. However, access, manipulation and analysis of biomedical data become significantly mandatory for medical disease, drug discovery. In this regard, computer technologies are being developed emerging new specialty, methods, prospects, approach to handle big data. Rather than this big data perspective is seems to be appeared increasingly in biomedical sector which have already a great obstacle for the practitioners to work efficiently. In this respect lots of research have been conducted still going on. It's a matter of blessings that recent computing technologies are doing well discovering new trends for disease detection as well as drug discovery. However, some barrier is still in action in this path. Realizing the worth of this computing technologies in biomedical science this work encompasses some exceptionally tremendous contribution in biomedical science, what they trended to do, their perspectives, challenges, way to avoid obstacles which can be a resource for the future further research in this field. Moreover, biomedical big data computing perspectives and contributions have been also concentrated in this work. Various machine learning approaches focused on various research and their drawbacks are also receptacle for the rest portion of this contribution.

Keywords: Genetics computing; Drug design; Big data analytics; Drug-Drug interaction; Machine learning; Gene-Gene interactions

Introduction

Impact and values of biological research are becoming dynamically more robust, mechanical and automated in data arrangement, analysis and measurement. Mining technique or algorithm is significant for retrieving meaningful information from the biological data. It is a dynamic and systematic demonstration. Since biological processing is considered the imperative part of the world computing, thus it requires more powerful methods and algorithms. Moreover, software and integrated tools are also important for this process. Machine learning is one of the key methods for handling biological datasets and very large DNA (Deoxyribonucleic acid) sequences [1-4]. Computers and other digital systems assist large biological data processing, thus discovery and development of new systems is essential with the rapid growth of large biological dataset. New research and computations are generating high volume of dataset in each and every moment. Moreover traditional approaches are unable to manage very large biological data with accurate and faster computations. Consequently, biological mining techniques which are hybrid mechanism with computer science, physics, chemistry, biology, mathematics, statistics, genetic engineering, molecular biology and biochemistry; become indispensable. Furthermore, sets of evolutionary computing algorithms can govern the large biological dataset processing [5-10]. These techniques control faster biological data processing with accuracy and perfections. Moreover, cloud computing, data sciences and bioinformatics are examples for popular new fields that deals the biological data processing.

Big data in biological processing is the common phenomenon in current industry and laboratories. Organizing and arranging information from these big dataset is a challenging issue as well as a key factor in knowledge mining. Statistical and mathematical illustrations are supportive for retrieving meaningful and hidden information. Data mining methods, approaches, processes, tools and formats are equally important for exact information generation [11-15].

In the age of wireless communications and faster digitization, very large bio centric information have been growing in exponential manner due to the rising of faster processing on microarray datasets. Moreover, DNA sequencing, RNA (Ribonucleic acid) synthesis, protein-protein interactions are also some prime factors that increase the datasets. Big data analysis techniques support these datasets to get the meaningful ideas from human and animal dataset. The growth of the data volume in the recent era is significantly huge compared to few years back. Recent growth is so rapid, which is almost twenty times more than three years back collections as reported in Figure 1. The primary assessments are done for twelve years as 2000 to 2012. In Figure 1, the X-axis shows the years and the Y-axis shows the datasets outcome for each year. All the four lines are merged due to the similarities among the datasets collections. For some initial years, there are very less changes from 2009 to 2012, but there are sudden changes from 2011 to 2012. Currently these changes are automated and reaching to the apex points for no returns [16-28].

One of the pivotal reasons behind the large biological datasets is the diversity of human, animal and plant life. Despite the variations, the biological details interrelated with the universal sets of living organs.

The chapter is organized as follows. Background study is demonstrated at section 2, followed by delineating the AD related associations in sections 3 including the gray matter functionalities of this severe disease. Cancer based computational analysis is performed

***Corresponding author:** Md. Golam Sarowar, Department of Computer Science and Engineering, East West University, Bangladesh, Tel: 09666775577; E-mail: Sojolewu6@gmail.com

Received March 02, 2018; **Accepted** March 13, 2018; **Published** March 30, 2018

Citation: Sarowar MDG (2018) Emergence of Automated Computing Technologies in Biomedical Disease and Drug Discovery. J Biomed Syst Emerg Technol 5: 117.

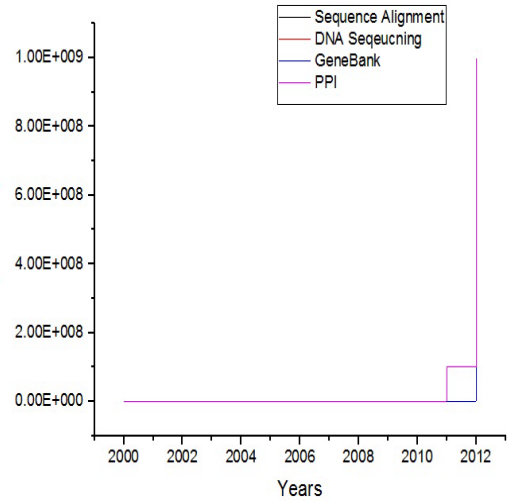

Figure 1: Various datasets growth from 2000 to 2012.

in sections 4. Big data and its impacts on health informatics are narrated in section 5. There are four subsections that also reflect the big data on health informatics applications. A afterward, a type 2 diabetes risk factor prediction is introduced in section 6. In section 7, the machine learning techniques based supervised learning analysis is addressed. Web Semantics is demonstrated in section 8, while the HIV-1 computational classifications are described in section 9. The findings of obesity under gene-gene interaction are discussed at section 10. Microbial communities and Bacterial Vaginosis are portrayed in section 11, followed by obesity analysis. Section 12 figures out the new part of Epigenetic datasets with Karhunen-Loeve transform. Finally, Section 13 introduces the Bayesian scoring system for in measuring the health related services.

Background

General mining can be groped in several ways; however most popular mining processes can be categorized as predictive-or descriptive- data mining as reported in Table 1. Complete dataset of any objects can be processed by these two mining groups. Image processing, signal processing, business data processing, DNA sequencing or protein interactions are easily manageable using one of these techniques. Predictive data mining techniques are used frequently to get faster and accurate data. Most of the predictive techniques are based on statistical processing as well as mathematical analysis. There are lots of mathematical models and techniques that are under predictive models. Irrespective of areas and subjects, predictive data mining approaches are imperative to obtain exact information from lots of datasets. Moreover, simulations and other computing are also easily adjustable by predictive mining. It enables ML approaches to learn and train datasets based on the historical data demonstrations and computing. These analysis and synthesis bridge biological mining from present to upcoming future.

There are set of predictive approaches which are frequently used in the digital era, such as the neural networks, principal component analysis, independent component analysis, particle swarm intelligence, self-organization map, regressions, support vector machine (SVM), classification and regression tree (CART), decision tree (DT), deep neural networks (DNN), discriminate analysis (DA), Bayesian Network (BN), Boosting (BT) and Random Forest (RF).

Alzheimer's Diseases Network Demonstrations with ADNI Cohorts

Alzheimer's diseases are one kind of neurological disturbance which occurs due to damages of brain cells. Though Alzheimer's diseases start lightly, it widens rapidly causing short time memory loss and cognitional degeneration. Long term process of this disease leads to dementia which is responsible for enormous damage of human brain's usual activities. Moreover, the aged community is largely afflicted by these heinous diseases, commonly saying it's a part of aging [29]. Basically, the authors tried to figure out the genetic factors which are basis reason behind the improvement of this disease using bioinformatics approach. An attempt to detect the main genetic factors which are liable for elevating and uplifting the AD was discussed. Once the main culprit can be found, it will be evolutionary steps to find cure for the corresponding diseases. For stepping forward various algorithms belong to bioinformatics can be embedded to identify the principal genetic codes behind the crucial diseases like Alzheimer. Ultimately, in this work the authors have referred genome-wide association studies (GWAS) on two Alzheimer's disease Neuroimaging Initiative (ADNI) outfit. Actually from the perspective of bioinformatics, the Alzheimer's disease Neuroimaging Initiative (ADNI) provides the practical and clinical reliable data for obtaining prevention, treatment cure for this kind of diseases. Rather than that the following specific tissue network-wide association study (Net WAS) have been applied on genome-wide association studies (GWAS) for identifying the incidental basic patterns lied on a functional network.

Various studies regarding AD have been illustrated that almost 70% damage is associated with genetics for this diseases [30]. Even one of the significant gene APOE (Apo lipoprotein E) which is the principal cholesterol server to human brain is engaged with the genes accused of spreading AD. The other similar genes to APOE are also obstacle to get cure for devastating diseases like Alzheimer [31]. More recent studies on Alzheimer's diseases have detected 22 as well as 11 high risky site on human body as main reason for late-onset Alzheimer's disease

Predictive Biological Analysis	Descriptive Biological Analysis
The primary goal is to separate the large dataset into small groups.	This process generates set of rules for controlling whole dataset.
Set of statistical analysis support to get the exact meaning of the desired items.	There are set of descriptive analysis that indirectly used the predictive mining algorithms. As for example Smith waterman and Needleman Wanch algorithms are applied to find the sequences under predictive environments.
All features associated in the training datasets are equally important for all data levels.	Group of training data clustering is used to handle the large volume of biological datasets. Popular clustering processes are frequently using to get the meaningful ideas.
Training and testing features determine the overall outcomes of the experiments.	Features are collected in a group rather being individual.
In each experiment, there must get some outcomes. If outcomes have probabilities greater than 70% it is said to be acceptable.	Experimental outcomes are determined by group results. Sometimes single results are not measured due to the excessive volumes of datasets exist in a group.

