Investigating Head &Neck Cancer Predictions Using Big Data: An Extensive Review

Abstract

Big data analytics is a cutting-edge method used to examine massive structured as well as unstructured data. It is applied to large data sets that are impossible for traditional data systems to be analyzed. Data are not only vast but also complex and variable, so it becomes very important to sort all of them so that it can be properly analysed. Big data analytics is implanted in many sectors like business, marketing, retail, research, automobiles, healthcare, etc. Big data analytics is crucial for the healthcare sector, from data collection to the development of new treatments. It also plays vital role in cancer care Description, Prediction and Prescription. Big data analytics technology is highly supportive in cancer care. Cancer was always a subject of research around the world, and the implementation of new technologies in it surely has revolutionized it. This survey aims to offer insights into the use of big data analytics in the health industry where the enormous data has been captured using various wireless sensors. Wireless biomedical sensors are smart devices that collect data related to monitoring patients' health and make them accessible remotely .It will also offer in-depth knowledge regarding the use of big data to the treatment of Head and Neck cancer (HNC). We illustrate the importance of our work through in depth literature Survey which emphasise on the contribution of recent research on big data in health care using latest wireless technologies. Further an elementary taxonomy for the HNC analytics is tabulated. We have also demonstrated graphically the percentage population of

Authors

Neetu Settia
Assistant Professor
Guru Tegh Bahadur Institute of Technology
New Delhi, India
Neetu_aug3@yahoo.co.in

Dr. Gagandeep Kaur
Assistant Professor
Guru Tegh Bahadur Institute of Technology
New Delhi, India
arora_gagan07@yahoo.co.in

Shreya Varshney
Student
Guru Tegh Bahadur Institute of Technology
New Delhi, India
ShreyaVarshney1402@gmail.com

Monica Bhutani
Assistant Professor
Bharati Vidyapeeth College of Engineering
New Delhi
monica.bhutani@bharatividyapeeth.edu

patient’s reached III-IV stage of HNC as well as the risk factors for the cancer outbursts.

Keyword: Big Data Analytics, Healthcare, Cancer, Genomics, Security, Application

I. INTRODUCTION

Big Data has become the topic of interest from the past few decades mainly because of its great potential and its flexibility to work for any industries. It has promoted the growth and development of many sectors that are blended with it. Same goes for healthcare sectors .In the past few decades, healthcare industries have obtained its massive growth in which big data performed a critical role. Digital technologies like electronic healthcare records, patient's portals, payer records etc. has made work easier for data collection. Methods like Association Rule Learning, Machine Learning, Genetic algorithms, Sentimental Analysis, Artificial Intelligence (AI), Social Network Analysis, and Regression Analysis are used to analyze data [1]. As the importance of data will increase, volume will also increase. According to the prediction of International Data Corporation, around 163 zetabytes of data will be produced till 2025 [2].Just like new trends continually replace the previous ones, recent trends in healthcare are shifting from treatment to prevention as a result of the growth in data volume. There is a constant improvement in healthcare by reducing curing cost and better drug delivery [3].Analytical models such as statistical, contextual, quantitative, etc., are applied to healthcare data and associated fields of awareness to assist factual decision-making for healthcare [4]. Big data carries a huge potential not only today but also in future. In the future of healthcare industries, predication analysis will be the reason of next revolution in medicine and statistics [6]. Big data can be used to determine pneumonia from chest X - rays and fundoscopy images can be used for diabetic retinopathy [7].

This survey focuses on one of the biggest problems in healthcare- Cancer. Cancers are grouped in one of the most complex diseases with high mortality rate. Cancer cells can develop in any part of the human body; the cancer discussed in this study is Head and Neck cancers. Squamous cell carcinoma is the most prevalent history in the Head and Neck Cancer (HNC), which is malignancies of the upper aero digestive tract that mostly affect the Oral Cavity, Larynx, Oropharynx, Nasopharynx and Hypopharynx [8]. HNC is one of the most complex and challenging maligned cancer, mainly because of diverse tumour subgroups that react to therapy differently, risk of recurrence, late detection, comorbidities, and challenging anatomical areas that result in negative outcomes [9]. The best course treatment for HNC patients is frequently determined by data on survival outcomes. The literature contains reports of HNC outcomes from either demographics or specific hospital-based studies [10].

