



Early Disease Detection through Data Analytics: Turning Healthcare Intelligence

Heta Hemang Shah
Independent Researcher USA
Heta2973@gmail.com



Corresponding Author

Heta Hemang Shah
Heta2973@gmail.com

Article History:

Submitted: 22-10-2023

Accepted: 22-10-2023

Published: 14-11-2023

Keywords: screening, analysis, information technology, big data, algorithms, artificial intelligence, machine learning, predictive modeling, wearable's, privacy concerns, voluntariness, ownership, bias, health access, patient-physician contracts, inequality, medical morality

ABSTRACT

The incorporation of data analytics into disease diagnosis is still in its primitive stage however, its capability to enhance efficiency and patients' health is tremendously exciting. The use of artificial intelligence (AI), machine learning, wearable technologies, and predictions analyzed can make healthcare systems enable early diagnosis of diseases and cost less to develop preventative health care. However, the change brings several issues of ethical consideration into the equation. Some of such challenges comprise of privacy and confidentiality of sensitive health information, challenges in getting consent and ownership of the patient data. Furthermore, this work discussed that algorithmic bias issue might expand the health disparities if solutions are not found, thus exacerbating inequalities in of health care treatment. The continued development of added AI tools should also be met with the preservation of trust in the doctor and the patient to prevent the man-made intelligence to lead into a replacement of human cognition. For these technologies to benefit the greatest number of people, the advanced technology solutions must be made readily available to avert further widening of the inequalities in health. These ethical questions are best understood and therefore shall be discussed in this paper, with a focus on privacy, consent, fairness, and accessibility. In assessing these issues, the concern of concern can use the big data analytics to enhance the illness diagnosis without violating the patient's privileges and/or discriminative the equal rights of health care.

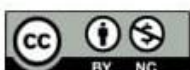
Brilliance: Research of

Artificial Intelligence is licensed under a Creative Commons Attribution-Noncommercial 4.0 International (CC BY-NC 4.0).

INTRODUCTION

Another reason for conducting early disease detection is that capable of finding out whether a person is at risk of developing a severe or life-threatening disease. The most important aspect of early diagnosis relates to the possibilities of beginning treatment as early as possible in the development of illness, thus enhancing the patient's recovery, minimizing the necessity for extensive therapies and keeping down the costs of medical procedures. Earlier detection, in the past, has been dependent on general check-up examinations, clinical examinations and symptoms presented by the patient [1]. Although these methods are important, diseases are often not discovered during their initial stages, when signs are non-existent. Over the past decade, the application of latest tools for data analysis in healthcare has significantly improved the chances of identifying diseases on an early stage. Many large sets of patients' data including the medical history, genetic predispositions, and vital signs as well as the social and physical activity status of individuals enables healthcare providers to detect patterns and risks that point to a particular disease before it manifests itself fully. This relied-on healthcare data approach is now a leading shift from diseases and illnesses diagnosis and treatment from a reactive state to a predictive individualized one [2].

AI enhanced diagnostic tools are now very far from what would have been considered possible just a few years ago. Despite the fact that machine learning algorithms, artificial intelligence (AI), as well as statistical and actuarial models that use predictive analytics, have advanced significantly in their analytical capabilities, this new development has made them capable of sifting through massive datasets with stunning accuracy. These technologies can identify certain interconnections between iconized genes, tips of the environment, and medical backgrounds that somehow elude diagnosis with macroscopic. For instance, the AI can review and diagnose problems like cancer, heart diseases, neurological diseases from images like x-rays and MRI that the human eye might fail to see. Data analytics also provides an active way of observing the progress of a specific patient through an integration of data from the wearable technology coupled with sensors and electronic health record data. These tools give health care professionals a broader picture of the patient status through the details of the vital signs, physical activity and sleep status. From this, analytics platforms are able to identify if a patient has gone off their normal behavior concerning health, provide a prompt if there is something





wrong before the condition deteriorates and the patient shows signs of ailment [3].

Another dramatic effect implied by data-driven disease diagnosis at early stages is individualized approach to patients. The insight derived from data analysis enables the healthcare providers to prescribe or recommend treatments that are personalized by the patient's genetics and lifestyle, previous and current health history. As opposed to a generic approach to treatments, this approach increases the chance of treatments being the most fitting and appropriate for each patient. Furthermore, data analytics results in preventive care which is when the patient is treated before the condition deteriorates. For example, those patients who are identified to have a higher risk for developing chronic diseases like diabetes, cardiovascular diseases among others shall be taken care of through the enhanced data systems [4]. It can allow for changes in the patient's behavior, medication and operations in order to stop the aggravation of the disease. It thus helps cut down costs lower or even eradicate expensive, urgent care that is usually so expensive but also has a positive impact on the lives of many patients.

The use of data analytics in early disease diagnosis can be a major game changer in healthcare systems globally. Early arresting diseases also acts to prevent overloaded hospitals especially the emergency answering units and the intensive care units, as the majority of the patients are usually in the final stages of their diseases. Further, early stage diseases are easier to cure, and therefore incur lower costs than when the diseases are advanced. Through preventions of severe diseases, an analyst protects the public health, increase their quality of life, and decrease the mortality rate. Using data obtained from EHRs, health systems can also optimize its resources effectively target high-risk population, and minimize hospitalizations costs. Due to the increase in the use of data analytics, early diseases diagnosis has been greatly enhanced because healthcare practitioners have better tools to use when trying to establish diseases at their initial stage. By integrating the data from different domains like genetics, life style, and real time data available on patient, the healthcare professionals can have earlier and better intervention. This analysis-centered, patient-tailored model not only benefit patients, but also lower the healthcare expenses and increase the effectiveness of the systems in place [5]. The future of disease detection is therefore bound to be centered on data analytics as the procedures continue affording care a more precise, proactive, and cost efficient frame work.

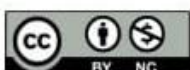
Data analytics has been hailed as the solution and salvation to many industries, including healthcare.

The use of big data analytics in healthcare is characterized extensively in handling the detection and diagnosis of diseases as well as in the planning of the proactive treatment processes. Subclinical to chronic pathophysiologic processes, under- and overuse disorders, as well as important variants that may be obscured by noise in large databases, such as patients' medical histories, imaging, genomics, vital signs, and continuous monitoring, can be picked up by data analytics and would be nearly impossible for HCs to notice empirically [6]. The use of such approach can enhance diagnosis procedures, make them faster, more accurate and efficient, which eventually leads to enhanced patient outcomes with increased attendance ratio. It is worth mentioning that within the past few years the healthcare industry has experienced a growing avalanche of data brought about by EHR systems, wearable devices, genomic sequencing, and much more. That's why, along with such growth of data, there appeared a requirement for such tools and techniques which would help for better analysis of these data. A growing portion of practical and often subjective decisions in the healthcare field is complemented with data analysis to improve its accuracy and efficacy.

Healthcare big data is therefore the large volumes of information which healthcare facilities produce such as patient data, diagnosis records, scans, and even life logs from wearable devices. Of great importance is the fact that through this data one can understand trends, risks, and diseases progression in a way that would otherwise be very hard. Data analytics make the following sense of the vast amount of information in order to improve patient care. Another tool in data analytics is a sub-discipline of artificial intelligence known as machine learning (ML) that uses algorithms to train computers on data sets in order to enable the computers to solve problems on their own. In healthcare, the machine learning algorithms are applied to patients' information and probability of the disease before it reaches an alarming stage is determined [7]. For example, machine learning algorithms can be applied on sizable datasets of patient records when constructing a model that provides probabilities for certain diseases, such as heart disease, diabetes or cancer, for future patients.

It can help identify diseases early through accrued patterns in real situations, concerning pictures like X-ray, MRI, and CT scans for different diseases like tumors, fractures or infections at their preliminary stages. These tools are useful to the extent that they help decrease diagnostic gaps, provide more rapid diagnosis, and allow for treatment at early stages when diseases are still easily manageable. One of the most use application for data analytics in healthcare is predictive analytics where an organization is able to use past data and statistical models to forecast future events. They can use this information to anticipate which patients are most susceptible to certain diseases because of their inherited traits and life styles and their environments [8]. If these risks are detected early, then patients can do some changes to their lifestyle or start taking medications or having more check-ups to decrease these risks of being severe.

For instance, some models can approximate the patient's risk factors for developing ailments such as diabetes, hypertension, and so on, whereby the doctor can modify or treat the ailment before escalating. The move towards more preventative oriented treatments doesn't only help to decrease the overall impact of chronic diseases; it also provides a more financially feasible solutions which show that early treatment is less costly compared to late-stage treatments. Real time; the final paradigm of data analytics in healthcare area is possible due to wearable devices, remote patient monitoring systems and IoT technologies. These devices gather a steady flow of information regarding the patient's heart rate, blood





pressure, glucose levels, oxygen saturation etc. The information is then communicated to healthcare givers where a patient's status can be closely observed and appropriate action taken [9].

Decision support systems spearheaded by the data analytics also help the clinicians by providing best practice guidelines within the healthcare facility for patients. These systems incorporate information from multiple sources to compute analysis, proposed diagnosis, and personalized treatment plan for a specific patient to reduce guesswork during treatment and enhance decision-making among clinicians. Big data is an important facilitator of personalized or precision medicine that attempts to innovate treatment based on genetic makeup, lifestyle, and environment of the patient. Using genomic analysis, the doctor can discover whether the patient has certain mutations that are likely to cause specific diseases or even markers that define a specific illness [10]. Medical reports can be tailored to enhance the healing process of a particular disease, also decrease side effects, and as a result enhance the quality of life of patients.

Data analytics is at the very forefront of transforming the healthcare industry since it redesigns the ways in which diseases are identified, diagnosed, and managed. By utilizing state of the art health information and communication technologies and specifically applications such as machine learning, predictive analytic and real time monitoring, healthcare delivery organizations are in a better position to deliver timely, personalized and accurate healthcare interventions [11]. Over the course, data analytics is estimated to be at the center of enhancing patient care, controlling costs in the health sector, and shifting towards preventive or personalized medicine. The continuous incorporation of Big Data Analytics in the healthcare sector is steadily leading to a revolutionary innovation of the healthcare delivery system to the advantage of the patients and the health care providers.