Table 1: Sources of data mining.

(LOAD) [32]. Here, from this work all the datasets are collected from online database system for identification of related components in the hippocampus tissue-specific network NetWAS detection process has been illustrated and the components belong in a few important GWAS p-values. In spite of this, the related components are used formally for classifications of AD gene federation. The NetWAS method is preferred for finding comparisons between the AD genes federation and the resultant genes after using none but GWAS.

Therefore, every network has the effectiveness to weep out the elusive aspects of complex and complicated diseases in biology and present crucial signals. Consequently, the GWAS remains ineffective for such kind of situations that's why new and improved methods are required for obtaining accurate and exact result [33]. Consequently, the NetWAS methods are improved in this work. For obtaining desired result, first of all the versatile gene-based association study (VEGAS) has been used to insert P-values from individual P-values. These P-values are used as the input elements in following stage. Using these P-values as input vector, the NetWAS reproduces features to differentiate the AD association genes more effectively. Here, the NetWAS method along with GWAS process is applied together for better performance and program efficiency than GWAS alone. The new developed method NetWAS here truly applied to invent the similarities between the gene associated data responsible for AD as well as to essence phenotype-associated network components. Basically, developed NetWAS which is one of the ML based algorithms is considered to recognize the symphony among the associated genes experimented before. Furthermore, the Network Interface Miner for Mutagenic Interactions (NIMMI) which combines protein-protein interaction data GWAS data for better and reliable performance [34] was introduced. So, a new ML approach for tissue-specific internal reaction to override the previous findings from the method GWAS was implemented. This work provides ideas regarding specific tissue features that may play prime role for determining the root cause for every devastating diseases like AD also can overcome various challenges that was quite impossible for the previous experiments. Thus, the source code and overall findings of this work can simplify the way to progress approach of various methods for best outcome and better efficiency.

Genetics Computing for Grey Matter Density in Alzheimer's Disease

The AD is one kind of forecast for the disease dementia. Basically, dementia doesn't stands for a disease; it's an overall long term process that describes various signs, symptoms associated with decline of a person's brain cells activity like memory loss, thinking skills and other day to day activities. Dementia is sometimes referred to "senility" or "senile dementia" [35]. Mostly, Alzheimer's is the outcome of aging. All over the world almost 44 million people are somehow affected by AD or related dementia. Approximately within affected people only 1 from 4 individuals are treated and cured from Alzheimer's [36]. In spite of being such devastating disease, the Alzheimer's has no exact known cure. In the current study, a full concentration on bioinformatics approach to the genetic analysis of grey matter density to result in deceased outset of the AD was provided. Various kinds of ML were carried out by assembling them together for better execution than previous works. Full concentration on gene factors movement and internal reaction behind them along with functional genomics data for entrancing biological relationship was specified. Various genetic studies found some basis gene factors which are responsible as outset of this risky disease. In the past decade, thousands of information studies which represented the genetic information regarding AD were conducted, but that information has become quite hard to follow for the researchers [37]. However, there are some more excessive genetic factors associated with AD but those can't be determined yet using usual methods of bioinformatics related to diseases gene factor recognition methods.

Considering all the undiscovered facts, this study is based on genome wide association study (GWAS) and applied on the datasets which belong to the AD Neuroimaging Initiative (ADNI) using grey matter density methods implemented process. Therefore, the authors implemented a new way to cope with polymorphisms and their regression to make an obstacle for rapidly growing AD. The main formal destination of this study is to determine the enhancement rate possibility of the memory loss which results in dementia. Therefore, if the rate is determined then better treatment for this disease can be determined which will be a great evolution in the history of medical science. The authors have also applied functional magnetic resonance imaging (fMRI) methods on approximately 818 peoples as well as 733 genetic data categories for experiment. After that both fMRI and GWAS has been embedded successfully which enhance the possibility of bringing to pass voxel-wise genome-wide association studies (vGWAS) for managing better opportunity to generate various mapping based problems. From the brain cells enormous types of phenotypes can be carried out. Consequently, here the authors preferred grey matter density approach to track out the responsible gene factors for AD. This study was mainly based on SNP-SNP interactions. Therefore, in the first stage the authors have engaged with using Quantitative Multifactor Dimensionality Reduction (QMDR) process to classify total number of genes along with SNPs, which can overcome the requirement of the first stage till execution. Basically, the QMDR helps to detect non-linear SNP-SNP interactions [38]. In the second phase, bioinformatics approach was applied on genes enrolled in the first phase to diminish the number of genes factors. Adopting more bioinformatics process the authors finally come to a specific solution.

From the last decades, several experiments have been crowned using various methods of bioinformatics related to identification of gene factors behind enormous devastating diseases. Nonetheless, all the previous works transpire only a low segment of gene-gene interaction or difference due to DNA sequence switch which is liable for rapidly increasing rate of demolishing disease like Alzheimer's. One of the major contentions behind this low invention is elaborate and complicated gene factors and constrains from the environment. Besides it's also one of the burning question for the scientists from the last decade to detect and characterize the main suspicious gene factors belong to prosperity of abundant human diseases [39]. Though one or two bioinformatics algorithms can be found hardly to deal with elaborate and sophisticated datasets, but still it's remain unbeaten both for the researchers and scientists for identifying the exact responsible gene factor for human diseases. Realizing this fact, a new way for squeezing the perplexity of the gene factors belong to AD was proposed. The findings of this work was quite noticeable than the previous work executed on AD. In spite of being better than previous works this work has one limitation which is it can easily illustrate one-way or two-way interactions but for more complex data current work is not so fruitful.

Iteratively Breast Cancer Analysis with METABRIC Dataset

Breast cancer is one of the supervise disease most of the time which occurs for the gene contains breast cancer's virus. Approximately for 5-10% possibility of being affected by this disease come directly

from the parents. The most liable genes for this heinous disease are BRCA1 and BRCA2 [40]. There are approximately 20 kinds of breast cancer is present at current era. The main diagnosis for some kinds of breast cancer can't be discovered yet but there's some cure is possible to apply to some kinds of breast cancer. From previous decades lots of researchers and scientists are performing various experiments to invent a full cure for the breast cancer. Moreover, various recent studies represents lots of microarray resolution to get a better way to identify the sub graphs for the cure of breast cancer [41]. From the perspective of clinical science, one of the heterogeneous diseases is breast cancers which obstacle for improving the diagnosis of tumors classifications clinically [42].

Gene factor evolution using microarray proves the complication of breast cancer disease. A lot of methods have been proposed by various researchers and scientists to identify the main culprit factors behind this devastating disease. Currently, multi-gene lists and single sample predictor models provide better performance to reduce the multidimensional complexity level of this disease. The incapability of some established model to deal with high dimensional data limits the opportunity of gaining desired result, however various new studies contributing a great role to compete with this mysterious disease. An evolution by creating a new iterative powerful strategy for computably biased subtypes and by enhancing class prediction while using METABRIC dataset were performed.

Typically, the traditional methods help largely for clinical decision making creating various discoveries. The PAM50 methods are used for assigning the molecular subtypes based on various gene expressions. Other various methods are also correlated to clinical diagnosis [43]. All the methods work on low dimensional datasets as well as individual sets of data. Therefore, in this work, the methods which has been implemented has a disquiet habit of using SSP models for the sake of detection of multiple gene factors responsible for improving breast cancer. Although the current proposed model automatically execute some extra computation inside it that's why learning mechanism is quite poor here. Consequently, there's a basic need to interpret the way for providing the exact predictions for clinical use. The Molecular Taxonomy of Breast Cancer International Consortium provides rich datasets for clinical research [44]. It provides some datasets for breast cancer disease but for some unavoidable reasons there's seems some disturbance on the proposed datasets. In spite of this, there seems some disturbance in subtype's level also that's why this work has been proposed to keep concentrate on such kinds of inconsistencies. The previous version of this work proposed for major review of this datasets and in this version a powerful approach has been brought forward to iteratively refinement of data along with assemble learning. So, this work is particular for clinic pathological markers and of course for the patient's life survival. The authors attempted to establish a strong and powerful approach using CM1 score to classify the cone of each group as well as effective learning techniques to strengthen the ability to determine the exact phenomenon. Besides, this work associated with an iterative computational process and statistical magnitude has been improvised here.