According to the study, around one–third of HNC cancers are detected in early stages, i.e., I and II stages [11]. Despite of development of advance surgery and treatment methods around 30-50 percent of stage III to IV HNC patients experience a relapse after 24 months of treatment, salvage treatments are unsuccessful in most of the cases [12- 13]. The National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) project efficiently gathers details on cancer occurrence and survivorship from population-based cancer registries (NCI) [14]. Big data analysis has put a huge impact on HNC treatment. HNC is 6th deadliest tumour around the world with 700,000 newly detected cases with 350,000 deaths annually [15].

The intent of this survey is to emphasise how big data analytics are employed in the treatment of cancer. Some of the significant contributions of this survey is as follows;

1. This survey comprehensively reviews in depth knowledge about big data analytics in healthcare and examines the current research in the field. (Section 2)
2. It discusses impact of big data analytics in health industries where origin of data that is used for study in cancer is explained. It covers all the 6v's, big data storage and processing systems which are briefly explained. Further an elementary taxonomy for the HNC analytics is tabulated. (Section 3)
3. Additionally, it examines head and neck cancer, including potential causes, implementation of big data analytics therein, and an explanation of the field's current state.(Section 4)
4. Further it is demonstrated graphically the percentage population of patient's reached III-IV stage of HNC based on pre-existing conditions namely Oral cavity, Oropharynx, Hypopharynx and Larynx. Also the Patient's percentage and risk factors for the cancer outbursts based on smoking, drinking, Oral hygiene and family history is shown graphically.(Section 5)
5. Further it provides insights of application of big data analytics. (Section 6)
6. Finally benefits, conclusion and future research is discussed.(Section 7-8)

II. LITERATURE SURVEY

Table 1: Literature survey shows contribution of various authors towards Big Data in Healthcare

S.NO	AUTHOR, YEAR	CONTRIBUTION
1.	2023,Hassani, Sahar, and Ulrike Dackermann [20]	Highlighting on the use of Big Data for History analysis to improve medical and healthcare services.
2.	2021, Tieˆn-Dung Ha`1 et al.[21]	Work toward finding big data as new approach towards biology and allows highlighting the use of big data for oncologist to understand and diagnose cancer.
3.	2020, Shah Nazirwt al.[19]	This paper studies crucial big data aspects in healthcare sector for efficient management.
4.	2020, Stefano Cavalieri et al.[22]	This article gathers and presents data on head and neck squamous cell carcinoma (HNSCC) and uses machine learning methods to explore the data.
5.	2020, Marta Bogowicz et al.[23]	This study focuses on combination of distributed learning and radiomics.
6.	2020, Md. IleasPramanik et al.[24]	Summarizing the key concepts of healthcare informatics and analytics in big data and analyzing it from its beginning till now
7.	2019, Sunil Kumar and Manindersingh [25]	Using a variety of tools from the Hadoop ecosystem to discuss the impact of big data on healthcare services.
8.	2019, Chiaojung Jillian Tsai et al.[27]	Outlines briefly how big data can be applied to cancer research and how knowledge gleaned from it can help in cancer detection.
9.	2019, Loredana G. Marcu et al.[28]	This article aims to discuss the state of art in big data utilization for customized therapy in head and neck cancers based on CT and PET imaging modalities.

10.	2019, Stefan M. Willems et al.[29]	The paper discusses the primary sources of big data particularly in (head and neck) oncology. It also mentions the necessity of combining multiple clinical, pathological, and quality-of- life data sources.
11.	2019, Fairuz Amalina et al.[46]	This paper discusses big data analytics as a method for dealing with complicated and unstructured data issues utilizing tools like Hadoop, Spark, and MapReduce.
12.	2018, Cary Jo R. Schlick et al.[47]	Analyzing computing technology to assist large amount of big data and its usage in healthcare sectors.
13.	2018, Carlo Resteghini et al.[63] .	Understanding how big data is applied in the field of head and neck oncology and how big data analysis is conducted.
14.	2014, M.S. Tiwana et al.[64]	This study focuses on providing major, long- term consequences of HNC that are site-specific.

1. **Big Data Analytics:** Data is a powerful resource that can be found in different ways. Big data means massive amount of both structured as well as unstructured data that shows an exponential increase in graph in different sectors with different complexities. Therefore, Big data analysis refers to analyzing and utilizing the growth of data properly. According to IBM [16] Big Data Analytics defines big data in terms of size that is far beyond its traditional database. Big data is not just about its volume but also its power. Big data is sometimes mistaken for simply being a lot of data, but it is much more than that. Data complexity rises as a result of the constant generation of new properties that make them distinct [17]. Hermon [3] proposed that massive is to be processed using data analytics to extract important information. Big data analytics can bring revolution to any sector that blends with it. Big data analytics focuses on providing study of raw data to extract information from it [18].