IMPORTANT TECHNOLOGIES WHEN IT COMES TO DATA DRIVEN DISEASE DIAGNOSIS

Disease diagnostics through the use of data starts on what we can refer to as several foundational technologies that allow the healthcare systems to gather, process, and understand big data related to individual and community health. All these technologies fuse to improve the specificity, speed and efficiency in the identification of diseases, and subsequent provision of treatment [12]. The most notable of these technologies are AI/ML as well as Big Data and Cloud Computing in Healthcare, and the Internet of Medical Things (IoMT).

Healthcare innovations, particularly disease detection, are leading industry indicators of Artificial Intelligence (AI) and Machine Learning (ML). Artificial intelligence means imitating human intelligence using machines and machines that systematically improve the performance of the system when presented with new data, without being programmed for the new data. With reference to disease diagnosis, AI and ML can capture, process and analyze huge volumes of data to come up with diagnosis probabilities on patients. Such an example of an application of ML is when using patient records of the past to estimate the probability of a patient contracting a disease. In the application of the oncology field, it is possible to ML models that can predict early stages of cancer from a collection of MID's such as X-ray images, MRI, or CT scans. Such models can be useful to detect emerging differences in body tissues that is characteristic of the primary cancer, sometimes even at the early stages that are hardly noticeable by naked eye [13].

Besides diagnostics, AI and ML can help in intelligent decision making in regard to the prognosis and outcome of a particular disease, as well as in the development of individualized therapeutic approaches that pursue better outcomes of an antithetic treatment. The use of AI in decision support systems enables professional staff in the health sector to make informed decisions based on analyzed data thus minimizing errors, improving on time taken to implement various solutions and increase rate of interventions. Big Data encompasses large, varied and complex data collected in healthcare organizations such as digital patient's health record, genomic data, data from wearable devices and patients' real-time vital sign data. Big Data is very effective in analyzing trends, alerting health risks and even identifying diseases in their early stages [14]. For instance, in a large population of patients the investigators can identify relationships between life patterns and disease progression that would be beneficial in estimating the risks of developing non-curable diseases such as diabetes or cardiovascular diseases. Moreover, Big Data can assess more 'epidemiological profiles,' the behavior and progression of disease, and estimate where they are likely to spread, which is important in cases of infectious diseases such as COVID-19.

Nonetheless, handling and analyzing Big Data needs considerable computer resources. Big data introduces the need for a platform on which to store, access and process such immense data, for which cloud computing seems to offer. Advancements in cloud for healthcare offers providers who use cloud platforms ability to access data in real time; to share information across institutions, or analyze patients' data at a distance. The calculation of the cloud promises scaling, security, and the ability to quickly analyze large volumes of data effectively to identify diseases. Such data are from various outlets including hospitals, research institutions, and wearable health devices to the cloud. This integrated model allows for a broader assessment of a patient's condition, and this is very important for purposes of early diagnosis of diseases [15].

Internet of Medical Things (IoMT) is defined as a convergence of Medical Internet of Things that allow for real-time flow of patient data through connected medical devices. Such gadgets are like fit trackers, smart insulin pumps, ECG/ECG, BP cuffs, and even more sophisticated devices like non-hospital based gadgets like remote ultrasound scanners. With continuous and real-time recording of patient's health status, IoMT is essential in early diagnosis of diseases. For example, smart watches and other wearables can monitor pulse, blood pressure, or glucose level, and notify a patient and their





doctor in case of a shift [16]. This kind of surveillance becomes useful in identifying early stages of diseases such as heart disease or diabetes when the conditions are not lethal. In the prolonged disease treatment procedure, IoMT devices can notify the attending medical staff if the patient's condition is getting worse, especially in asthma or hypertension treatment, to prevent re-admission.

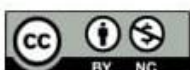
Apart from curve monitoring, IoMT enables monitoring of the patient's compliance with recommended treatments and rehabilitation processes after the treatment is complete. Such data can also be used in conjunction with other prediction systems in endeavouring to predict possible future health problems to permit treatment to be better tailored and probable problems to be prevented. One more widely applied technology used in data-driven detection of diseases is genomics, technology referring to the study of specific person's genes and their relation to various diseases. Data analysis interfaced with biological data is progressively used to scan through the genomes of patients, although genetic markers and mutations indicating the risk of developing certain diseases can easily be detected [17]. By sequencing a patient's genome, doctors and other caregivers are able to get information about genes relative to diseases including cancer, genetic illnesses, and other diseases. It can also be used for screening, prevention and put personalization is possible by using the genetic signature of the patient. For instance, genetic diagnostics are widely applied in the diagnostics of breast cancer and may detect mutations of the BRCA1 or BRCA2 genes to take informed decisions on treatment and prevention.

However, bioinformatics applications enable the study of genomic differences in various populations in order to determine how genetic differences affect diseases. It plays a crucial role in determining genes associated with particular diseases and risk factor for the disease in early stages and designing strategies for their treatment [18]. There is another important technology identified as Natural Language Processing (NLP) useful in disease diagnosis from unstructured textual information. Doctor's prescriptions, patients' files, trial results, and articles are usually composed in natural language and cannot be analyzed by a human being by sight. Text sources are analyzed and decoded by NLP algorithms, where signs of diseases, and risks are indicated. For instance, NLP tools can automatically review a patient's chart for the following: searching for specific keywords or patterns for the doctor's description that is suggestive of early sign of a disease, even where the patient has not volunteered it. This in turn helps healthcare providers to refer patients on risk factors so that they can be attended to early [19].

So, the integration of the four fundamental technologies such as AI and ML, Big Data and Cloud Computing, IoMT, genomics, and NLP has created a way to revolutionize screening disease earlier. Combined, they allow healthcare providers to tap data in ways never before possible and so promise to speed up the diagnostic process and improve treatment outcomes. The applications of these advancements will only progress as time goes on and expanding the standing of these technologies in healthcare systems will only increase our capacities for early disease identification and to improve overall healthcare [20].

Diseases which can be detected by Data Analytics

Data analytics has brought a significant change to the healthcare industry since different diseases can be diagnosed from the outset. To provide context, many patients' data sources including EHRs, biosensors, medical imaging, and genomic data can be harnessed to identify diseases early enough, when they are easily manageable. The advanced settings of data incorporation in diagnosis embody artificial intelligence (AI), machine learning (ML), and predictive modeling that have improved the recognition of a large variety of ailments: chronic, infective, and genetic. In this work, the various types of diseases that can be identified through data analytical techniques are discussed.



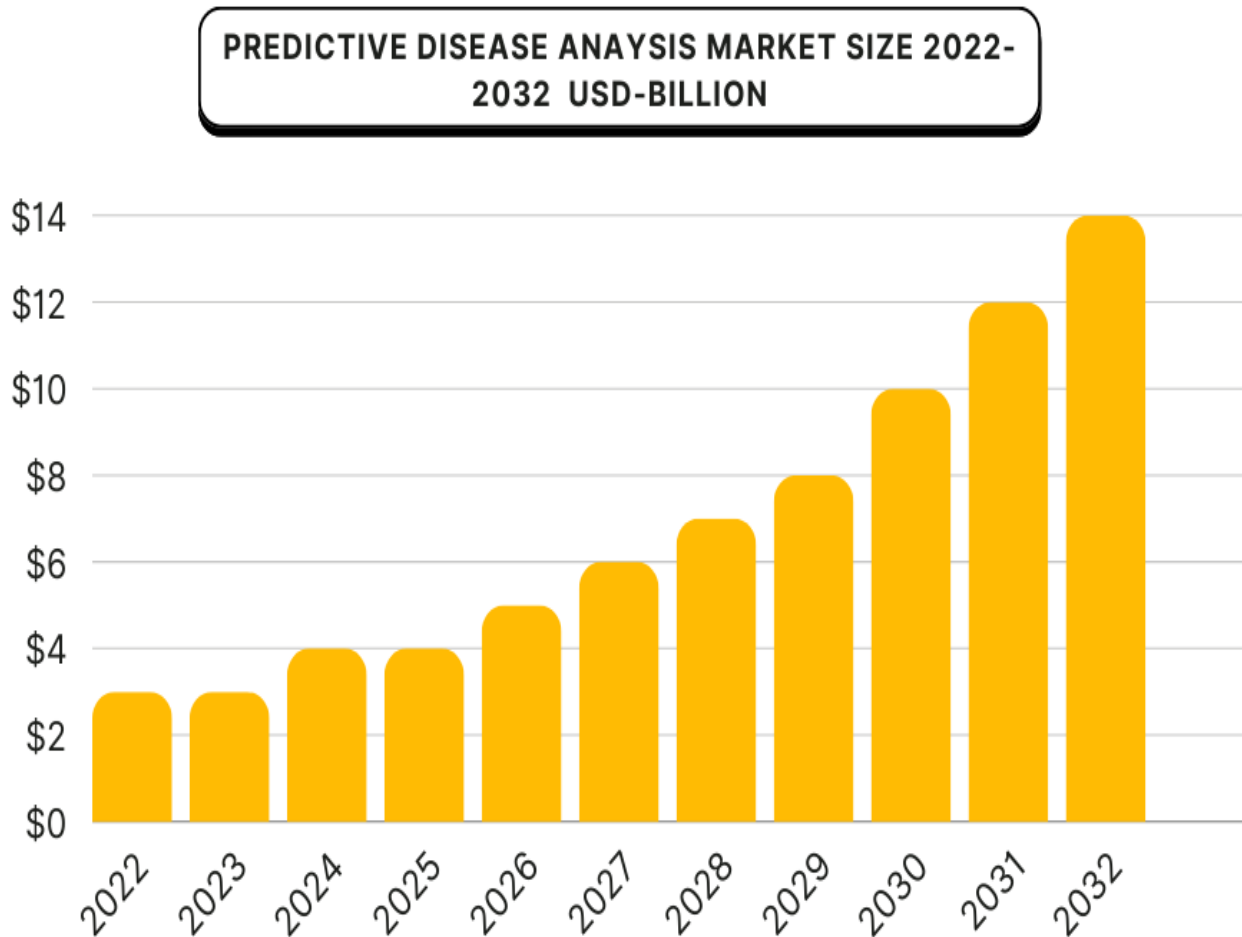


Figure: 1 showing predictive analysis market size 2022-2032

Chronic Diseases: Conditions like diabetes, heart disease, and hypertension are the world's leading causes of death, to begin with. Since many chronic diseases are slow to develop, very limited indicators of its onset are usually available [21]. Using data, it is possible to detect endangered people, early diagnosis of these diseases, as well as to control them.