From bioinformatics, the ensemble learning with various algorithms results in a great extent. Actually the main advantage of ensemble learning is it can easily comprehend decreased over fitting as well as improvise performance of classification. There's a lots of ensemble approaches among which select-bagging and select-boosting are the main approaches to work efficiently and in a faster way. Although, the iterative approach was used alone or along with CM1 score then the outcome was quite disappointed, whereas the iterative approach with combination of an ensemble learning mechanism provides faster and efficient performance than others [45]. Besides, practically this work's improvised methods titled iterative method with CM1 score and ensemble learning approach represents a great effectively for foretelling more accurate and exact sample subtypes in the METABRIC breast cancer dataset. Consequently, from the outcome of comparing this method to traditional process and finding the new supervised approached which has been implemented in this work, a great opportunity for finding the best performance and accurate gene expression which will be evolution for medical science to obtain one way to compete with the breast cancer disease.

Integration for Various Cancer Genome Predictions

Cancer is the phenomenon of unusual growth of human cells along with capability of being aggressive for other cells creating lumps or masses just like tumors though all the tumors are not capable of spreading everywhere, such as the benign tumors which are not spreadable to the other parts of the body. There's approximately 100 types of tumors are present in various human cells. Each and every cancer disease is identified by the human cell which is primarily affected [46-48]. Meanwhile, present days for cancer all the improvised versions of modern technologies next-generation sequencing and microarrays already have disclosed enormous number of genomic features like DNA copy number alterations (CNA), mRNA expression (EXPR), microRNA expression (MIRNA), and DNA somatic mutations (MUT). Therefore, lots of exploration for a particular type of this genomic data produces various types of prediction biomarkers in cancer. Various predictive biomarkers have mentioned for research basis on various number of biological components simply like genomic, proteomic, metabolomics, pathological, imaging and psychological features from where the genomic biological features have been used in a great extent although here National Cancer Institute and the National Human Genome Research Institute plays the main role [48]. In this work, the authors have implemented a very new approach to discriminate the prediction power of models proposed for various complex cancer data types, genomic data types and algorithms. For execution of the experiment the authors illustrate three different algorithms form EXPR, MIRNA, CNA, MUT this types of data and keep processing through integration for bounding the data together as close as possible. Further, the Constraint particle swarm optimization contribution as well as various components of biological connection which contains protein-protein interaction information has been illustrated.

For implementation purpose of this work, the authors have indicated their full concentration on a variable feature selection algorithm which is known as Cox proportional hazard model. In addition, the constraint particle swarm optimization process was also interpreted based on biological behavior. Completing various iterations using bootstrap beta coefficients have been detected going through the log-likelihood (NFS and CPSO) or a penalized maximum likelihood function. The used CPSO was basically associated with biological behavior of flock. Therefore, these particles belong to the swarm of particles representing the positions and velocity. This CPSO randomly set the positions vectors and velocity together. The velocities and positions are updated automatically updated analyzing their performance and types of elements. The authors have used 4 types of datasets which are online based associated with almost 100 types of elements including EXPR, MIRNA, CNA, and MUT in the TCGA assortment while accession. The CPSO was used to evaluate basis components of the sophisticated dataset using its ability to produce various user defined elements that is used as the basic survival model.

The network feature selection model is also used to detect protein-protein interaction network for evaluation of feature selection. Therefore, it is also used for producing poly-cancer biomarkers. This NFS established best performance while the datasets were more complicated and sophisticated to deal with. It evaluated any kinds of data and represented the desired result [49] and also successfully explored the complicated data with noticeable success rate. They have also referred least absolute shrinkage and selection operator (LASSO) to select the elements while the number of features is greater than the number of datasets. This algorithm performed the best efficient process of finding the basic features for detecting cancer disease. Using similar algorithms all the models were evaluated and were compared using the concordance index (c-index) for obtaining the better performance.

This work differentiated the predictive level of various features genomic for characterize genomic related data. The integration process of enormous genomic datasets produced higher class models of datasets which are more powerful than that from single survival data simply like mRNA. From the source of genomic data, the mRNA gene reproduced stronger highly preserved models for integration of cancer genes data sets. In addition, the authors have selected the models that enabled to bind all the four types of gene expression. According to the authors' prediction, the CAN and miRNA data pursue the mRNA data for better ability to perform during alternative data seems to be poorly calculated. The main risk of calculation of various types of data hold back and the basic level of agreement is seemed to be associated with censoring data.

Big Data Computing

Big data computing is an overgrowing technique for mining multidimensional data from scientific discovery along with various large-scale structures. The data which is collected from various scientific experiments or movements require efficient data management, analyze, providing, manipulating to reach in a goal [50]. The recent IDC report predicted that there could be increment in the big data fields 20 times more than current situation [51]. Recent development in semiconductor and cloud computing technologies enlarge the way to manipulate big data, store necessary information, analyzing data, preserve significant components etc. in a faster and quite cheaper than ever. All the preserved data can be accessed in an affordable cost, which was quite impossible even a decade ago. Therefore, big data organizes the valuable data from rapidly growing datasets, lots of variations, large volumes, frequently changing multiple datasets in a constant time.

The big data analysis system is designed in such a way that it can identify any meaningful data from a vast crowd of data. Big data technologies are currently gaining the opportunities in medical science, bioinformatics, health informatics, computer science, management system and lots of fields. Various recent articles have been published regarding survey or statistical; information of benefit of big data computing like mobile devices, tablet computers, internet of things, cloud computing etc. [52,53]. Current demand of big data computing widens the spectrum of big data analysis in various domains. The most significant fields are scientific exploration, health care, governance, financial and business management analysis, web analytics, internet of things along with mobile health informatics, and bioinformatics. Compared to big data computing, service-oriented technologies such as cloud computing is also capable of storing data, analyzing and manipulating large size of data but sometimes it becomes challenging working with cloud computing, that time big data computing is must.

In [54], the authors kept their attention for particular examples of big data computing experimenting practically by lots of researchers and scientists. As bug data computing is has huge advantage at present time that's why the authors described the importance of bug data computing and the basic concept of it. Traditionally, big data system refers to row column wise distribution of various data but when the size overflow the capacity of such containers then it is indicated as big data architecture. Overall architecture of big data is demonstrated in Figure 2. Figure 2 illustrated the big data computing along with closer innovation or discussion on architecture, technologies, tools, mobile technologies, health technologies and many more paradigms working behind big data computing.

Big data estimation with next generation in drug design

From the past decades, big data technologies have been used widely medical science and health-care informatics research. A huge numbers of data sets have been gathered and generated using various bioinformatics approach for the sake of research in the fields of medical science [55]. For clearance with the development of clinical technologies and improvised version of medical science it's quite easy now for determination of exact gene or DNA sequence in a couple of seconds evaluating millions of data traditionally, which was quite impossible for the researchers and scientists. Typically, the outgoing features of bioinformatics and health science complex, complicated and huge size of data flourish the researchers to engage with the research work of big data analysis in bioinformatics. It also involves big data analysis for latest achievement of medical technologies, computational intelligence, data mining, machine learning, ensemble learning etc. [56]. Consequently the fields of bioinformatics ensures the improvement of algorithms, databases, computational data mining, and machine learning nowadays excavation of big data problem in bioinformatics is a faster engaging process in the biomedical research. These kinds of verities of source in biomedical data as well as applying next-generation informatics algorithms on it can help to evaluate a new era of precision medicine. Modern informatics or next-generation informatics approaches provide a great occasion for sharing their findings interpreted from their latest investigations using enormous types of informatics approaches. Besides, the amount of data in medical science is exploring in a great speed. In order to compete with this increasing amount of dataset, algorithms of informatics science to explore the hidden discoveries from them were performed. In spite of all this there's a lots of obstacle with big data among which store, search, analysis, processing, sharing, viewing, discovering knowledge from those data though exploring knowledge from these type of big data has become burning question for the scientists and researchers of last decades [57].

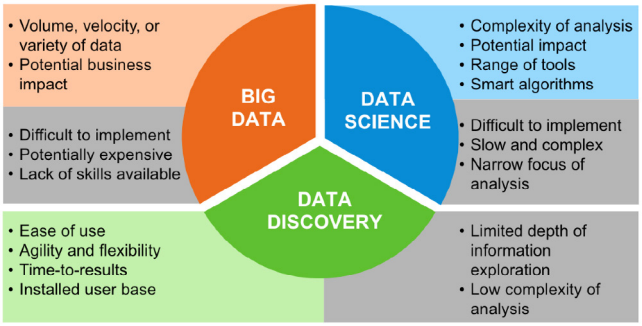

Figure 2: Overview architecture of big data computing.

Traditionally there stay various types of sources for genomics and proteomics information. Each and every source has own styles, mythology though most of source use ontology like genome ontology. Precision medicine asserts the entire requirement needed to acquire the best clinical outcome. Therefore, precision of medicine data refers to analyze, interpret and integrate the increasing number of unequal sources [58]. Sometimes it becomes crucial for the researchers and biologists to integrate or to analyze the data from various sources along with improvised organizations or geographies for simplify the development of progressed therapies for better treatments outcome. The benefit of machine learning, supervised learning, and data mining was used to replace the current traditional use of various algorithms. Informatics approaches for analyzing, integrating, detaching the medicine legibility for better performance of current medical science were introduced in [58]. The authors differentiated between the traditional strategies and next generation informatics approaches to evaluate data of medical science for making a pathway to cope with any kind of diseases and explore new drugs analyzing the output produced using various modern informatics approaches. Consequently, for going towards the goal of precise medicine and healthcare in current system needs to accomplish the bioinformatics approaches along with development of the capability of storing accurate information regarding genomic data. Additionally, more emphasizes are needed to characterize genetic factors which are badly need to be accessed for development of health outcomes [59].