III. OVERVIEW

Big data consist of large volume of data with much complexity, high velocity and variability that advance the methods and technology to collect, capture, control and analyse information. Big data is the data whose scales, complexity and diversity needs new algorithm, technology and structure for visualization and management to provide easy and efficient solution in health care, commerce, and other fields . Big data is broadly divides into six categories as shown in figure 1. 6 V's of big data are listed below.

- **Velocity:** It is an analysing process speed. It plays a crucial role in live streaming and processing real time data. This includes social network, audio-video, map visualization, transactions etc. Traditional methods are not enough to analyse new generation's big data.
- **Volume:** It is the huge amount of data that is extracted and process from different applications and many social media platforms like Google, Games, Netflix, and Facebook etc. It is predicted that 2.5 quintillion bytes of data is created everyday which will continue to increase in near future. To handle, store and process this amount of data has become challenge for data analysist.
- **Variety:** It is the data type that is obtained from various sources and platforms. This includes data from corporate sources, sensors, mobile devices, social networks.

- **Value:** It is considered as an important angle for big data measurement by determining data that the data is useful or not. But performing the task can be challenging because of the current complexities of big data.
- **Veracity:** It is the quality of data that is to be analysed. During the investigation, data could be missing, corrupted, noisy, or dirty. Therefore, it is necessary to discard and unrelated data which becomes a difficult task in case of large volume of data.
- **Complexity:** It refers to the challenges that occurs during the processing of data that includes finding interconnections among data from different sources [26].

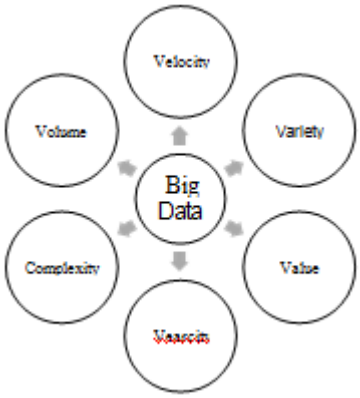

Figure 1: 6V's of Big Data

Big Data Storage System: Large amounts of data, or "big data," are stored, managed, and retrieved using a form of storage system called "big data storage." It enables sorting and storage in a manner that it can be easily in a manner that can be easily accessed by serious applications working on big data. Storage technologies now a days are only restricted to manage and store because of its large volume. Traditionally RDBMS are used to cope up with structural data. But this traditional system cannot be used to store and manage big data. Therefore, reliable source is required to collect and manage massive data. System like Hadoop Distributed File System (HDFS) [30], Open stack swift [31], and Google File System (GFS) [32] have been suggested for handling big data. According to IBM [33] modern device (terminals, computes) connected themselves to storage devices either through network or directly. Users tell computes how to access data from various storage devices system like HDFS, GFS, and Open Stack Systems are made up of network connected. Storage devices that offer scalability, virtualization, and distribution to handle large amount of data efficiently [34].

Processing Systems: As a complexity of data increases, the volume of data too increases with it. Thus, to store and manage massive amounts of data, processing systems were proposed to make enormous data meaningful as well as useful. For Example: data scientist utilizes processing systems for machine learning and data mining, data intensive text processing, large scale social media analysis, assembly of large genomes. Platforms used for big data processing systems are Spark [35], Samza [36], Hadoop [37], Flink [38], and Storm [39]. These all platforms are Apache software based open-source projects [40]. These platforms under Apache software foundation contain three layers mainly: Cluster manager, processing engine, and processing frame form [41]. Processing system put forward distinct

characteristics that can be accommodated in different big data scenarios. Flink and Spark can be used for batching and streaming; Samza and Storm can be used for string processing; Hadoop can be used for batch processing. System processing is done on continuous data which must be rapidly processed. Processing in batches is done in the cases where the data is stored in gigantic files. Data analysis is one of the crucial steps for organization to develop their business steps and marketing strategies by changing raw data into human-readable format (documents charts and graphs) [42]. Processing enormous amount of data necessities, a very high-performance computing environment i.e., simple to operate and can be tuned with linear scalability. Processing of big data contains several sub-stages. At every stage data processing and transformation takes place to produce outputs [43]. An Elementary taxonomy for HNC Analytics given in Table 2 explains the classification based on the level of data collection for HNC patients, their analysis using various data processing tools followed by the prescription.