Diabetes: Machine learning can be used to predict patients, who are more likely to develop type 2 diabetes by computing the relative risk with factors like age, BMI (Body Mass Index), family history, blood sugar, and diet. The AI models can compute all these factors and with them estimate the likelihood of the onset of diabetes, thus with early treatment in form of lifestyle changes and medication or others, can the onset of the disease be effectively delayed [22].

Cardiovascular Diseases (CVD): This is true because data analytics can also be used in the early detection of heart ailments, attacks or strokes. Artificial neural networks are capable of taking into account such variables as blood pressure, cholesterol, family dispositions, and even such aspects of a person's life as smoking and physical activity [23]. Some of the steady state characteristics also serve as predictive biomarkers and include ECG data, heart rate variability and in some cases imaging data like the echocardiograms.

Hypertension: Hypertension may not present any symptoms before it worsens to a critical level affecting heart health, kidney function and leading to a stroke. Through wearable devices that keep checking blood pressure regularly, integrated with artificial intelligence-based analysis technologies, healthcare givers will be given a chance to act appropriately should there be rising trends in fluctuations of blood pressure in their patients' systems thus handling hypertension becomes easier [24].

Infectious Diseases

Bacterial and viral infections are some of the most common and can easily affect a number of people within a given



population. This has made early diagnosis one key pointer in the containment of the diseases, saving many lives and indeed offer the affected individual proper attention as they need. Big data is highly beneficial for pattern recognition combined with the ability to predict disease prevalence and, hence, prepare for them [25].

COVID-19: Data analysis played a vital role of identifying, controlling and tracking the COVID-19 outbreak hence playing a crucial role. Some even used predictive models and machine learning algorithms from data from reports of public health, test results, and contact tracing apps to calculate and determine disease spread and hotspots and even help shape healthcare policies [26]. We have also used artificial intelligence in the examination of chest X-Ray and CT scan to discern the signs of COVID-19 to improve on diagnosis and treatment.

Tuberculosis (TB): Tuberculosis is still as the leading infectious disease killer in the world and more so to the emerging nations. In particular, data analytics can be used for early TB outbreak detection and for making prognosis on future cases. Specifically, machine learning algorithms that cut through hundreds of data points from chest x-rays, laboratory tests, and patients' histories can recognize signs of TB before it becomes fully diagnosable. Gender as well as other determinants of disease risk are obvious for the purposes of choosing areas prone to TB outbreaks to focus efforts on.

Malaria: Malaria is yet another infection where data analytics is used to identify and prevent the disease. Statistical models based on weather conditions and patient travels, together with symptoms, can be applied for malaria epidemic prediction [27]. Studying these data points allows for efficiency in destining resources to the areas they are needed most and in offering timely treatment in order to forestall the spread of the infection.

Cancer

This paper focuses on cancer as being among the leading causes of deaths globally; screening plays a vital role in the enhancement of prognosis. Using data analytics to flag signs of cancer early is another aspect that has been successful in early detections reducing the result on patients with cancers at early stages.

Breast Cancer: Screening for breast cancer can be very useful when done early enough. It is worth stating that sometimes the naked eye is unable to see certain tumors or abnormalities in Mammogram images this is where AI algorithms come in handy. In another algorithm, patient record including family history, age, genetic charter, and lifestyle characteristics can predict possibility to get breast cancer [27]. In order to apply early diagnostics and intervention, the possible candidates are outlined in order to go further through a special algorithm.

Lung Cancer: In the identification of lung cancer, machine learning is used to scan computed tomography scans and chest x-rays for suspected tumor appearance. These systems can diagnose initial-stage lung cancer, which may not even exhibit signs yet. Similarly, computer-based models that assess success probabilities on genetics and environment enable consideration of high risk patients for screening more often and earlier for lung cancer [28].

Skin Cancer: The utilization of data analytics has been most frequently used in the detection of melanoma, as well as in the identification of other types of skin cancer. While the treatment of deep learning-based models in big amounts of dermatologic images allows physicians to diagnose skin cancer with high accuracy, this approach can also detect carcinoma that clinicians might fail to see [29]. When incorporated into normal skin examination, outpatient care can also diagnose skin cancer at early stages.

GENETIC AND RARE DISEASES

There is one principal reason why the early diagnosis of genetic diseases can hardly be effective – the given diseases are either rare or the first signs of which do not pose a risk to the patient's health bear minimal similarity to the symptoms of genetic diseases. However, the development of genomics and bioinformatics has made it possible to use analytics to find out if there are genetic mutation and or rare disease early enough so as to help the patient.

Genetic Disorders (e.g., Cystic Fibrosis, Down syndrome): Since the availability of genetic tests, data analytics is applied to define such genetic markers as are associated with different hereditary diseases. Based on the relations above, genetic models can estimate the possibility of genius to pass some hereditary conditions genetically and to their descendants. Genetic screening detects these disorders at early stages hence hopeful management and prevents family planning with a plan that endangers the lives of members [30].

Rare Diseases: Since they are uncommon, patients are hardly examined and may not receive diagnoses for a long time. Data analytics is used to diagnose such conditions from symptoms, medical histories and, genetic makeup and profile. Using the advanced search technologies, a doctor can find differences in great lots of seldom occurring pathologies even if she/he cannot reveal such differences with the help of his/her experience. By developing molecular 'fingerprints' of diseases from thousands of patients at early stages of the disease, this leads to more effective treatments for rare diseases. Data Analytics is becoming ever more essential in identifying diseases at their early stages, and it is used in matters concerning chronic diseases such as diabetes and cardiovascular diseases, Top of infectious diseases, and some genetic diseases [31]. Through such tools as AI, machine learning, shot and long-range predictive models, and genomics, risk, diagnosis, and personalized treatments can be found. The benefits in enhanced disease diagnosis has major implications in the future of technology and in particular holds great promises in the future prosperity of the healthcare industry as a whole.



SOURCES OF INFORMATION ON DISEASE PREDICTION AND OR DIAGNOSIS

These data sources have become an important tool in the digital health care setting in that so much data is produced through different paths to predict and diagnose illnesses. When combined and evaluated, different datasets provide holistic information on a patient's condition, the presence of precursors of diseases, and an individual approach to therapy. This section aims at discussing the major forms of primary data which can be used in determination of diseases; these are; EHRs, wearable health devices data, imaging data, genomic data, and patient generated data. Electronic Health Records (EHRs) are currently one of the richest sources of information in the modern healthcare system [32]. Using these records, a physician is able to view the patient's medical history, diagnoses, prescriptions, treatments, drug allergies, pathology results, and more. First, EHR has complete, lifelong information about the patient, which is critical when developing disease models and diagnosing diseases.

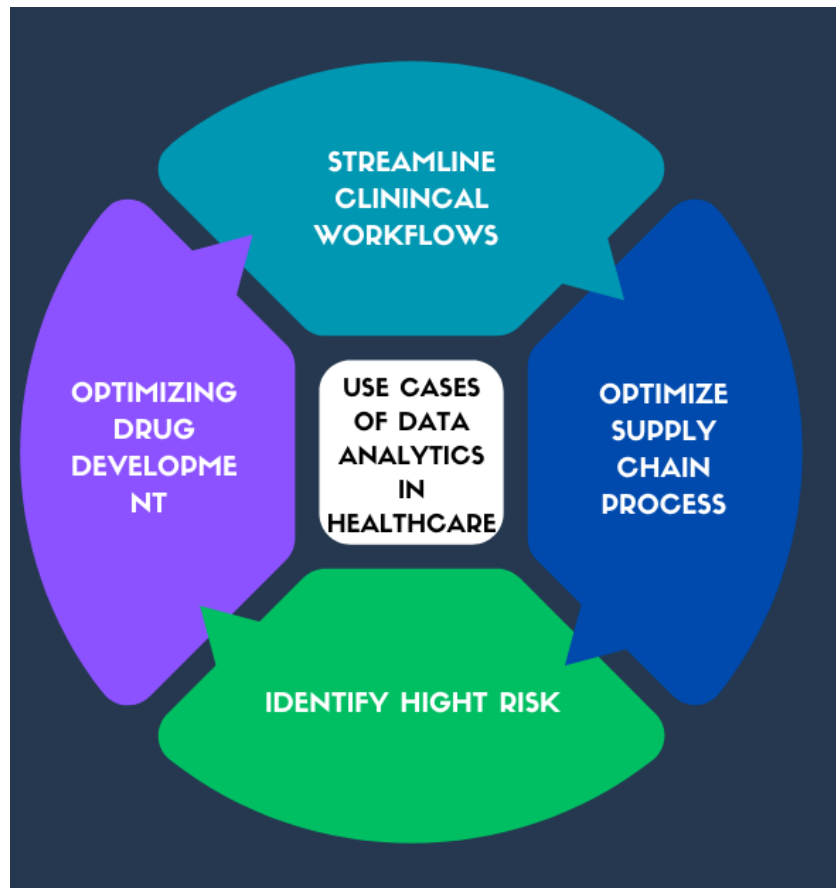


Figure:2 USE cases of data analytics in healthcare

EHRs are useful in tracking the health status changes in the patient over some period since they are a rich source of data. For instance, through data collected over time, the statistical models will be used to do risk assessment for chronic diseases like; Cardiovascular diseases, diabetes, and hypertension. These records can explain symptoms of early diseases or forecast potential health problems through Machine learning algorithms that detect patterns within a patient's records. That is why, EHRs can be employed for tracking the effectiveness of provided treatment. In such a way, there is a possibility to compare different treatment options and choose the best ones based on the outcomes of experience [33]. Moreover, EHRs bring efficiency in sharing and using patient information across multiple care givers which leads to accurate diagnosis as well.