Big data in health informatics

Health informatics is associated with health care technologies to develop health care system more reliable and take action in constant possible time. Its basic elements are basically engaged with information science, computer science, social science, behavioral science, management science, and others [60]. Actually the health informatics is involved with the resources, devices, storages, backups with other computers, clinical guidelines, and information/communication system [61,62]. At present around the world the number of population is increasing at a great rate, with this increasing rate of population the health informatics data for each and every person is increasing rapidly creating a vast amount of data to manipulate. For gaining knowledge from this vast amount of data the health informatics badly need a potential limitless process to cope with this kind of data. The most challenging consequence working with vast data is investigating this data in a reliable manner. The main use of health informatics is to develop various bioinformatics process to illustrate vast medical data in an easy way. In addition, the health informatics cooperates with the population data which is executed in a particular subfield [63]. The health informatics is now a best way to the research field for working with big data and manipulate them. A big velocity for big data is happen when new update of current population's health care system is providing to data center in a great speed by means of health sensor or mobile devices [64]. Figure 3 illustrated the big data analysis in a particular way describing the subfields of health informatics showing the overall described [64].

various efficient process along with powerful hardware make the heath informatics capable of analyzing big data in shortest possible times were introduced for the capability of increasing computational algorithms. Therefore, the basis goal of health informatics providing answer of any patient's frequently asked question regarding health issue in a constant possible time via mobile devices, internet services, tablet computers etc.

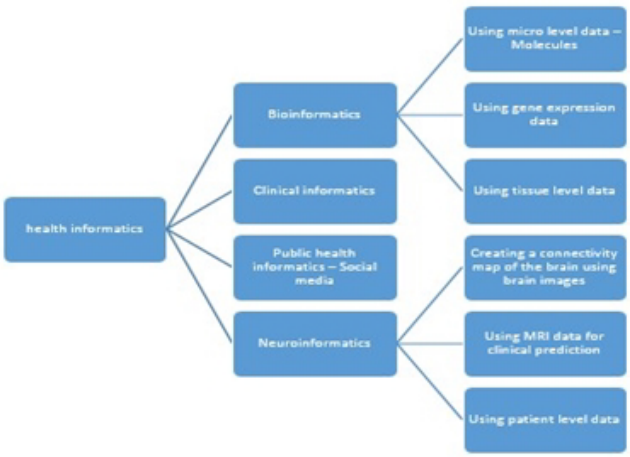

Figure 3: Overview of health informatics in big data analysis.

Big data analytics for mobile health monitors

Mobile health monitors refers the use of improved modern technologies like mobile device, computers for monitoring health issue and providing alarm for risky situations [65]. Presently, the mobile health has risen as sub-segment of eHealth, use of information technologies like mobile phones, tablets, communications technologies, and patients monitoring [66]. Mobile health technologies stands for the device collecting dynamic patients health information and immediately inform the condition to the practitioners, researchers, doctors, advisors to take particular steps in every steps. It also monitors the symptoms, signs of various diseases via mobile phone technologies [67]. Nowadays, healthcare systems are quite crucial for the vast amount of population. The overall health system of population is quite low. Consequently, an improved healthcare system along with better health care management is must for the government to ensure human right that's why it's require better health management [68]. By the means of latest modern technologies, necessary components needed for establishing dynamic health care monitoring system that considered as mobile health monitors are essential. All the modern technologies are considered the basis foundation for this type of monitoring system. National Surveys of Population Health (NSPH) helps all the way for establishing such kinds of expensive work. Therefore, a significant correlation between mobile data services and computerized system can lead health system more reliable in the current situation. Nonetheless, in every country there's a lot of population for which we need a vast amount of complex data or big data manipulation. Sometimes this issue becomes challenging to tackle the big data problems. The spreading of mobile health technologies spectrum is shown in Figure 4 [69].

Big data application in biomedical research and healthcare

Big data is a most challenging task for the active researchers and it indicates investigating, analyzing, storing data, visualizing in lower dimension of higher data. It generally extracts desired value from data [70]. Usually big data represents the data sets beyond the ability of general tools and software to deal with for getting desired result [71]. For this type of context new approaches are required to find discoveries from a huge set of data. Therefore, big data requires a set of new techniques with new forms of integration to form better performance [72]. Recently, parallel computing which was proposed by Google has created a wide vision of working with big data. Cloud computing is also

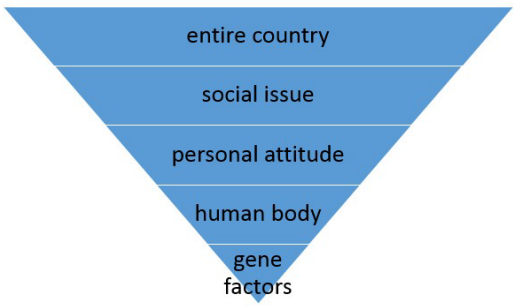

Figure 4: The overall mobile health technologies process.

an approach of dealing with biological big data because of its ability to develop system ability, and to speed up the system. Cloud computing also diminishes system requirements. Biomedical research indicates analyzing the biological data in molecular or macromolecular stages. Next generation sequencing technologies and big data techniques in bioinformatics assist to access into biomedical research data sets. The hadoop and mapreduce technologies are playing a great role recent years for dealing with the vast amount of data. Big data has great significant efforts in clinical informatics. Clinical informatics refers to health care technologies in medical science along with all kinds of health related issues. Overall there's a lot of use of big data in various fields like clinical science, biomedical research, imaging informatics etc.

Type 2 Diabetes Computations of Risk Factors Predictions

Form the last decade with the improvement of modern science the dynamic health record has become available everywhere in the world and for symptoms of any risky disease constant action can be easily estimated in recent time. The availability of type 2 diabetes causes an enormous harm to human. A recent study predict that all over the world in the year 2000 the total number of patients affected in diabetes was approximately 171 million and it will be almost 366 million within 2030 [73]. Recent statistics on type 2 diabetes predict that the major patients are affected by type 2 diabetes [74]. All the recent research regarding diabetes 2 illustrate that only regular physical exercise and weight loss can be the reduction process of diabetes. All the other research also confirmed this [75-77]. The diabetes patients are being selected for any kind of experimental works depending on their obesity and glucose level. The past popular known models for better risk prediction models of diabetes are ARIC [78], San-Antonio [79], AUSDRISK [80], and FINDRISC [81]. Although the mentioned models are popular but the main obstacle or constraint is time consuming as well as more costly. That's why with the improvement of modern technologies new approaches have become must to be implemented anyway.

By the sake of various ML algorithms, all the liable gene factors or symptoms as well as risk factors can be determined. Clinical and practically observed data actually helps to accomplish precautionary steps for type 2 diabetes.

A Supervised Learning Process to Validate Online Disease Reports

All the bioinformatics algorithms that determine the main risk gene factors or gene variations behind each and every disease of medical science must require a large number of data to manipulate. For this situation, the pathogen distribution model is widely used for its high predictive ability of any factors. It also demonstrates largely to create diseases maps depending on each disease's variation and liable gene factors. Production of data for these methods comes from online health based data system such as Genbank. One of the major problem of online data base is all the data may not be valid and there is no such method for validation of dynamically providing online based data. Depending on environmental and socio-economic condition the occurrence possibility of each disease in a particular position is defined by the location of disease occurrence.

Targeted learning in healthcare research

From the beginning of availability of big data electronic health care technologies and claims data sets are arising in a great extent for answering the drug safety measures all over the world. The ever-growing rate of these current technologies towards investigational data sets such as genomic information, laboratory results, and radiologic images [82]. These big data sets in health care technologies are increasing day by day and the questions are widening. Compared to this trend the new approaches are not improved enough to get rid of this trend. The old parametric modeling approaches are inadequate for analyzing the coefficients of the big data source. Therefore, the illustration of these coefficients is largely depends on various covariant. All the traditional approaches feel obstacle for the higher dimensionality of big data. They can't convert this higher dimension to its lower one [83]. Realizing this context, big data problems can be solved using a new approach namely targeted learning (TL).

Targeted learning (TL) is a current approach for dealing with big data problems which is implemented using semi parametric approach along with supervised machine learning mechanisms. This approach targets on higher dimensional data and helps previewing lower dimension of higher dimensional data. The specific focus of targeted learning algorithms is to minimize the bias of any targeted parameter and make it simple to reach in a discovery point. Basically the targeted learning (TL) algorithm is the combination of two bioinformatics approaches which are Super learning (SL) and targeted minimum loss- based estimation (TMLE). Actually the TMLE is applied on data which is analyzed by Super learning (SL) before. 1st the big clinical data is manufactured by SL, after that TMLE is used. Because of this combination the targeted learning (TL) approach outputs sound findings analyzing big health data. Nowadays the TL is basically applied in a wide range of spheres like genomics, precision medicine, health policy, and drug safety. Therefore this paper illustrate the significant contribution because of the combination of two bioinformatics approach named Super learning (SL) and targeted minimum loss-based estimation (TMLE).