Table 2: An Elementary Taxonomy for HNC Analytics

Head and Neck Cancer Analytics			
	Description	Prediction	Prescription
Questions	What happened? What is happening?	What will happen? Why will it happen?	What should I do? Why should I do?
Enablers	Public health care Statistics Socio Economics Data Medical Equipment Data	Data Mining Data Intensive text Processing Social Media Analysis	Optimisation Simulation Analysis
Outcomes	Well defined methods to collect, manage and analyse Information	Accurate projection	Assemble of large genomics

Application: Building a real-time huge data system that works effectively presents several challenges, including real-time event, data transfer and data collection, information finding, analytics, decision making, and answers to design an effective application. Effective and efficient development strategy reduces risk and improves quality of life [44]. Big data has a crucial role in developing business strategy and enhancing overall compound interest of the users. Thus, big data came as a dominating field across the globe. According to a survey performed by Wikibon, big data generates revenue in the market with increase from $42B in 2018 to $1030B in 2027 attaining the overall growth rate of 10.48%. Industries and others sectors are now switching towards big data application to help uncover market trends, custom preferences, hidden patterns, unknown correlation and many more. Big data has a very important part in various fields namely manufacturing, media entertainment, healthcare, government, and IoT using wireless sensors [45]. There are many sectors that are still untouched by big data. Soon big data will surely influence them by performing output forecasting in the most efficient way they are now. It will surely put a great impact on the users around the world.

IV. BIG DATA FOOTPRINTS IN HEALTH INDUSTRY

Big data on its own is a great system, it cannot be explained in few words. It is one of those sectors which has great potential associated with it. It can bring great changes towards prediction of accurate information and patient diagnosis in healthcare sector [46]. Big data is frequently used in medical education to improve the profession of medicine. Diversities are always observed in medical health related data such as Biomedical signals, Electronic Health Records (EHRs), Sensing data using wireless sensors, Biomedical Images, social media, and Clinical test. Therefore, major uses of big data related to health sectors can be named as: Clinical Decision Assistance, Public health management, Disease surveillance, Epidemic control, etc. [48]. This inclusion of big data in healthcare sectors give positive results and came out with many benefits such as Early detection of diseases with smart healthcare systems through big data analytics (EHRs, Fault detection, Mobile health, Cost efficiencies.) [49]. Future of medical industry enormously depends upon big data and its analytics. More there will be efficiency in managing big data more efficiently medical industry will work.
Big data analytics aids medical professionals to diagnose patients and can provide accurate treatments for the same. Collecting and processing huge amount of data helps data analysts develop a track for patients that will directly put a great impact upon healthcare systems. Frameworks regarding the impacts as shown in figure 2 are:

- **Right Providers:** Healthcare professionals contribute significantly to the system by gathering and organising data (public healthcare statistics, socio-economic data, and medical equipment). Processing of such data can help in development of technologies of treating patients in efficient way [48-50].
- **Right Value:** Healthcare workers must pay attention towards their patients whether they are getting right treatment or not. This can be done by checking and removing waste, destroying misinterpretation, and improving resources reading patients recovery [50-51].
- **Right Living:** It refers how an individual can work upon themselves by managing their health and well-being. When an individual is in their best state, they can make decision and make better choices by choosing right tracks such as preventive care, exercise, diet, etc. [48-52].
- **Right Innovation:** It refers to how healthcare workers and researchers can recognise new diseases and develop cure for the same.
- **Right Care:** This track makes sure that the patients are receiving right treatment from the available data [53].

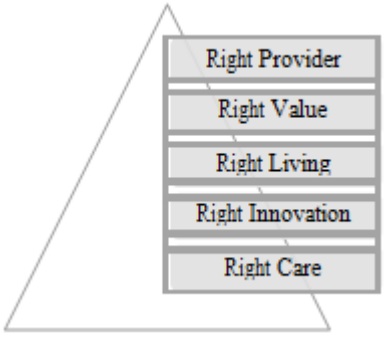

Figure 2: Framework of Impact of Big Data on Healthcare

Objectives: Big data analytics does not only revolve arounds storing, processing, and protecting data but also helps in life saving causes such as cancer. Cancer on its own is a critical disease, hence extra ordinally integrate. To treat cancer patients, one needs to go through enormous amount of previous data which becomes a difficult task for doctors. In order to speed up the discovery of more efficient treatments for this disease, researchers are creating big data analytics approaches that could speedily extract pertinent awareness from cancer data. [54]. Data is been studied and analyse to develop new treatment and therapies that can lead to more accurate treatments. With the help of big data analytics cancer recursion, response, and progress prediction is no longer a difficult task. Thus, help in increasing the pace through accurate treatment, hence, saving life of many [55].