Smart watch and fitness trackers, and remote patient monitoring devices are another crucial part of the care delivery system, nowadays. These devices are always monitoring a host of other health parameters including heart rate, blood pressure, oxygen levels, glucose level, sleeping number, and movements. Compared to traditional biometrics, Wearable's are real-time and continuous, and therefore, give actualization of a patient's health and alert of possible development of severe conditions. For instance, smart wristbands or smart watches that monitor heart rate, daily activity levels, and performance reveal signs of emerging cardiovascular trouble such as an irregular heartbeat, high blood pressure. With sensors, smart watches fitted wearers can monitor if their heart is racing and give the signs to seek for medical help. For diabetes, wearable devices designed to capture blood glucose that will help identify changes that may warrant changes in diet or insulin intake [34].



However, Wearable devices are most helpful in chronic disease situations. They supply the data that can be used continuously, therefore such diseases as asthma, diabetes, or hypertension can be well managed. While using real-time monitoring, the patient and healthcare provider can make appropriate adjustments affected by the monitoring results like changes of the dosage of drugs or alterations in life activities. Medical imaging is a very useful and invaluable tool in diagnosing different diseases including malignant tumors or neurological diseases at an initial stage [35]. These are the comparison of diagnostic tools and techniques, X-ray, MRI, CT scan and ultrasound are some examples of diagnostic imaging techniques that provide structural images of human body thereby identifying any unusual sign of diseases.

This paper has shown that the use of data analytics in medical imaging systems has enhanced the diagnosis of diseases particularly cancer. To illustrate, contemporary deep image recognition approaches can look at radiology scans to diagnose tumors, often in their nascent stages, and often overlooked by human physicians. For detecting breast cancer, the play of Machine Learning results in making mammograms and the machines are helping radiologists in improving their ability to identify malignant growths. In the same manner, the applications of AI in the neuroimaging is to identify the symptoms of neurological diseases like Alzheimer's, multiple sclerosis and even strokes. ML algorithms look at the patterns of the level of atrophy or similar things and can then correlate that with big data sets to find that some of the beginning signs of these are present and then the doctor can intervene [36].

Other collections of information are combined with the imaging data including genetic and demographical information to give more complete picture of the patient. Such approach is used to ensure that the various data sources bring out comprehensive and accurate identification of diseases. Genomic data helps identify an individual's genetic structure with such strong implications in various diseases. In the past few years enhanced technologies of genomic sequencing allow for determination of entire human genomes and genetic markers that may indicate whether a person is prone to certain health conditions like cancer, atherosclerosis or certain genetic diseases. Genomic data is essential for the diagnosis particularly of diseases in the personalized medicine environment. The patient's biomarkers can be examined to determine chances of developing certain diseases so healthcare practitioners can recommend more frequent testing and possible preventive actions [37]. For example, defect in the BRCA1 and BRCA2 genes may lead to the development of breast and ovary cancer and people with these recessive genes can be closely observed for early signs of the mentioned diseases.

It is instrumental in the enhanced generation of targeted therapies. Because of medical genetics, researchers and clinicians are able to now adapt treatment approaches that are specific to the gene change of a disease. Subsequently, this approach can increase the effectiveness of treatments and at the same time reduce the side effects. Dr honorable brought it to our notice that as the direct-to-consumer genetic testing services are becoming common, more people are having their genetic details hence enabling them to make right choices on their health. Currently, genomics is an active area with even wider integration to other databases will only augment the outcome of predicting, diagnosing and managing diseases [38]. Self-measurements mean the patient self-reports the condition, symptoms, behaviours and health status of the patient. This data is normally accrued through surveys, questionnaires or using mobile health applications; however the data gives insights that may not be obtained by other forms of data.

For instance, data on a patient's diet, exercise routine, and stress levels may show how well a patient is coping with life, whether he or she has a poor diet that may lead to obesity, diabetes, or heart disease. Likewise, some symptoms that patients can recall, including fatigue, pain, or breathing issues, may inform a clinician of possible health complications that may not show up in tests. Furthermore, self-report data is incorporated into the treatment of chronic diseases and illness. It also gives the patients chance to report on their condition/ side effects of a certain treatment in real time thus helping the healthcare provider in making right decisions with regard to the patient. Such information can be fed into prediction models so as to predict the progression of the disease and recommend appropriate action to take [39].

Combination of EHRS, wearable health, body scan data, genomic data, patient health report data, and imaging data has provided a richer and more accurate approach towards disease prognosis and diagnostics. Through the help of complex analysis of patients' data with the help of information technologies, the doctors will be able to diagnose diseases at earlier stages and take individual approach to a patient condition [40]. As these data sources continue to develop and are integrated with each other, the components of those schemata and paradigms are also going to improve in predicting, diagnosing, and managing diseases for better patient outcomes and effective and efficient healthcare outcomes.

OPPORTUNITIES AND BARRIERS SURROUNDING ANALYZING DATA FOR EARLY DISEASE DETECTION

Problems and limitations associated with the application of data analytics for the early diagnosis of diseases hold some possibilities. While some of these challenges may block the adoption of a universal big data strategy in healthcare, others can downgrade disease prediction models. This section summarizes important issues and barriers to early disease detection by using data analytics encompassing data reliability and accessibility, privacy and security, algorithm bias, absence of standardization, and requirement of skilled interpretation [41].

Missing Data: Indeed, in most cases, basic clinical data of patients is either not available or is incomplete and this poses great challenges for most of the predictive models. For instance, if a patient's medical history summary does not contain some data including family medical history or recent laboratory test results then algorithm cannot produce a full range of risks [42].



Inconsistent Data: The information used in healthcare may be obtained from various organizations like different hospitals, clinics, and laboratory. However, these may utilize different formats, terminologies or systems of recording information hence varying incompatibilities which hinders efficient synthesis and analysis of the data.

Data Availability: In some cases, the quality of the data could be poor if health care services in some regions such as low and middle-income countries are poor. It becomes challenging to apply data-driven disease detection strategies whenever there is no adequately developed health information technology in such regions; the situation increases inequalities in health. Another major issue with the application of data analytics for disease diagnosis is keeping the patient data and information sought private and secure [43]. Health care sector is one of the most vulnerable to cyber-crime and violation of patient data is a nightmare to all.

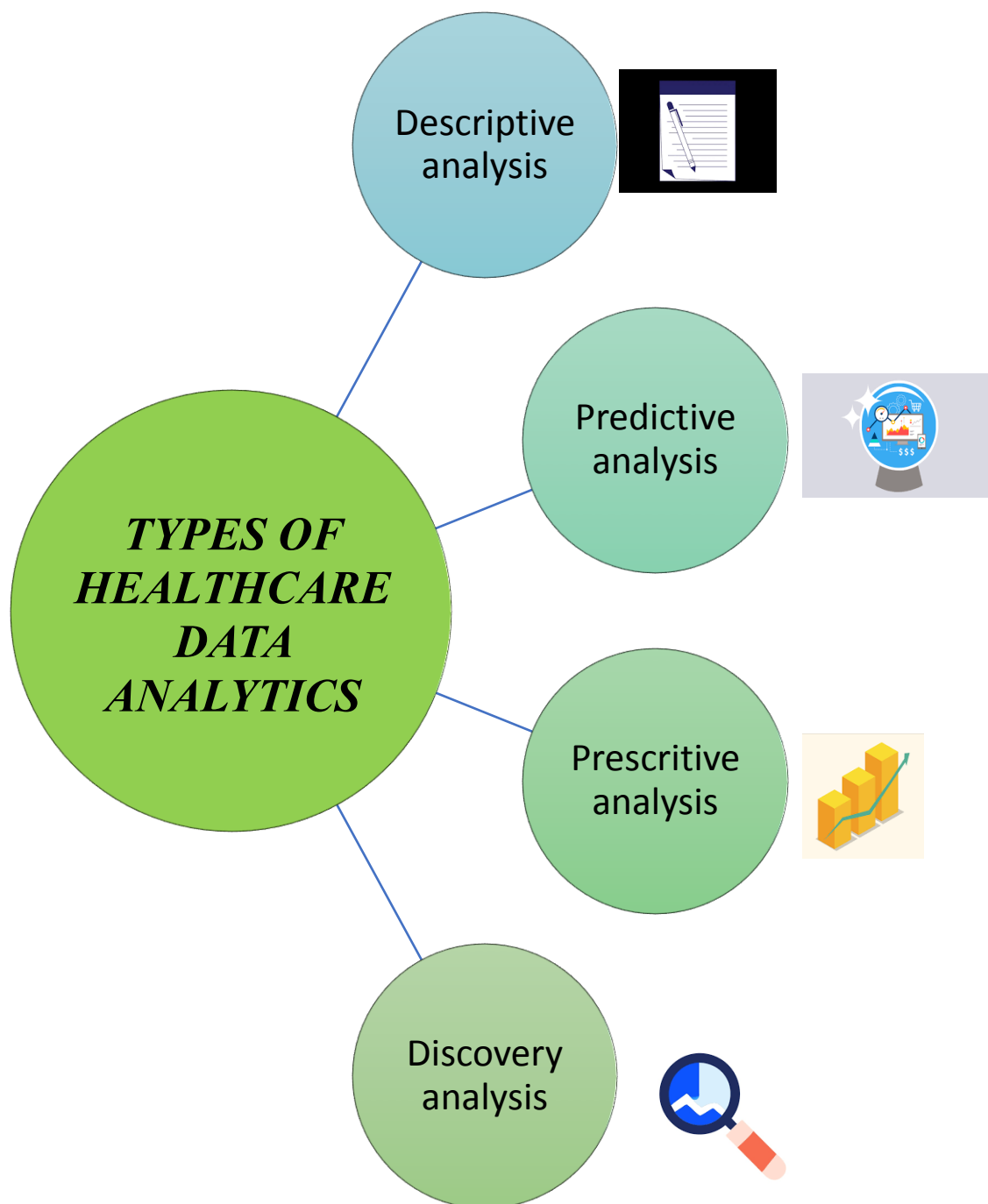


Figure: 3 showing types of healthcare data analytics



Patient Confidentiality: Information that is incorporated in a healthcare setup is usually confidential in that it may include data on the patient, his/her medical history, diagnosis among other genetic data. People remain skittish on the privacy of their health information as more health information is provided to third-party entities or moved to cloud-based storage. Realization of data analytics solutions has hence call for strict measures that respect patient's privacy and uphold confidentiality of their health information [44].