Lumping versus Splitting for Mining in Precision Medicine

In recent years, the rise of data severe biology, advancement in molecular biology and technological biology along with the way the health care is delivered to the patient of current world paves the biologists working on these new diseases to find a particular cure using precision medicine models [84]. Medical science is utilizing the advantage of modern improvised data mining methodologies in various aspects of detection of liable DNA codes or gene factors for better outcome of devastating diseases which are invented recently. It's one of the most challenges for medical science finding prevention or cure from such kinds of recent invented diseases and most of the time it's remain unbeaten. Consequently, thousands of people die because of not finding the exact gene factors or DNA codes working behind the disease. With the renovation of bioinformatics, a new movement has

been seen in medical science because of getting the related genes factor corresponding to each and every disease [85]. Therefore, nowhere the data mining is needed for precision medicine. The capability of representing each and every disease's risk for treatment purpose is one of the correlating factors for precision medicine. This is achieving a great measure from the last era for ongoing technology improvement. A lot of current powerful approaches rely on unilabiate and linear evaluations that can be easily avoid the structure of complicated criterion [86]. One of the successful example for the precision medicine model is to drug development involving the drug Crizotinib, a vanquisher for the MET and ALK kinases which started the practical improvement with a widen number of population in a great extent.

Conceptually, the fact regarding precision medicine, prevention and treatment techniques creates isolated differences into various contexts. Traditionally, blood pumping has been widely supervised for blood transfusions from more than a century. However, various medical technologies improvements, broaden use of bioinformatics, invention of various powerful methodologies in information technologies along with development of large-scale datasets, and strong background for invention of methods for evaluating patients make a new era. Nowadays, it's badly needed for the medical technologies broaden the uses of precision medicine for guiding clinical experiment [87]. For achieving the goal of broaden, the use of precision medicine in clinical experiment some steps need to be initialized. One of the key features is using data mining techniques for all experiment purpose of precision medicine. From the vision of complicated data sets along with strong loci here preferable methods may be unable for performing the accurate operation on the datasets. Several studies tried to prove that using accurate types of data mining and making working portion for each disease will be fruitful for the scientists and researchers determining subtype graphs in low cost along with less time consumption. With the determination of subtypes for various diseases can lead to a great opportunity to keep compete with the increasing rate and approaches of development of diseases along with accurate liable factors of that particular disease and also if this happens then it will be start of a new era from the perspective of medical science. Rather than that through using enough specific small groups and perfect small types of data it is possible to improve precision momentously for precision medicine models. Thus, the biological data mining process has a great effect in biological large data to play a sensitive role in finding responsible culprit accurately. It also can be used for many clinical and practical contexts to ensure better treatment for the patients and also for exploring reason behind recent invented diseases along with devastating diseases like diabetes or cancer.

Mining Drug-Drug Interactions on Semantic Web Technology

Almost from the last decade drug-drug attraction gets most priority for showing its tremendous effect on patients. It got the concentration of medical scientists for various dangerous symptoms. Approximately 770,000 people all over the world died or seriously get injured because of these adverse drug events [88] in every year. In this modern era for an easy simple condition a patient is prescribed lots of drugs clinically long term process of which leads the patient to a dangerous way [89]. For prevention of various recent discovered diseases a large number of drugs are explored in every year and new unfavorable symptoms are seen to be harmful for the patients compared to the benefit of that drug [90]. It's almost out of control for the physicians to keep trace of unfavorable consequence of drug-drug reaction in medical science. Unfortunately, drug-drug interaction is badly effect patients. For sure drug-drug interaction is a major providing factor for some kinds of unfavorable and unfortunate converse drug effects. In some practical reason the genetic expression tests for classifying the risk gene factors for the patients is potentially devastating for prescribing to some extent. It seems to be in a large extent especially to the developed countries with approximate 20-40% [91]. Mainly, unfavorable adverse drug events are the consequence of provocation of various drug-drug addictions [92]. A sufficient knowledge should be mandatory for prescribing or taking the medicine for both the clinical association and patients. The new improved process regarding investigation of various diseases risk factor for patients should be broadening to decline the death rate for such diseases. To improve genetic tests for finding the suspicion of various diseases DDI-induced ADEs is required to diminish the risk of prescribing harmful medication [93]. Providing information investigating the DDIs data is full of challenge for medical science but using improved informatics approach it is quite easy to beat the challenge.

TaxKB: A knowledge base for new Taxon-related drug discovery

Last decade was the evolutionary era for medical science. Lots of clinical approach and widens of informatics approaches are the main reason behind the uplift of medical technologies. Although enormous number of diseases weep out with creating devastating situation but the modern medical technologies anyhow created a way to cope with such kinds of diseases for finding a cure [93]. Vast number of new drugs has been invented using various bioinformatics still processing. One of the main significant discoveries is taxon related drugs. Paclitaxel and docetaxel are considered the best performance giving anticancer taxon related drugs though they have lots of bad side effects [94]. So, basically Texans are most frequently working natural substance which comes from a potential group of chemotherapeutic drugs which has the ability to make an obstacle to the way of improved cancer disease. The clinical statistics, singular machinery to protest cancer gene for diminishing growth rate makes the Texans related drugs differ from the other related drugs. Besides it is considered most hopeful treatment of cancer disease over worldwide [95]. Realizing all the facts from the last decade's additional research is going on taxon-related drug discovery for protest deleterious diseases like cancer as well. Moreover, some promising findings from various taxons related works paves the researchers, scientists, drug manufacturers to contribute to this field for their future prosperity and development in the field of cancer. There are several reasons behind the taxon in cancer drug discovery. The discovery is like a breakdown in the progress of all types of cancer diseases still now which is used most widely to be only cure for cancer worldwide being fully natural-based substance [96]. Furthermore, the improvement of novel taxon variants as well as broadening anticancer activities will minimize both uplift cost and treatment cost of cancer disease worldwide forever. The cancer disease also can be indicated as deadly disease which can only be cured improving the emphasis on invention of various taxon-related drugs.

Actually there's a lot of cancer including breast, neck, cervical cancer, gallbladder cancer, brain cancer [97] and many more but it is the matter of regret that there's only one fruitful treatment produced from taxon family. Furthermore, Texans are medically associated with hydrocarbons along with four types of isoprene generated by Yew tree. Thus, the basic structure of taxon contains tetracyclic core frame called baccatin 3 along with four specific rings called ring A, ring B, ring C and ring D shown in Figure 5. Several studies have been applied on finding accurate structure of taxon so that investigating the structure new treatment for cancer can be easily found. In [98], the authors

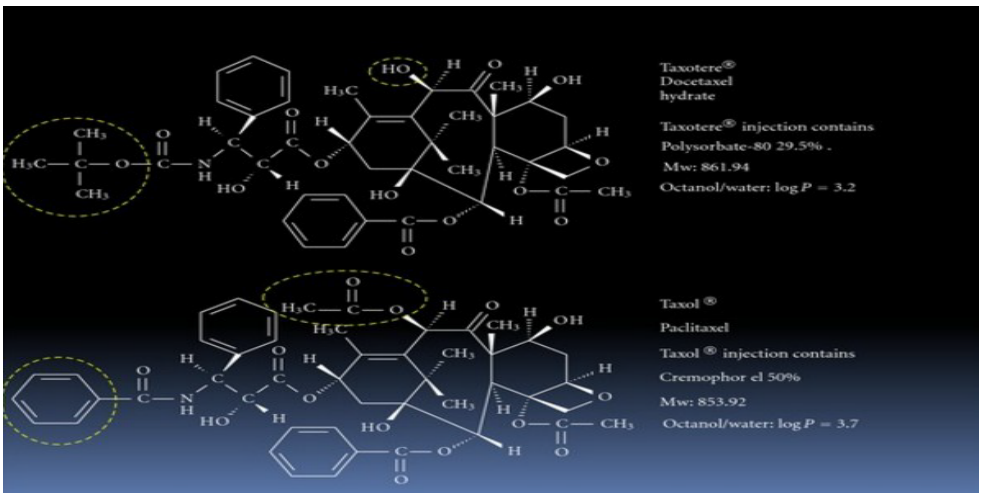

Figure 5: Basic structures of taxane.

have used TaxKB which is an online based multi-processing analyzed datasets stored for the benefit of various research work associated with taxon. The statistics indicate that obtaining cure is costly, time consuming almost take greater than ten years to execute full process accurately. For diminishing the time as well as developing cost the TaxKB is incomparable. TaxKB is a monochromic information server for inventing natural drugs components ever. From the last decade it is the only recourse to work faster processing of discovering something new.