Sources of Big Data in Cancer Genomics

- **Commercial and Private Cancer:** Cancer related big data analytics differ from other big data analytics. It solely depends upon cancer related information and real time data collected. It generates feedback and helps in giving accurate prediction of the same. Several businesses have started to gather and evaluate real-time data from clinical notes, billing data, and medical records to provide feedback and predictions for cancer care. One such use is the cancer cloud that Flatiron Health has launched. There are 250 cancer clinics in the Flatiron Health network, and there are 1.5 million active patients [56].Additionally, two cases of health support organisations and private zone databases are MarketScan by Truven Health and the Kaiser Permanente Clinical Research Network [57].

- **National Population – Based Cancer Database:** Major data related to cancer research is collected by National cancer institute (NCI) through Surveillance, Epidemiology and End results (SEER) program and National Program for cancer Registration (NPCR). SEER fetch and issues cancer cases and surveillance data from public cancer entries that represents more than 30 % of the public in United States. The information gathered by SEER are: Cancer characteristics (Tumour cell types, some biomarker and genomic details on tumours, biological and clinical aspects), patient's demographics (gender, age, birthplace, race, and ethnicity),stages of disease (I, II, III,IV), patient's outcomes (Viral status and cause of death) and treatment details(operations, radiation, chemotherapy etc.) [58]. Most of the data capturing of US population is done by SEER, NCPR. With the help of this data researchers all around the world can evaluate cancer trends for treatment and prevention at the right time. Merging data and analysing it, helps the researchers to develop data sheets to track the risk of occurring, recurring, and dying of the patients through cancer further developing treatment and preventive measures of the people around the world.

- **Database for the Cancer Genomics and the other- 'omics':** In 2018 NCI's Centre for Cancer Research established Cancer Data Science laboratory for producing computational algorithm for integrating and studying data for patients and cancer omics laboratory [59]. The Cancer Genome Atlas (TCGA), an association between the National Cancer Institute and National Human Genomics Research Institute is amongst the largest publicly accessible and collaborative genomic dataset. In 2019, TCGA has listed and genomically characterised tumours for more than 30,000 people with almost 100 distinct kind of cancer by exploring large number of genes and

tracking down 3,150,000 mutants. TCGA generated enamours data and is used all over the world for recent cancer studies. For investigating transcriptomes and genomes, it employs a multi-omics approach and several techniques [61]. Data collected from different sources have help researchers globally to have an efficient approach toward cancer. Analysing data and fetching relevant data information through it, has made the works of clinical a lot easier as the data compiled is much easier to study that that of raw ones.

V. HEAD &NECK CANCER (HNC)

Head and Neck Carcinoma refer to variety of cancerous tumours that grow around the larynx, throat, nose, mouth, and sinuses. They are mostly squamous cell carcinoma. These types of cancer generally start from flat squamous cell which makes thin layers of tissue on the surface of Head and Neck structures. According to a survey, it shows that in USA, Head and Neck accounts for around 4% of all the cancers. Evaluation shows that in 2023 around 66,928 people were diagnosed with head and neck cancer [76]. Head and neck carcinoma is one of the most critical tumours around the world with the survival rate around 50%. From the last few decades there is a constant improvement in treatment methods but they are still far away from increasing survival rates. The purpose of using big data analytics along with wireless technology in medical industries is prediction analysis by making accurate prediction towards prognosis of diseases. Research on head and neck carcinoma through big data analytics is very limited as shown in figure 3, however there are constant growth in treatment and medical field that will continue to grow in next few decades by improving survival rate.