Data Security: The combination of EHRs, wearables, and genomics data make it challenging compared to isolating single sources of patient data. Security of big data is exceptionally important and presenting protection against access or cyberspace threats. When it comes to storing and processing patients' data, as the use of digital and cloud applications for various enterprises increases in the field of healthcare, sophisticated encryption, and control procedures together with security check-ups should be applied [45].

Regulatory Compliance: Health care information is normally governed by certain rules and regulation like HIPAA in United States or even GDPR regulation in the European Union. Obligation to these regulations means that results from the application of data analytics tool will respect privacy and security of patients. This means that both the healthcare service providers and the technology developers are trapped in these regulatory frameworks to avoid being on the wrong side of the law and penalties that accrue with it. That is why one of the biggest concerns related to data-driven disease detection is algorithmic bias. When the machine learning model takes training data fed to it, and that training data is bias whether through demographic or imbalanced representation, the model will replicate and reinforce this bias, which makes it both unfair and inaccurate in its predictions [46].

Bias in Training Data: When the training data sets adopted in arriving at the predictive models do not conform to the generality of patients, some patients may not be captured, or they may be categorized wrongly. For instance, deep learning systems that are developed using samples from a particular age, gender, or ethnocentric population will not be optimal in ferreting out similar traits in other generations, gender, or population [47]. This can lead to variation in rates of diagnosis with some individuals being told that they are negative while in real sense, they are positive or that they have a given ailment

Healthcare Disparities: This means that algorithmic bias can actually increase existing disparities in healthcare, in marginalized communities in particular. For example, in using data to train a model for predicting future ailments, it may work for the rich who benefit from better healthcare, not for a poor person or someone whose risks are different or who cannot afford health care [48].

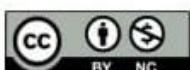
Bias in Interpretation: Algorithmic systems may initially be trained using more representative datasets, yet prejudice that lies in the post-processing of outcomes may result in misidentification or discrimination. Doctors practicing in health care sector may contribute to calibration of an algorithm by prostrating more credit on predictions made out of it thus fueling biased provision of treatments. In healthcare, there is no single format, template, or format of data collection; there is no unified approach to its processing and sharing another crucial issue for the implementation of data-driven disease detection tools on a large scale [49]. Various healthcare givers, facility or even countries may use diverse data format, codes, as well as terminology hence; creating a problem on how to compile data from different sources and authors.

Data Formats and Terminologies: Healthcare data resides in many format; numerical data and text data; numerical data includes structured data such as set results while; text data include unstructured data such as clinical notes, images. These variations can act in a way that would make it difficult for data analytics systems to analyze and understand the information uniformly. This concerns the structuring of formats, identification codes (including systems for diseases such as ICD), and terms to report in general as well as specialty medicine in order to develop models of prediction that can be applicable in different health facilities [50].

Interoperability Issues: Most of the healthcare facilities have implemented various EHR systems, which do not integrate, or share details with other systems. This lack of interoperability can cause long times for diagnosis and treatment as well as problems in compiling the patient data. While attempts to create interoperable healthcare information exchange languages continue and progress like HL7 FHIR, there are issues surrounding implementation. Even though, such approaches as data analytics and machine learning will give certain suppositions, it is still impossible to use them instead of human experience [51]. Risk models designed display probabilities or risk scores but these need to be understood in the context by those who take care of the patient.

Contextual Knowledge: While data analytics can help use the data to spot out the risks or abnormalities, it is essential to note that an analyst cannot use that information to make clinical decisions without having all the relevant information. For instance, the predictive model might come up with some signals, such as high risk to develop a specific disease, but the healthcare provider will have to take into account other factors such as the patient's lifestyle, other exposures and patient preferences [52].

Clinical Judgment: To date, even with the help of some progressive technologies, such as Artificial Intelligence and Machine Learning, doctors and other qualified specialists must always rely on their clinical reasoning abilities in order to either confirm or modify the evaluation of a client's case. It is important, thus, to consider computer findings in light of the entire picture of the patient. In other words, algorithms may occasionally contain false positive Or false negative





results. There's a vast potential for the use of data analytics in early diseases diagnosis but there are several issues and limitations that have to be addressed in order for the possibility to work. These are; problems of data quality, data privacy and security, the question of whether algorithms used are bias, the problem of standardization and the fact that the data produced is qualitative in nature and will require interpretation by other humans [53]. In order to answer these challenging questions, the sophisticated interaction between healthcare workers, data experts, managers and policymakers and IT specialists shall be provided for a director's AI instruments to be trustworthy, non-biased and used in healthcare without negative ethical implications. Finally as these barriers are being revamped, data analytics is poised to continue to be a major player in disease diagnosis, patient management and enhancement of health care globally.

TREND ANALYSIS OF EARLY DISEASE DIAGNOSIS USING DATA ANALYSIS

There exist grounds to state that Early Disease Alert & Detection is a very active and rapidly evolving field of research and development, which is supported by big data and data analytics, machine learning, artificial intelligence (AI) and many others. Technologies, which are on the cusp of development, bear the potential to revolutionize the health care segment by enhancing the diagnosis in addition to the mode of recovery. This section describes how future advancements are sequentially progressing in disease diagnosis and detection by using data analytics such as artificial intelligence, predictive analytics and wearable, and personalization of medicine, and global health management of health information [54]. AI and ML are to support disease diagnostic in the nearest future and are believed to be a cornerstone to future advances in the field. And as this list of these applied technologies advances further, more and more of them will be able to distinguish subtle intricate details in large unstructured patient databases that may escape real-time observation by the clinician. When the disease parameters are early detected by artificial intelligence the identification processes of diseases will be more efficient. The diagnostic mechanisms associated with conditions of patients will be enhanced and the clinical decisions making associated with patient's conditions will also be more efficient [55].

Deep Learning Models: In particular deep learning the precision of the disease prediction models is anticipated to benefit from it. Taking medical image, genomic data and patients history as examples, huge amount data enables deep learning models to capture the relationship of data and learn patterns for early detection of diseases such as cancers and neurological disorders or heart disease [56]. For instance, deep learning models in radiological images may be used to identify pathologic finding, tumor, fractures, or even brain, lesion accurately.

Natural Language Processing (NLP): Thanks to NLP the likelihood of using unstructured data such as taking from clinical notes, patient's records, and scientific publications is likely to rise in the future. Presently, with the help of NLP one can extract the required details from the text that challenges the physician or scientist and at the same time enhance the integration of the unstructured content with other data base containing formalized data [56]. This will boost the working of disease detection because the doctor will have over view of the patient.

Explainable AI: Thus while AI models are getting and more complex in terms of the depth of data comprehended, there is increased pressure in models to be explainable or at the very least willing to give a hint into the way they are arriving at their conclusions. In future AI systems such techniques of explainable AI will be incorporated; they shall provide the health care givers with all the information as to how an algorithm came up with that particular reasoning of it's a prediction [57]. Thereby increasing confidence in using the AI generated diagnostic tools for diseases and eliminating the misuse of the tools in healthcare organizations. Diagnostic tools shall continue to constitute one of the crucial determinants in as far as probabilities of development of one or another disease in a certain individual before the disease becomes evident shall be concerned. In cases where there are many patients involved predictive analytical models may tell the chance of a patient developing a certain illness in future, thus the doctor treats the patient before the condition becomes worse.

Risk Stratification: Screening tools include those based on hereditary, likely lifestyle, previous diseases or illness record that will help determine who is likely to develop some diseases. For instance, close ones of cancer or cardiovascular Diseases are given the chance to be observed during the beginnings of such [58]. Healthcare systems will remain to extend risk stratification models that serve as base directions in determining which patient may require more of check-ups or preventive measures to be performed on.

Proactive Health Management: There will also be a great improvement on the field of health management by the application of put models of preventive health crises due to chronic diseases. For instance, the big data system can onboard statistics from wearables, EHR, or other data sources to identify if a patient's condition is likely to deteriorate or if the complication is emerging. For instance, diabetic and asthmatic patients will be able to receive event-triggered notification and also have their treatment recommendation change according to their real time individual predicted models, which leads to low hospitalization rates. Smart watch devices, fitness tracking devices and sensors developed into garments will continue to hold the key to early disease detection [59]. These devices are always monitoring and monitoring various health features of a person including the rate at which the heart is pounding, the pressure of the blood, level of glucose, activity level and so on, and these could then be examined for deviations from the standard patterns for the diseases in their early stages.

Real-Time Health Data: Next generation wearables would then feature enhanced and advanced characteristics that address the particular consideration for monitoring and documenting a patient's status. For example, smart watches with a number of sensors in the newer models could already detect an irregular pulse rate or any fluctuations and increase or





decrease in pulse rate that may flag a probable CVD. Similarly, smart watches are allowing diabetics to monitor their glucose levels on devices that they can wear and act appropriately before they worsen [60].