HIV-1 Protease and Reverse Transcriptase for Multi-label Classification

The HIV (human immunodeficiency virus) is one of the deadly disease from the last three or four decades. There's no fruitful cure for this disease. The researchers and medical science can be considered failed regarding prevention of HIV. Antiretroviral therapy is only way to manipulate the patients for a limited time but this is not any long term cure for this heinous disease. The high mutation rate of HIV virus is the main obstacle for medical science to create any cure for the disease. Though many experiments have been proposed but none is enough effective till now. The HIV has several variants and thereby for prevention only single drug is insufficient because the variants are capable of ignoring against action of the drug. HIV affects various human cells including helper T cells (specifically CD4+ T cells), macrophages, and dendritic cells [99]. The prevention process can be improved either scanning the responsible mutants or scanning the resistive capabilities of resistive drugs [100]. Several expert groups have been working on both genotypic and phenotypic consequence of genes for HIV although genotypic is faster and cheaper than phenotypic.

In [101], the authors have concentrated on protease and reverse transcriptase cross-resistance information for improved drug resistance prediction by means of multi-label classification. Therefore, 2 protease sequence and 715 reverse transcriptase sequences along with specific genotypic and phenotypic data have been manipulated using various machine learning techniques including binary relevance classifiers, classifier chains, and ensembles of classifier chains. In this study, for acquiring best performance and efficiency the author's intention goes to multi-level classification models along with cross-counteraction intelligence to portend the two of the best resistive drug classes used for antiretroviral therapy for the disease HIV-1. The two basic drugs are named as protease inhibitors (PIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). Completing overall process the authors have successfully achieve a stage that is quite predicted to be accurate compared to other investigation till today.

For the current situation analyzing the genetic basis of HIV-1 is quite must for medical science for improvement of antiretroviral drugs for competing against HIV-1 virus. This supervised process was affected negatively for the mutation problem related to cross-resistance information with each and every antiretroviral drugs class. Approximately 20 types of antiretroviral drugs have been approved for HIV-1 treatment but none of this is beneficial enough to keep track on HIV-1 genes expression [101]. The authors conclude their overall described discussion with a positive feedback on whether or not the cross-multiplication information can be used for the better performance of various drug resistance prediction using machine learning algorithms and they have practically represents that multi-level classification process can be fruitful enough for improving accurate prediction for various resistive drug samples of HIV-1 disease.

Characterizing Gene-Gene Interactions in a Genes for Obesity

Recently, obesity becomes common problem for almost everyone both in developed and developing country as well. It's spreading at a great rate. Most of the recent research predicts that obesity and over weight is the main reason for diabetes, cancer and premature death. The high appearance rate of overweight and obesity will pave the people's health to a great challenge [102]. Therefore, the appearance of overweight is informed from everywhere in the world. It's a big issue for every country now [103]. Recent studies have proven the necessity of epistasis or gene-gene interactions for explaining the harmful factor of overweight and obesity. Various network-based models play a great role to explain the basis reason. It is also challenging for the researchers to analyze the data pairwise here the network based algorithms play a great tribute as it has the ability continuing pairwise operation. Recent years the medical science moves their concentration on gene-gene interaction along with SNPs for obesity. There are 12 robust gene working with obesity which are BDNF, ETV5, FAIM2, FTO, GNPDA2, KCTD15, MC4R, MTCH2, NEGR1, SEC16B, SH2B1, and TMEM18.

In [103], the authors have implemented a network based approach which is called statistical epistasis network (SEN) which is expert to classify the SNP-SNP interaction in each and every gene associated with obesity or overweight. The interactions which exceed a specific terminate point build the SEN network. Genes never react in isolation but instead as part of complex regulatory networks. All the SNPs used here come from the gene. The authors detected 709 SEN of SNPs along with 1241 SNP-SNP interactions. Binding the entire obtained sample a specific dyadic gene has been determined. One of an in cRNA SNP also has been detected here after executing SEN using multiple network measures. So basically this study represents a traditional properties for identification of genetic interaction from genome based arrays and also invention of various nodes of biological substances to diminish the obesity problem specifying the accurate gene responsible for overweight or obesity problems. As the obesity problem is main obstacle for strengthen the living period of life so the medical scientists, researchers, followers have given their full concentration on this type of research to prevent the patients from deadly diseases.

Microbial Communities and Bacterial Vaginos is under Machine Learning

Advancement in bioinformatics and informatics sequencing technologies paves clinical science to come forward to studying microbial communities and bacterial vaginosis. Microbial communities are basically a collection of hundreds or thousands of various complicated types of bacteria. Generally, researchers preferred the 16S rRNA for bacteria identification purpose in a microbial community. Microbiomes are associated with human's health and disease related sophisticated work. For example the gut microbiomes help to digest food and works for obesity in a human's body [104]. On the contrary bacterial vaginosis which is usually referred as the vagina microbiomes are consists of various types of biological bacteria although all of them a few are quite dangerous for human body. The execution process for vagina microbiome is differing from men to women [106,105]. Microbe types of bacteria are the main culprit for being bacterial vaginosis disease and almost over the world the women are highly affected by this heinous disease [104].

The precast signs of this disease are discharge, irritation, odor and so on. The disease is also linked with preterm birth and rising rate of different types of suspicious STDs [105]. For completion of bacterial vaginosis lots of bacteria works together, it's not possible for only one bacterium. BV is most dangerous while pre-birth period of children. In [105], the similarities, dissimilarities, affection, abandoned, contradictions everything has been described using two important bioinformatics approaches. The authors have focused on relationship between bacterial viagnosis and microbial communities. For microbial community, the subsets of every community are justified with various bioinformatics algorithms. For classification and identification of relationship between bacterial vaginosis and microbial communities the logistic regression (LR) and random forests (RF) have been used. For some incident the LR and RF can't determine the discovery from complex typed datasets for those types of data the authors realize and implemented some new features for simplicity of identification. Most of the cases the RF and LR methods conclude with high accuracy if five or six features are added where the other methods are unable to maintain high accuracy while complex datasets. It is also indicated here that how the findings or relationship indicators for bacterial vaginosis and microbial community changes for adding new features of LR and RF. There's a strong connection between the bacterial vaginosis and microbial community. Besides the microbial community may have internal error as well as noise with it. The authors successfully represent the relationships between BV and MC by adding some features to machine learning methods of bioinformatics. In the initial stage, it is not fully clear this classifier models are accurately find any differences between the proposed correlations between BV (bacterial vaginosis) and MC (microbial community). However, the ML determines great accuracy finding similar patterns, therefore here the ML methods can be used to detect the similar patterns of bacterial vaginosis and microbial community. Obtaining percentage of similar and dissimilar patterns the correlations can be easily substituted. Besides it also indicate the differences portion is major or minor. Consequently, the relationships between the bacterial vaginosis and microbial community can be measured by manipulating the data using logistic regression (LR) and random forests (RF).

Epigenetic datasets using the Karhunen-Loeve transform

The main concept regarding gene expression or various biological processes simply like cellular division and differentiation are at all disturbed by epigenetic regulations regarding gene expression [106]. There are a lot of variations in epigenomes datasets than genomes data. From last decade it has been challenging for the researchers and scientists to generate ideas working behind the epigenomic datasets using general informatics approaches because of the poor performance of general approaches. In recent time, genome expression wide model of epigenetic information have brought a new significant way to explore uncovered information regarding tissue and cell. Some genome expression wide models of epigenetic datasets are histone modification, DNA methylation and open chromatin [107]. To address these challenges, various particular approaches are improvised by the researchers. Therefore, for hundreds of these types of models the opportunity paves the way to be suitable to uncover the liability from quite complex way. The computational epigenetics actually refers to solving or manipulating the epigenetically data by means of bioinformatics. Due to the recent blast in the area of epigenetic data now computational methods are must to explore the necessity information investigating these epigenetic datasets [108].

Till now 100 distinct epigenetics datasets have been discovered whereas many more remains uncovered yet [108]. That's why this paper is on uncovering remained datasets by means of Karhunen-Loeve transform method. The basic phenomenon for the contribution of the work in [108] is to invent the discoveries which are undiscovered yet. For obtaining the goal the authors here preferred using Karhunen-Loeve transform method. The authors implement an approach that works as functional principal component, one of the finite karhunen-Loeve transform. Afterword by means of these methods the positive correlation between H3K4me3 and H3K36me3 as well as H3K9ac and H3K36me3 are detected. Then, highly negative relations between H2A. Z, H3K4me3, and many more are determined. Now, demonstrating all the positive and negative components the complete process is executed. Therefore, the authors have represented a complete method for the explosion of various novel patterns of epigenetic datasets by the help of functional component analysis.