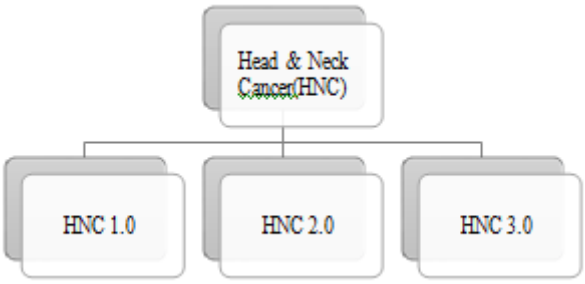

Figure 3: Overview of Head & Neck Cancer

1. **HNC 1.0:** Resteghini, C et al. [63] proposed the study about Head & Neck Cancer and its related forms. The study was done on patients by giving various questionnaire and medical data was gathered through an online case report that was named Open Clinical Platform. All the Datasets regarding their treatment, pathologic features, follow ups, and toxicity of Head & Neck squamous cell carcinoma (HNSCC) patients had been collected. This study included 1537 HNSCC patients having III-IV stages. From which 1086 i.e., 70% was retrieved between2008-14and rest of the 451 cases was retrieved between2015-17. Data was collected from following countries: 670-43% from Dutch, 151-10% from German, and 716-47% from Italian Cancer centres. In this study 1096 i.e., 71% were male patients. The study population was made up of (41%-624) Oro Pharynx Cancer

patients, (28%-429) oral cavity cancer patients, (20%-314) Larynx Cancer patients, and (11%-170) Hypopharynx Cancer patients (Fig. 4).

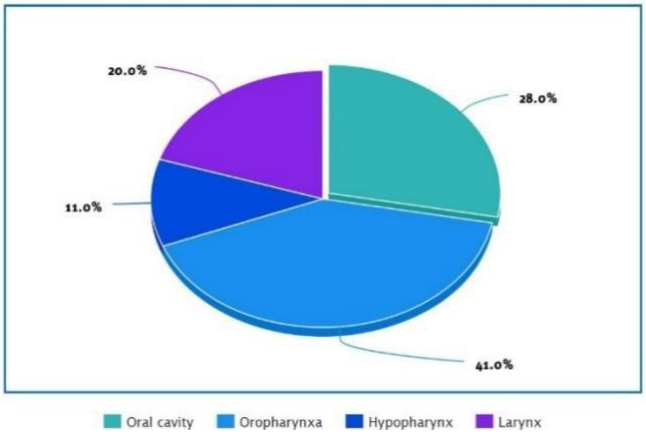

Figure 4: Percentage Population of Patients Reached III-IV Stage of HNC.

The key reasons for cancer outburst among the patients were Drinking Alcohol, Family History, Poor Dental Hygiene, and Smoking. According which (58%-900) were current or former drinkers, (25%-381) were holding records of family history, (11%-176) were diagnosed with poor dental health and (76%-1170) were current or former smokers (Fig. 5).After the treatment, at the times of follow ups (17%-257) patients were tested for Locoregional reoccurrence, (10%-151) distant relapse, and (3%-49) tested for both. Secondary primary malignancy was detected in (4%-59) cases. (<1%-7) related case for primary disease failure. In the cases of relapse-secondary primary cancer/reoccurrence 178 patients were given curative care whereas 213 patients were given palliative care. In total (39%-599) patients expired, (61%-938) were present till the last follow ups, and (83%-782) were disease free.

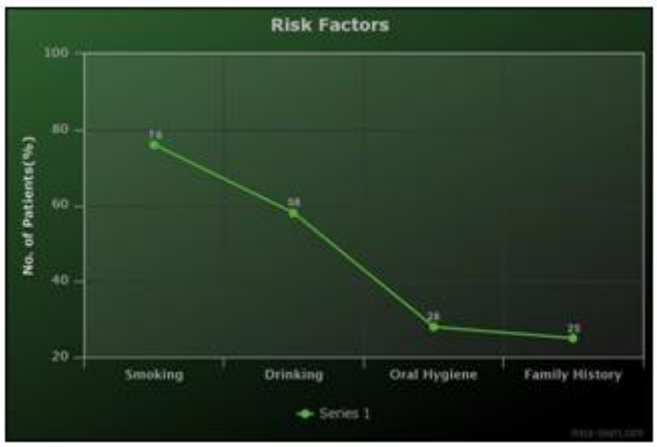

Figure 5: Patients Percentage and Risk Factor for Cancer Outburst.