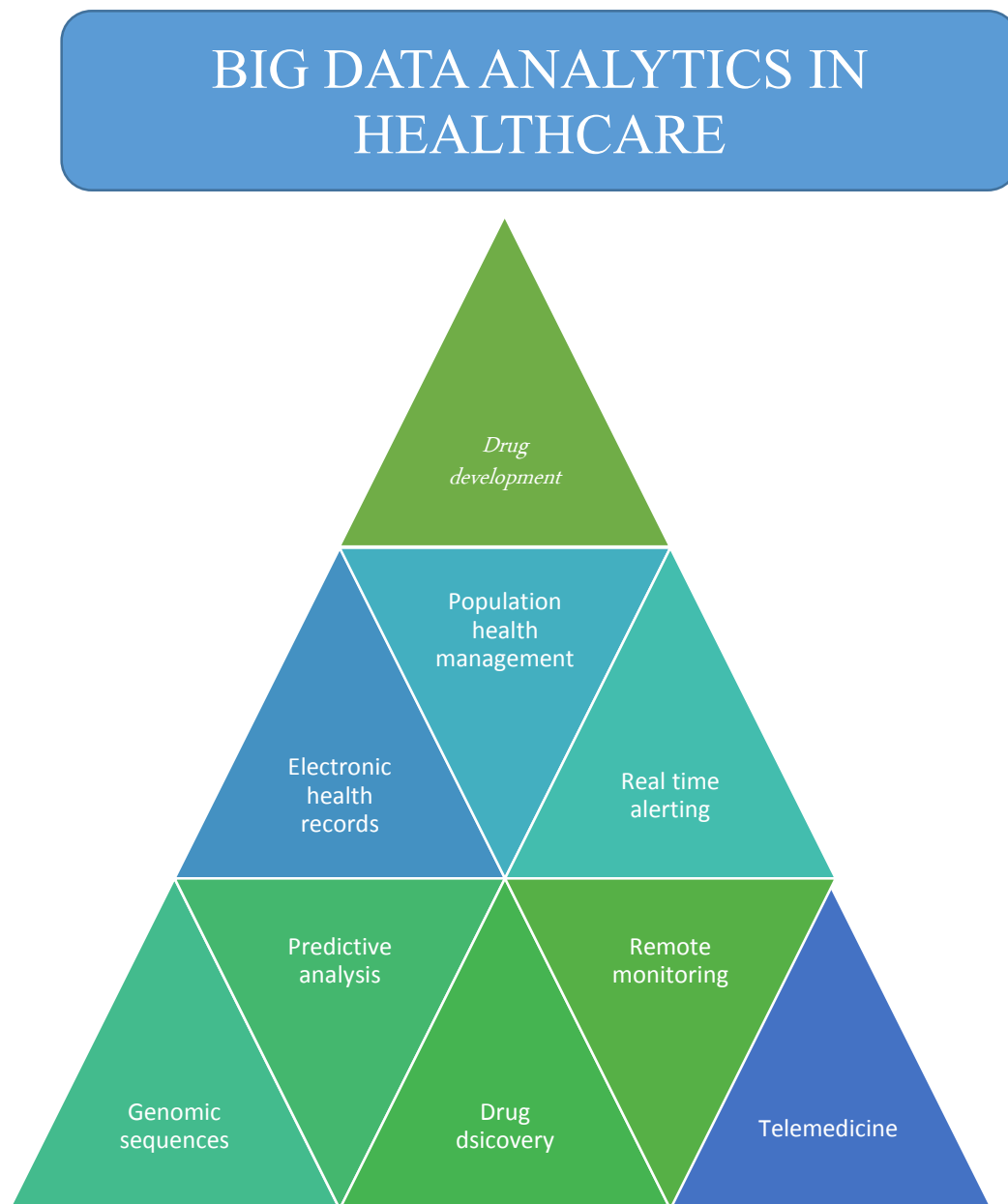


Figure: 4 showing big data analytics in healthcare

Integration with AI and Cloud Computing: The smart clothing of the future will have a direct integration with interfaces and a cloud that can itself tackle the information processed by these devices. Through outsourcing of computing services, there is creation of mass storage of health data and therapists can use the data to supplement clients' check-ups from a distance [61]. Next time data will be collected and algorithms will search for early sings of disease so that it can inform the patient and the doctor that one is ill.

Personalized Health Insights: Wearable's will include opportunity to share the real time health data on Total Herb s-Protein. For instance, a wearable may recommend shift of lifestyle or inform a patient about modifications in his or her exercise routine, stress or sleep patterns. These individualized recommendations assist patients to develop a healthy living and support patient management and thus enhances early disease diagnosis. Once again, the advance of disease screening in early stages will directly depend on the emergence of personalized medicine as its major component [62]. Personalized medicine is when using specific genes, life circumstances, and setting when providing medication. Another drive such as information technology and an advance in genomic such as bioinformatics will one day make for efficient diagnosis of



the disease and its treatment.

Genomic Data Integration: The uptake of genomic data in early signs of diseases will progress further because the cost of genomic sequencing is becoming higher. The study recognizes that through a patient's DNA, healthcare practitioners can ascertain any particular gene abnormalities or tendencies towards particular diseases which can be diagnosed early and accommodate. For instance, a genetic test for mutations in breast cancer risk genes like the BRCA1 and BRCA2 makes it possible to prevent or treat cancer, admitting early signs of breast or ovarian cancer [63].

Targeted Treatment Plans: This will also mean that an individual is likely to get a certain treatment protocol based on the genetic and molecular disposition. For example in oncology people can get individualized based therapies Such as molecular alterations in their cancer cells Thus boosting on the likelihood of success and reducing adverse effects. With help of current genomic, clinical and the accumulated lifestyle current predictive models, and new doctors will have the opportunity to create individualized diagnostics and treatment methods that will have less effect on a patient's body [64]. Disease early detection shall entail benefits from the globalization of healthcare care systems together with the approaches to collaborative researches. For instance, when the doctors upload their gathered health data, and they collaborate with the researchers on researching various topics, they can both enlarge their scope and receive the information on the incidences of the diseases and risky factors in different populations.

Global Data Pools: The proposed work on the formation of the global health data archives will allow the researchers assemble the big, divers, various samples that might represent the population and healthcare systems of the world. This will assist in identifying disease patterns and the factors of risk to health that are not easily distinguishable if the data is from a mere restricted geographic region. For instance, sharing of data in the global setting can provide information on how some factors such as environment or culture impacts disease outcomes and Umber which requires quality health systems intercessions across the world [65].

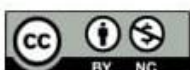
Collaborative AI Research: As AI and machine learning are in their progressive advancement the sharing of such data and models across the institutions will fathom the ability to advance better disease detection algorithms. This leads to the enhanced models, which have the ability to work across different population and different setting that would make early disease diagnosis across the world. Since the prognosis of early diagnosis based on data analysis looks good, there are many technological developments yet to arrive in the near future. AI incorporation with machine learning algorithms would improve the diagnosis model through powerful database combination and wearable technologies and continuous monitoring feed, real-time health data for disease treatment. The further enhancement of completion with genetic profiles, digital personification that allows the creation of extensive therapeutic strategies, as well as the exchange with an international health information contribute to the performance of research work [66]. These technologies are in still in the process of being developed and they will enhance one another to transform healthcare systems not merely by increasing chances of diagnosing diseases at early stages but reducing cost on health, and enhancing the overall quality of life of patient's across the world.

THE ETHICAL IMPLICATION OF EARLY DISEASE DIAGNOSIS: A STUDY OF THE USE OF DATA ANALYTICAL TOOLS

The purpose of this paper is to synthesize literature to develop an answer anticipating increasing ethical concerns when data analytics increases in the healthcare industry with a focus on early disease identification. Despite the fact that these technologies has the potentials of improving the health of patient, it is marked with some of the following ethical dilemmas which has to be well thought through. This section describes its focus on different ethical issues concerning with accurate early disease diagnosis through data analytics as follows: privacy and data ownership of NSF, data sharing, data consent, the algorithm bias issue, fairness in immaterial healthcare delivery, and shifting doctor-patient relationship contours [66]. The major ethical consideration related to data analytics for early disease diagnosis is right to privacy particularly in relation to health information. Today EHR is present, wearables, and health apps, and all of these things generate a massive amount of personal data. The usage of such data should be made appropriately, and secrecy of such information is highly important.

Data Security Risks: The incorporation of digital health tools creates potentiality of risks of unlawful admittance and Cyber-crimes and data break-ins. If the information is breached, health information is good for identity theft, discrimination or it can cause harm if/read by the patient. Especially, we established that communication and storage of patients' information must be safe and encrypted; hence, health institutions, technology companies, and the government have to address this issue [67].

Access Control: Because of privacy of health information, another issue is who in a certain organization or a hospital can get to see patient's information. Whilst the healthcare providers need such information in order to make the right decisions, a patient may not always be sure or wish for others to gain access to the information, let's say, a third party developer or the insurers. Of primary importance to the protection of the patient right to control who can gain access and use health data is that these paper should state clear contractual requirements and consents. The last ethical dilemma which was previously discussed under this category is informed consent. Even the concept of reaching meaningful consent gets entangled with the emergence of AI and machine learning algorithms popularized in the current era together with the use of wearables and self-tracking technology [68]. Patients who are diagnosed with data are generally unaware of how





this data will be used or the consequence of being involved with big data diagnostic processes.

Complexity of Consent: At a psychologically measurable level, patients may have relatively poor awareness of the manner and processes in which they contribute data in their relationships with healthcare organizations. Because of information hiding used in many AI algorithms they can also be called “black boxes” as the user may not know how these particular algorithms reached the final decision. It is just as important, correspondingly, that patient’s privacy is respected and that patients have enough knowledge and understanding of the outcomes of sharing the information with the system [69]. Others are; whether the data will be used for other purposes for instance in research? And how long the data will be stored for?

Dynamic Consent Models: Often setting used for data analytics purposes will be dynamic, meaning that the tools will develop different forms depending on the further possibilities of the training AI models. It is not enough to have a consent model that can only be operationalized at large and inflexible such as consent. Consent must be active and contextual in accord with the specifics of the circumstances. Other ethical principles could need to evolve to dynamic consent to adequately apply the dynamic patient consent model where in patients have to make a new ‘permit’ or ‘refuse’ each time that the technology is improved or each time a new purpose for their information is made. Ethical considerations in the utilization of data analytics include: who owns the control of health data [70]. As with other data generated by patient’s self-driven activities such as using a health app on a smart device, having a fitness tracker or engaging in personal DNA testing, patients may not own the data physically.

Ownership of Data: At present there is debate on who owns information derived from such health gadgets and tests, the patient, or the organizations or companies, collecting that information. Patient data can be used for commerce, or for research representing health systems, scholars, and technology businesses may supply data, capitulate and violate it.