A Bayesian Approach to Learning Scoring System

In today's world the public expense in health is one of the major challenging factors for the government to compete with it. Therefore, in developed and developing countries, the health ministry moves their attention to monitories their health systems as well as improve it. Some countries thought to be create a new health system. Approximately between the time 1993 and 1997 two reports on health

system effectiveness of 191 countries was published by World Health Organization (WHO) [103-108]. That time they used a general model to represent the statistical view of health system. The present artificial intelligence and computer science as well as bioinformatics play an excessive role to cope with this kind of situation. Scoring system basically represents a simple numbering system like one digit representation of various kinds of diseases. For example on a medical scoring system we can illustrate cancer with 1 point, diabetes with 2 points, pulmonary disease for 1 point, bacterial vaginosis 2 points and we assume 3 points for clinical signs and symptoms of deep vein thrombosis (DVT). These points are not experimentally or accurately distributed. By the discussion of all the senior doctors or medical science's scientists or researchers this points generally assign for simply interpreting the risk factor of a particular disease. Most of the time scoring system is a linear system combination of the digits 1, 2 or 5.

In [108], the Bayesian approach learnt the scoring system in medical science. There are a lot of medical scoring systems that have already proceeded successfully. Some of the popular scoring systems are $ABCD^2$ score, ACR score for rheumatoid arthritis, Aldrete's scoring system, Alvarado score, CHA2DS2–VASc score, Child-Pugh score, Cormack-Lehane classification system, Eagle score, Framingham Risk Score, Geneva score, and Glasgow-Blatchford score [108]. In this article the authors contribute this paper with implementation of Bayesian scoring system. All the digits of this process are small significant numbers like 1, 2 or 5 and all the coefficients must be nonzero.

Conclusion

Health informatics ensures faster and accurate medical data and symptoms processing. There are several analyses for the support of mining desired information from large volume of raw data. Drug design, big data analyses, diabetes factors predictions, cancer gene analyses, machine learning based scoring system, semantic web synthesis, Epigenetic internal functionalities, type -1 HIV intersections, computational obesity simulations and Microbial Communities and Bacterial picture are the some areas have been sketched here. Each and every area is the key filed that control the better life of human being. Computational mining and simulations help to get new them and dimensions in these sectors. Accurate measurements are vital for better health. Now-a-days, robots are frequently used to complete the critical operations of the human body. Moreover lots of devices are using for controlling exact amount of chemical for drug design, disease identifications, pathological interactions for HIV monitoring and organic chemical reactions. So, large data mining approaches for scientific measurement are essential for ever.

References

1. Aenenhaus A, Philippe C, Guillemot V, Kim CA, Vincent F (2014) Variable selection for generalized canonical correlation analysis. Biostatistics 15: 569-83.
2. Alon (2003) Biological networks the tinkerer as an engineer. Science 301: 1866-1867.
3. Fatema AS, Pabel M, Vaquerizas MJ, Jaoquin D (2005) Babelomics A suite of web tools for functional annotation and analysis of groups of genes in high throughput experiments. Nucleic Acids Res 33(Web Server issue): W460-464.
4. Stojadinovic A, Bilchik A, David S (2013) Clinical decision support and individualized prediction of survival in colon cancer Bayesian belief network model. Ann Surgery Oncolology 20: 161-174.
5. Yoel A, James BP, Sune L (2011) Link communities reveal multiscale complexity in networks. Nature 20: 761-764.
6. Michael A, Catherine BA, Judith BA, David B, Heather B, et al. (2000) Gene ontology tool for the unification of biology. The Gene Ontology Consortium Natural Genetics 25: 25-29.
7. Valadimir B, Andrej M (1998) Pajek Program for Large Network Analysis. Connections 21: 47-57.
8. Charles BA, Stephen KE, Washburn PM (2012) Affinity purification of protein complexes for analysis by multidimensional protein identification technology. Protein Expression and Purification 86: 105-119.
9. Gray BD, Christopher HW (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC bioinformatics 4: 1-2.
10. Bobby B, Chris S, Mike T (2003) Osprey a network visualization system. Genome Biology 4: R22.
11. Gordan C, Timothy HF (1988) Distance geometry and molecular conformation. New York Wiley.
12. Kim CA, Debra R, Chiristele RG, Philippe B (2008) A sparse PLS for variable selection when integrating omicsdata. Stat Application of Genetic Molecular Biology 7: 35.
13. Kim CA, Martin GP, Christele RG, Phillippe B (2009) Sparse canonical methods for biological data integration application to a cross platform study. BMC Bioinformatics 10: 34.
14. Dongjun C, Hyonho C, Sunduz K. Sparse Partial Least Squares (SPLS) Regression and Classification.
15. Han C, Eunjun L, Yu L, Trey I (2007) Network based classification of breast cancer metastasis. Molecular System Biology 3: 140.
16. Venkateswara CR, Jing W, Julian DA (2007) Using Fly Atlas to identify better *Drosophila melanogaster* models of human disease. Natural Genetic 39: 15-20.
17. Andrew CA, Bobby BJ, Heinicke S (2013) The BioGRID interaction database. Nucleic Acids Research 41: D816-D823.
18. Ming CS, Jaiwei H, Philip YS (1996) Data mining an overview from a database perspective. IEEE Trans. Knowledge and Data Engineering 8: 866-883.
19. Pankaj C, Jaewoo K, Jiang Y, Goo LM (2008) Microarray data mining using landmark gene guided clustering. BMC Bioinformatics 9: 92.
20. Ivan CG, Roland K, Lennart O, Alexnder S (2007) Semi supervised learning for the identification of syn expressed genes from fused microarray and in situ image data. BMC Bioinformatics 8: 10S3.
21. David C, Antonio MF, Robin H, Marija M, Joel W, et al. (2014) The Reactome pathway knowledgebase. Nucleic Acids Res 42: D472-477.
22. Tibor C, Agnes K, Sandor B (1998) Comparison of partial least square method and canonical correlation analysis in a quantitative structure retention relationship study. J Biochem Biophys Methods 36: 131-141.
23. Dandrade R (1978) U Statistic Hierarchical Clustering. Psychometrika 4: 58-67.
24. Dahlquist DK (2004) Using Gen MAPP and MAPP Finder to view microarray data on biological pathways and identify global trends in the data. Bioinformatics, Chap. 7, Unit 75.
25. Zhang FFX, Brodley EC (2003) Solving cluster ensemble problems by bipartite graph partitioning. In Proceedings of the 21st International Conference on Machine Learning. New York ACM Press: 182-189.
26. Fruchterman MT, Reingold ME (1991) Graph Drawing by Force Directed Placement. Software Practice and Experience 21: 1129-1164.
27. Frey JB, Delbert D (2007) Clustering by passing messages between data points. Science 315: 972-976.
28. Shayne (2008) Alzheimer's Disease Research Paper University of perpetual Help System DALTA- Molino Campus Molino 3 Bacoor CaviteS Y 2008-2009.
29. Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, et al. (2011) Alzheimer's disease Lancet 377: 1019-31.
30. Lambert JC, Verbaas ICA, Harold D, Naj AC, Sims R, et al. (2013) Meta-analysis of 74 046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45: 1452-1458.
31. Younghee L, Haiquan L, Jianrong L, Ellen R, Ikbel A, et al. (2013) Network models of genome-wide association studies uncover the topological centrality of protein interactions in complex diseases. J Am Med Inform Assoc 20: 619-629.
32. Akula N, Baranova A, Seto D, Jeffrey MAN, Singleton N, et al. (2011) A

Network-Based Approach to Prioritize Results from Genome-Wide Association Studies.