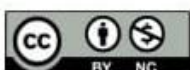
Access to Data: Those who possess webs of control and access to health data is also interpreted to as owning the records in question. It became apparent that while privacy is at risk where those who own such data have no control on how personal information is dealt in, public information has to be used appropriately with the consent of the public. The situation leads to the scenario that, at the currently, there should be clear policies on the use, sharing and Energy Plus commercialization of health data to expound the rights of individuals [71]. Algorithms are major ethical considerations in data analytics largely due the question of health. First disease detecting models and AI algorithms relies on these data which can themselves be prejudiced in terms of demographics of sample, or from certain categories of population available in the existing medical information systems.

Discrimination and Health Inequality: Training with AI creates models with prior knowledge and where bias is provided it can only make matters worse regarding unequal access to healthcare patients. For example, they observed that a model developed by learn from white middle-class groups – source – is far less reliable when diagnosing groups of patients of ethnic minorities, or lower socio-economic status, and consequently a chance for the right diagnosis or timely intervention was missed. This could also widen the ailments of these population and vice brigade parity the discrimination and health status [72].

Transparency and Accountability: As the subject area, early disease detection is a precisely described problem of the AI and data analysis and, depending on the tool and applications used, either the tool itself or its applications must be of the AI class. Transparency can also be used in preventing biases in the models such that they can be corrected if they are seen [73]. Those who have developed such algorithms at the moment need to be answerable for fairness of their models, as well as to test their algorithms on other datasets, prior to their practical use in clinics.

Bias in Data Collection: Another way that bias can appear is in healthcare data management Base on what has been analyzed above. For example, if some population (for instance, poor, rural) is missing in medical trials or health datasets, they will have low accuracy in giving specified population segments. Ethical issues must also ensure that the data collected in health care SYSTEMs addresses and captures all or the vast majority of the following population segments. This makes equity in health another good ethical consideration now that data analytic tools are part and parcel of disease diagnosis. Using diagnostics, wearables and developing AI and other frameworks on early disease signals will be helpful for many from a population health perspective, but can harm equity by widening access to health care facilities if not provided to all population groups [74].

Loss of Trust: Cohost and lack of ability to predict, let alone, diagnose a disease at an early stage, means that the patient loses all credibility of both the application and the doctors who used it. To keep trust and confident into the healthcare system we have to make sure that the doctor patient relationship is at the heart of HC delivery with support of AI and Data Analytics. The application of data analytics for early disease detection poses the following ethical considerations which are explained in the paragraphs below [75]. To achieve the maximum for these technologies and observing the rights of patients no unfairness, the four factors are; privacy, transparency, consent, and equity the health care systems must consider while implementing these technologies. Moreover, several general concerns should be addressed when developing AI in particular; there is no recognizing of bias during the algorithms’ creation, the provision of technologies for all the possible patients, and maintaining the special status of healthcare professionals involved in diagnosing and prescribing. The above presented ethical issues should therefore be taken into account to ensure that efficiency and effectiveness of disease identification achieved through data analysis in the healthcare sector is - for the greater good.





CONCLUSION

Application of the big data analytics on preserving and enhancing the comprehensive disease detection has the possibility of revolutionizing the health care since it offers opportunities for offering high precision in the diagnosis of the diseases, quality health of patients and minimizing the cost of the health care services. But let not ignore the fact that brought this technology came a long some questions ethical question as well. Privacy and confidentiality, informed consent and ownership of data play central roles in helping or hindering patient rights in the total process. Moreover, if the above-stated problem of the algorithmic bias is left unsolved, it can deepen social disparities and produce negative consequences for the amplification of the primary healthcare deficiencies for more vulnerable groups. When it comes to improving the community's ethical considerations, disparities in healthcare endure because it appears that new smart technologies including AI and wearables are going to advance, while some communities will still be left without access to such devices. Considering the haphazard employment of the latest technologies in screening for health risks which can only exacerbate existing Health inequality therefore it has to be made available and accessible especially for the financially incapable in procurement and other pecuniary related comprisals.

One issue goes only to the relationship between doctor and patient. Since several of the tools and data analytics that have been predicated on AI have tremendously enhanced the diagnosing abilities, introduction of any AI tools should not detract from the work of the health care professionals. Patients should not lose confidence in their doctors and other healthcare givers that their care is personalized and that AI platforms which are being used are actually helping with clinical decisions. In the end, it remains impossible to perform an analysis of the early diagnoses of the diseases brought about by big data without considering ethical matters at each stage. Overcoming dilemmas such as data privacy, data consent, data equities, formulaic bias, and the evolution of doctor/patient relationship will be helpful in making pathways for the growth of technology for the combination of power on both sides of the doctor and patient. When effectively adopted, these technologies could improve efficiency in health care delivery and easy, accurate and effective early disease diagnosis for every individual. On balance therefore it is worthwhile to note that as witnesses to the potential of the H4H paradigm and the potential that data analytics holds for re formations in health care, a careful ethical approach will still need to be applied in order to ensure that the re Formations occasioned by these innovations will be the kinds of re Formations that can improve on the positions of positive equity and fairness and that will not compromise the privacy of the patients. If we can confront these ethical considerations, we can advance toward creation of a health care positive in purpose and ethically more acceptable from social, health and fairness perspective that is the intent behind the global call for early detection of diseases towards saving lives and emancipation of needless sufferings from individuals in all the countries of the world.

REFERENCES

1. Wrenn JO, Stein DM, Bakken S, Stetson PD (2010) Quantifying clinical narrative redundancy in an electronic health record. *J Am Med Inform Assoc* 17:49–53
2. McInnes BT, Melton GB, Zhang R, Pakhomov S (2011) Evaluating measures of redundancy in clinical texts. In: *AMIA annual symposium proceedings*
3. Cohen R, Elhadad M, Elhadad N (2013) Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. *BMC Bioinform* 14(1):10
4. Silva HB, Brito P, da Costa JP (2006) A partitional clustering algorithm validated by a clustering tendency index based on graph theory. *Pattern Recognit* 39(5):776–788
5. Dongkuan X, Tian Y (2015) A comprehensive survey of clustering algorithms. *Ann Data Sci* 2(2):165–193
6. Cong S, Han J, Padua DA (2005) Parallel mining of closed sequential patterns. In: Grossman R, Bayardo RJ, Bennett KP (eds) *KDD*. ACM, New York, pp 562–567
7. Ahmed S, Coenen F, Leng PH (2006) Tree-based partitioning of data for association rule mining. *Knowl Inf Syst* 10(3):315–331
8. Ceglar A, Roddick JF (2006) Association mining. *ACM Comput Surv* 38(2):5:1–5:42
9. He J, Rong J, Sun L, Wang H, Zhang Y, Ma J (2018) D-ECG: a dynamic framework for cardiac arrhythmia detection from IoTbased ECGs. In: *International conference on web information systems engineering*. Springer, pp 85–99
10. Ma J, Sun L, Wang H, Zhang Y, Aickelin U (2016) Supervised anomaly detection in uncertain pseudoperiodic data streams. *ACM Trans Internet Technol (TOIT)* 16(1):4–91
11. NaliniPriya G, Kannan A, AnandhaKumar P (2012) A knowledgeable decision tree classification model for multivariate heart disease data-A boon to healthcare. In: Li Z, Li X, Liu Y, Cai Z (eds) *ISICA*, vol 316. Communications in Computer and Information Science. Springer, Berlin, pp 459–467
12. Alizadehsani R, Habibi J, Bahadorian B, Mashayekhi H, Ghandeharioun A, Boghrati R, Sani ZA (2012) Diagnosis of coronary arteries stenosis using data mining. *J Med Signals Sens* 2(3):153–159





10. Alizadehsani R, Habibi J, Sani ZA, Mashayekhi H, Boghrati R, Ghandeharioun A, Khozeimeh F, Alizadeh-Sani F (2013) Diagnosing coronary artery disease via data mining algorithms by considering laboratory and echocardiography features. *Res Cardiovasc Med* 2(3):133–139
11. Karaolis MA, Moutiris JA, Hadjipanayi D, Pattichis CS (2010) Assessment of the risk factors of coronary heart events based on data mining with decision trees. *IEEE Trans Inf Technol Biomed* 14(3):559–566
12. Kunjunnair AP (2012) Erratum to: “Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules and decision tree rules”. *Cent Eur J Comput Sci* 2(1):86
13. Husain W, Adnan MHM, Rashid NA (2012) Hybrid approaches using decision tree, Naive Bayes, means and euclidean distances for childhood obesity prediction. *Int J Softw Eng Appl* 6(3):99–106
14. Shirazi SH, Umar AI, Haq N, Naz S, Razzak MI, Zaib A (2018) Extreme learning machine based microscopic red blood cells classification. *Cluster Comput* 21(1):691–701
15. Huang YC (2013) Mining association rules between abnormal health examination results and outpatient medical records. *Health Inf Manag J* 42(2):23–30
16. Naseer A, Rani M, Naz S, Razzak MI, Imran M, Xu G (2019) Refining Parkinson’s neurological disorder identification through deep transfer learning. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-019-04069-0>
17. Kunjunnair AP (2015) Classification of Parkinsons disease using data mining techniques. *J Parkinsons Dis Alzheimer Dis* 2(1):4 138. Suganya P, Sumathi CP (2014) A novel metaheuristic data mining algorithm for the detection and classification of Parkinson disease. *Indian J Sci Technol* 8:1
18. Tsao C-W, Yeh J-Y, Wu T-H (2011) Using data mining techniques to predict hospitalization of hemodialysis patients. *Decis Support Syst* 50(2):439–448
19. Razzak MI, Naz S (2017) Microscopic blood smear segmentation and classification using deep contour aware cnn and extreme machine learning. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 801–807
20. Patil HK, Seshadri R (2014) Big data security and privacy issues in healthcare. In: 2014 IEEE international congress on big data, pp 762–765
21. Slawson DL, Fitzgerald N, Morgan KT (2013) Position of the academy of nutrition and dietetics: the role of nutrition in health promotion and chronic disease prevention. *J Acad Nutr Diet* 113(7):972–979
22. World Health Organization (1990) Diet, nutrition, and the prevention of chronic diseases. Report of a WHO Study Group. Diet, nutrition, and the prevention of chronic diseases. Report of a WHO Study Group, p 797
23. Willett WC, Koplan JP, Nugent R, Dusenbury C, Puska P, Gaziano TA (2006) Prevention of chronic disease by means of diet and lifestyle changes. *Disease Control Priorities in Developing Countries*, pp 833–850 11. Policy brief on ageing no. 3, older persons as consumers (2009)
24. Policy brief on ageing no. 6, health promotion and disease prevention (2010)
25. Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R, Riley WT (2011) Health behavior models in the age of mobile interventions: are our theories up to the task? *Transl Behav Med* 1(1):57–71
26. Griffiths F, Munday S, Friede T, Stables D, and Holt TA, Thorogood M (2010) Automated electronic reminders to facilitate primary cardiovascular disease prevention: randomised controlled trial. *Br J Gen Pract* 60(573):137–143
27. Dasah JB, Kuranchie P, Amoah AGB, Adjei DN, Agyemang C (2015) The effect of electronic reminders on risk management among diabetic patients in low resourced settings. *J Diabetes Complicat* 29(6):818–821
28. Waqialla M, Alshammari R, Razzak MI (2015) an ontology for remote monitoring of cardiac implantable electronic devices. In: 2015 international conference on computer, communications, and control technology (I4CT). IEEE, pp 520–523
29. Waqialla M, Razzak MI (2016) an ontology-based framework aiming to support cardiac rehabilitation program. *Procedia Comput Sci* 96:23–32
30. Jiang Y, Shepherd M, Maddison R, Carter K, Cutfield R, McNamara C, Khanolkar M, Murphy Dobson R, Whittaker R (2016) Text message-based diabetes self-management support (SMS4BG): study protocol for a randomised controlled trial. *Trials* 17:179 *Neural Computing and Applications* (2020) 32:4417–4451 4445 123
31. Alaleh Z, Hollmann Markus W, Frits H, Benedikt P, Jeroen H, Polderman Jorinde AW, de Groot FA (2016) An automated reminder for perioperative glucose regulation improves protocol compliance. *Diabetes Res Clin Pract* 116:80–82
32. Fischer HH, Fischer IP, Pereira RI, Furniss AL, Rozwadowski JM, Moore SL, Durfee MJ, Raghunath SG, Tsai AG, Havranek EP (2016) Text message support for weight loss in patients with prediabetes: a randomized clinical trial. *Diabetes Care* 39:1364–1370





33. Daghistani T, Al Shammari R, Razzak MI (2015) Discovering diabetes complications: an ontology based model. *Acta Inform Med* 23(6):385
34. Gerber BS, Stolley MR, Thompson AL, Sharp LK, Fitzgibbon ML (2009) Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. *Health Inform J* 15(1):17–25
35. RJ, Block TJ, Hopkins D, Carpenter HA, Dolginsky MS, Hudes ML, Palaniappan LP, Block CH, Block G, Azar KM (2015) Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. *J Med Internet Res* 17(10):e240
36. Steinhubl S, Kim S, Bae WK, Han JS, Kim JH, Lee K, Kim MJ, Kim JY, Oh S (2015) Effectiveness of 6 months of tailored text message reminders for obese male participants in a worksite weight loss program: randomized controlled trial. *JMIR mHealth uHealth* 3(1):e14
37. Morgan P, Callister R, Collins C, Hutchesson MJ, Tan CY (2015) Enhancement of self-monitoring in a web-based weight loss program by extra individualized feedback and reminders: randomized trial. *J Med Internet Res* 18(4):e82
38. O'Grady JS, Thacher TD, Chaudhry R (2013) the effect of an automated clinical reminder on weight loss in primary care. *J Am Board Fam Med* 26(6):745–750
39. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M (2015) Mobile health devices as tools for worldwide cardiovascular risk reduction and disease management. *Circulation* 132:2012–2027
40. Moreland EC, Volkeneing LK, Lawlor MT, Chalmers KA, Anderson BJ, Laffel LMB (2006) Use of a blood glucose monitoring manual to enhance monitoring adherence in adults with diabetes: a randomized controlled trial. *Arch Int Med* 166(6):689–695
41. Bonato P, Chan L, Patel S, Park H, Rodgers M (2012) A review of wearable sensors and systems with application in rehabilitation. *J NeuroEng Rehabil* 9:21
42. Zhou H, Hu H, Harris ND (2006) Wearable inertial sensors for arm motion tracking in home-based rehabilitation. In: Arai T, Pfeifer R, Balch TR, Yokoi H (eds) *IAS*. IOS Press, Amsterdam, pp 930–937
43. Ahmed E, Yaqoob I, Hashem IAT, Khan I, Ahmed AIA, Imran M, Vasilakos AV (2017) The role of big data analytics in internet of things. *Comput Netw* 129:459–471
44. Augustyniak P (2011) Personal wearable monitor of the heart rate variability. *Bio-Algorithms Med-Syst* 7(1):5–10
45. Para'k J, Tarniceriu A, Renevey P, Bertschi M, Delgado-Gonzalo R, Korhonen I (2015) Evaluation of the beat-to-beat detection accuracy of pulse on wearable optical heart rate monitor. In: *EMBC. IEEE*, pp 8099–8102
46. Villalba E, Salvi D, Ottaviano M, Peinado I, Teresa AM, Akay A (2009) Wearable and mobile system to manage remotely heart failure. *IEEE Trans Inf Technol Biomed* 13(6):990–996
47. Zhang Z, Zheng J, Ha C (2016) Design and evaluation of a ubiquitous chest-worn cardiopulmonary monitoring system for healthcare application: a pilot study. *Med Biol Eng Comput* 55:283–294
48. Wang Z, Jiang H, Yang K, Zhang L, Wei J, Li F, Chi B, Zhang C, Wu S, Lin Q, Jia W (2013) Lifetime tracing of cardiopulmonary sounds with ultra-low-power sound sensor stick connected to wireless mobile network. In: *NEWCAS. IEEE*, pp 1–4
49. Koo HS, Michaelson D, Teel K, Kim D-J, Park H, Park M (2016) Design preferences on wearable e-nose systems for diabetes. *Int J Cloth Sci Technol* 28(2):216–232
50. Hosseini V (2015) Algorithm and related application for smart wearable devices to reduce the risk of death and brain damage in diabetic coma. *J Diabetes Sci Technol* 10(3):802–803
51. Esser P, Steins D, Dawes H, Collett J (2014) Wearable accelerometry-based technology capable of assessing functional activities in neurological populations in community settings: a systematic review. *J Neuroeng Rehabil* 11:36
52. Wu X, Zhu X, Wu G-Q, Ding W (2014) Data mining with big data. *IEEE Trans Knowl Data Eng* 26(1):97–107
53. Joshi S, Nair MK (2015) Prediction of heart disease using classification based data mining techniques. Springer, New Delhi, pp 503–511
54. Smitha T, Sundaram V (2012) Classification rules by decision tree for disease prediction. *Int J Comput Appl* 43(8):6–12
55. Anooj PKN (2011) Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules and decision tree rules. *Cent Eur J Comput Sci* 1(4):482–498
56. Persi Pamela I, Gayathri P, Jaisankar N (2013) A fuzzy optimization technique for the prediction of coronary heart disease using decision tree. *Int J Eng Technol* 5(3):2506–2514





57. Liu X, Fu H (2014) PSO-based support vector machine with cuckoo search technique for clinical disease diagnoses. *Sci World J*. <https://doi.org/10.1155/2014/548483>
58. Anto S, Chandramathi S, Aishwarya S (2016) An expert system based on LS-SVM and simulated annealing for the diagnosis of diabetes disease. *IJICT* 9(1):88–100
59. Vadicherla D, Sonawane S (2013) Decision support system for heart disease based on sequential minimal optimization in support vector machine. *Int J Eng Sci Emerg Technol* 2(2):19–26
60. Subbalakshmi G, Ramesh K, Rao MC (2011) Decision support in heart disease prediction system using Naive Bayes. *Indian J Comput Sci Eng (IJCSE)* 2(2):170–176
61. Tseng W-T, Chiang W-F, Liu S-Y, Roan J, Lin C-N (2015) The application of data mining techniques to oral cancer prognosis. *J Med Syst* 39(5):1–7
62. Khan SU (2015) Classification of Parkinsons disease using data mining techniques. *Parkinsons Dis Alzheimer Dis* 2:4
63. Kong L, Wiederhold BK, Gao K, Wiederhold MD (2013) Clinical experiment to assess effectiveness of virtual reality teen smoking cessation program. *Stud Health Technol Inform* 19:58–62
64. Luo L, Xie Y, Zhang Z, Li W-J (2015) Support matrix machines. In: *International conference on machine learning*, pp 938–947
65. Zheng Q, Zhu F, Qin J, Chen B, Heng P-A (2018) sparse support matrix machine. *Pattern Recognit* 76:715–726
66. Razzak MI, Naz S, Zaib A (2018) Deep learning for medical image processing: overview, challenges and the future. In: *Classification in BioApps*. Springer, Cham, pp 323–350
67. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E (2015) Deep learning applications and challenges in big data analytics. *J Big Data* 2(1):1
68. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
69. Phan NH, Dou D, Wang H, Kil D, Piniewski B (2015) Ontology-based deep learning for human behavior prediction in health social networks. In: *BCB*. ACM, pp 433–442
70. Razzak I, Imran M, Xu G (2018) efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks. *IEEE J Biomed Health Inform*. <https://doi.org/10.1109/JBHI.2018.2874033>
71. Langley P, Sage S (2013) Induction of selective Bayesian classifiers. *CoRR*, arXiv: 1302.6828
72. Boule' M (2007) Compression-based averaging of selective Naive Bayes classifiers. *J Mach Learn Res* 8:1659–1685
73. Provost F (2000) Well-trained PETs: improving probability estimation trees, CDER working paper #00-04-is, Stern School of Business, NYU, NY, 10012
74. Boule' M (2006) Regularization and averaging of the selective Naive Bayes classifier. In: *IJCNN*. IEEE, pp 1680–1688
75. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D (2007) Top 10 algorithms in data mining. *Knowl Inf Syst* 14(1):1–37