33. Alzheimer's Statistics (2017) Alzheimers.net Centers for Disease Control, United States.
34. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007) Systematic meta-analyses of Alzheimer disease genetic association studies the AlzGene database. Nat Genet 39: 17-23.
35. De R, Verma SS, Drenos F, Holzinger ER, Holmes MV, et al. (2015) Identifying gene gene interactions that are highly associated with body Mass Index using quantitative multifactor dimensionality reduction (QMDR), BioData Min 8: 41.
36. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, et al. (2001) Multifactor dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69: 138-147.
37. Breast cancer. wikipedia
38. Perou CM, Sørlie T, Eisen MB, Rijn MVD, Jeffrey SS, et al. (2000) Molecular portraits of human breast tumours. Nature 406: 747-752.
39. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27: 1160-1167.
40. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, et al. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. The Lancet 365: 671-679.
41. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, et al. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486: 346-352.
42. Dittman DJ, Khoshgoftaar TM, Napolitano A (2013) Selecting the Appropriate Ensemble Learning Approach for Balanced Bioinformatics Data, Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference: Pp: 329-334.
43. Gao K, Khoshgoftaar T, Wald R (2017) Combining Feature Selection and Ensemble Learning for Software Quality Estimation, Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference: Pp: 47-52.
44. The MNT Editorial Team (2016) Cancer What you need to know. Medical news today.
45. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. (2010) International network of cancer genome projects. Nature 464: 993-998.
46. Wang Q, Zhang J, Song S, Zhang Z (2014) Attentional Neural Network Feature Selection Using Cognitive Feedback.
47. Dean J, Ghemawat S (2008) Map Reduce simplified data processing on large cluster. Communications of the ACM 51: 107-113.
48. Gantz J, David R (2012) THE DIGITAL UNIVERSE IN 2020 Big Data, Bigger Digital Shadow s and Biggest Growth in the Far East. IDC 3: 1414.
49. Chen CLP, Zhang CY (2014) Data-intensive applications challenges techniques and technologies a survey on big data. Inform Sci.
50. Chen M, Mao S, Liu Y (2014) Big data survey. Mobile Networks and Applications 19: 171-209.
51. Timo Elliott's Blog (2017) Digital Business & Business Analytics.
52. Jake L, Min W, Deepika GK, Yiqing Z (2016) Big Data Application in Biomedical Research and Health Care A Literature Review. Biomedical Informatics Insights 8: 1-3.
53. Xin D, Donghui W (2015) Big Data Analytic Technology for Bioinformatics and Health Informatics Special Session at 2015 IEEE Symposium on Computational Intelligence in healthcare and e-health (IEEE CICARE 2015).
54. Emdad K (2014) Addressing bioinformatics big data problems using natural language processing help advancing scientific discovery and biomedical research. Modern computer applications in science and education: Pp: 221-228.
55. Matt B (2018) Integrated Informatics Driving Translational Research & Precision Medicine. Cambridge Healthtech Institute's 23rd International Conference. Moscone North Convention Center San Francisco CA.
56. Carter TC, Max MH (2016) Challenges of Identifying Clinically Actionable Genetic Variants for Precision Medicine. J Healthcare Engineering: Pp: 1-14.
57. Health informatics. Wikipedia
58. O'donoghue J, Herbert J (2012) Data management within mHealth environments Patient sensors mobile devices and databases. J Data and Information Quality 4: 5.
59. Mettler T, Raptis DA (2012) What constitutes the field of health information systems Fostering a systematic framework and research agenda. Health Informatics J 18: 147-156.
60. Chen J, Qian F, Yan W, Shen B (2013) Translational biomedical informatics in the cloud present and future. BioMed Res Int 8.
61. Martin M (2013) Big C data social media combo poised to advance healthcare. HPC Source: Pp: 33-35.
62. Adibi, Sasan (2015) Mobile Health a technology road map. Springer. ISBN 978-3-319-12817-7.
63. A world of witnesses (2008) A survey of mobility A world of witnesses The Economist. EpiSurveyor software in public health monitoring in Africa.
64. Nick M (2013) Charting the Future of Capacity Building for mHealth Tech Change.
65. Schatz B, Marsh C, Patrick K (2012) Research challenges in measuring data for population health to enable predictive modeling for improving healthcare. ACM SIGHIT Rec 2: 36-41.
66. Jiang Y, Liao Q, Cheng Q (2012) Designing and evaluating a clustering system for organizing and integrating patient drug outcomes in personal health messages. AMIA Annu Symp Proc: 417-426.
67. Cavanillas, Jose M, Curry E, Wahlster, Wolfgang (2016) New Horizons for a Data Driven Economy Springer.
68. Magoulas, Lorica R, Ben (2009) Introduction to Big Data.
69. Snijders C, Matzat U, Reips UD (2012) Big Data Big gaps of knowledge in the field of Internet. Inter J Internet Sci 7: 1-5.
70. Wild S, Roglic G, Green A (2004) Global prevalence of diabetes estimates for the year 2000 and projections for 2030. Diabetes Care 27: 1047-1053.
71. Centers for Disease Control Prevention (2014) National diabetes statistics report Estimates of diabetes and its burden in the United States, Atlanta, GA: US Department of Health and Human Services.
72. Lindstro MJ, Louheranta A, Mannelin M (2013) The Finnish Diabetes Prevention Study (DPS). Lifestyle intervention and 3-year results on diet and physical activity. Diabetes Care 26: 3230-3236.
73. Li G, Zhang P, Wang J, Edward WG, Wenying Yang, et al. (2008) The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: A 20-year follow-up study. Lancet 371: 1783-1789.
74. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, et al. (2006) The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance. Diabetologia 49: 289-297.
75. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW (2009) Two risk-scoring systems for predicting incident diabetes mellitus in US adults age 45 to 64 years. Ann Intern Med 150: 741-751.
76. Stern MP, Williams K, Haffner SM (2002) Identification of persons at high risk for type 2 diabetes mellitus. Do we need the oral glucose tolerance test. Ann Intern Med 136: 575-581.
77. Chen L, Magliano DJ, Balkau B, Stephen C, Zimmet PZ, et al. (2010) AUSDRISK An Australian Type 2 Diabetes Risk Assessment Tool based on demographic lifestyle and simple anthropometric measures. Med J Aust 192: 197-202.
78. Lindstro MJ, Tuomilehto J (2003) The diabetes risk score: A practical tool to predict type 2 diabetes risk. Diabetes Care 26: 725-731.
79. Groves P, Kayyali B, Knott D (2013) The big data revolution in healthcare. Boston McKinsey Quarterly.
80. Bellman RE (1957) The theory of dynamic programming. Rand Corporation technical report.
81. Susan DH (2012) Toward Precision Medicine A New Social Contract. Transl Med 4: 127.

82. The future of health begins with All of us (2018) US Department of Health & Human Services National Institutes of Health United states.

83. Collins FS, Varmus H (2015) A New Initiative on Precision Medicine. N England J Med 372: 793-795.

84. Reducing and preventing adverse drug events to decrease hospital costs (2001) Agency for health care research and quality 8: 1-20.

85. Siobhan D, Angela F, Moray N, Martin W, Shaun T, et al. (2015) Drug-disease and drug-drug interactions systematic examination of recommendations in 12 UK national clinical guidelines. BMJ : 350

86. Caterina P, Antonello DP, Chiara G, Chiara C, Giacomo L, et al. (2013) Pharmacokinetic drug drug interaction and their implication in clinical management. J Res Med Sci 18: 601-610.

87. Becker ML, Kallewaard M, Caspers PW, Visser LE, Leufkens HG, et al. (2007) Hospitalisations and emergency department visits due to drug-drug interactions a literature review. Pharmacoepidemiol Drug Saf 16: 641-51.

88. Daly AK (2013) Pharmacogenomics of adverse drug reactions. Genome med 5: 5.

89. Manu M, Parul S, Harinder S, Gajendra PSR, Subhash MA (2012) NPACT Naturally Occurring Plant based Anticancer Compound Activity Target database. Nucleic Acids Res 41(Database issue): D1124-D1129.

90. Verma RP, Hansch C (2010) QSAR modeling of taxane analogues against colon cancer. Eur J Med Chem 45: 1470-1477.

91. Fauzee NJS, Dong Z, Wang YI (2011) Taxanes promising Anticancer drugs. Asian Pac J Cancer Prev 12: 837-851.

92. Song L, Chen QH, She XE, Chen XG, Wang FP (2011) Conversional synthesis and cytotoxic evaluation of novel taxoid analogs. J Asian Nat Prod Res 13: 787-798.

93. Niko B, Martin D, Mark O, Klaus K, Daniel H, et al. (2003) Estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res 31: 3850-3855.

94. Soo YR, Jonathan T, Gauhar W, Asa BH, Douglas L, et al. (2007) Genotypic predictors of human immunodeficiency virus type 1 drug resistance 103: 17355-17360.

95. Kelly T, Yang W, Chen CS, Reynolds K, He J (2008) Global burden of obesity in 2005 and projections to 2030. Int J Obes 32: 1431-1437.

96. Ogden CL, Carroll MD, Kit BK, Flegal KM (2014) Prevalence of childhood and adult obesity in the United States 2011-2012. JAMA 311: 806-814.

97. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, et al. (2009) A core gutmicrobiome in obese and lean twins. Nature 457: 480-484.

98. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, et al. (2011) Vaginal microbiome of reproductive age women. Proc Natl Acad Sci 108: 4680-4687.

99. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UM, et al. (2012) Temporal dynamics of the human vaginal microbiota. Sci Transl Med 4: 132-152.

100. Koumans EH, Sternberg M, Bruce C, McQuillan G, Kendrick J, et al. (2007) The prevalence of bacterial vaginosis in the United States, 2001-2004 associations with symptoms sexual behaviors and reproductive health. Sex Transm Dis 34: 864-869.

101. Wiesenfeld HC, Hillier SL, Krohn MA, Landers DV, Sweet RL (2003) Bacterial vaginosis is a strong predictor of *neisseriagonorrhoeae* and *chlamydia trachomatis* infection. Clin Infect Dis 36: 663-668.

102. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74.

103. Jason E, Manolis K (2015) Large scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nature Biotechnology 33: 364-376.

104. Bock C, Lengauer T (2008) Computational epigenetics. Bioinformatics 24: 1-10.

105. Lim SJ, Tan TW, Tong JC (2010) Computational epigenetics the new scientific paradigm. Bio information. 4: 331-337.

106. Evans DB, Tandon A, Murray CJL, Lauer JA (2000) The comparative efficiency of national health systems in producing health: an analysis of 191 countries GPE Discussion Paper Series no 29. World Health Organization Geneva Switzerland.

107. Tandon A, Murray CJL, Lauer JA, Evans DB (2001) Measuring overall health system performance for 191 countries GPE Discussion Paper Series 30, World Health Organization Geneva Switzerland.

108. Category Medical scoring system. Wikipedia.