2. **HNC 2.0:** Big data analytics plays a crucial role in HNC during storing, processing cancer related data. HNSCC is a diverse set of disease, each with its own biology, treatment option and results [63]. Thus, producing a huge amount of data. As a result, there are many applications of big data in HNC:

    - **Genomics:** The genomics study of HNCs can leverage a variety of big data analytics techniques Support Vector Machine (SVM), a widely used method in cancer research. This machine learning technology was used to establish a prognostic prediction model dependent on genomics data analysis. These results were shown for Oral Cavity Cancer, HPV Positive, Oropharynx Cancer, Larynx Cancer, and mix subsites of HNC treated with Chemo radiation after surgery.
    - **Technology for Optical Imaging and Enhancement:** In the evaluation of patients and the execution of surgical procedures, clinical and visual examinations are critical. Different light sources or bioactive enhancers have been used to build optical imaging technologies [63]. Nonetheless, these gains are significant. Big data technology should not be used. The importance of data and images during a clinical evaluation is comparable to that of radiological imaging. A few instances of algorithms that can enhance physicians' ability to diagnose. There are several studies of this use case, analysing preoperative,postoperative outcomesas well as intraoperative. Training and simulation simulators Virtual Reality based presurgical planning is being developed and could be used in the future. Soon, it will be valuable for both therapeutic and educational purpose.
    - **Radiation Therapy:** It is well known and effective therapy for HNCs. The aim of this therapy is to attain an inflated likelihood of confined tumour control while lowering the danger of adverse effects on regular cell. Several factors, including as delayed radiation reactions, tumour management, and adversities, affect the likelihood of normal tissue complications (NTCP). A decision-support system was built using big data to mimic treatment outcomes and NTCP. These approaches incorporate forecasting data from various medical sources to reach the top most level of precision in predicting cancer cases and recurring cases, considering the increasing complication of radio therapy [65].
    - **Radiomics:** Radiomics is the utilisation of numerous quantitative imaging features to comprehensively assess the phenotypes of tumours. In HNCs, it is the most researched field in analytics research. The early works in radiomics research focused on computer tomography (CT) imaging [65]. The basic non-contrast-enhanced CT scan is the cornerstone of radiomics development in the HNC area and is utilised for radiation planning. In HNCs [65] and certain subsites, such as the oropharynx [67-68], examples are targeted at providing a predictive biomarker. Another study [66] focused on this subsite, using CT-based radiomics to identify HPV status.

3. **HNC 3.0:** Big Data was never a common source of research in HNCs.' There are very less researches done on HNC through big data analytics. One reason can be because of its complexities and heterogeneity. Researches through big data are based on traditional analysis and study of design. However, Genomics and Radiomics that is one of the big data applications is gaining popularity among researchers [90]. Radiomics is helpful in HNC for planning doze and preventing toxicity during treatment, but HNC genomic research has not yet produced therapeutic targets.

Given the low incidence rate of HNC, Multi-Disciplinary Treatment (MDT) requires extensive and specialised supportive care as well as high skills, which are lacking in most hospitals. HNC is a group of complicated diseases with atypical presentations [1]. Thus, HNC is consider to be a challenge for clinical and researchers having 50% of survival rate. Big data analytics is welcome by many practitioners but many of them were against it, mainly because of their concerns towards loss of autonomy and privacy, unanticipated consequences, and methodological aspects [69]. Traditionally, medicine science is based on hypothesis which goes under several clinical trials. On the contrary, big data analytics works upon the collected information which are extracted by processing the data which does not have any relation hypothesis. Traditional researchers and practitioners are opposed to big data analytics tools and analytics mostly for this reason. Aside from the fact of opposition many researchers have welcomed it in a positive way. In the future, there will researches done on HNC using big data analytics, developing new technology for the treatment of this complex heterogeneous disease and increase the survival rate among the patients.

VI. BIG DATA-BASED HEALTHCARE APPLICATION

Big Data Analytics is an emerging field in healthcare sectors. As the world moving towards digitalization, it becomes necessary for everyone to go with the flow. Healthcare sector is one of the most sprouting sectors with USD 3, 32,391.42Million worth market size in 2020. In the expansion of health sectors, big data has a very important role by providing real-time processing, treatment methods for curing cancer, improving healthcare policies, preventing human error, etc. Big data analytics helps in providing towards a precise problem, thus, applied in various sectors. Some of the big data healthcare applications are listed below.

1. **Healthcare Insurance:** Big Data Analytics alone offers large number of benefits to healthcare insurance providers such as detecting frauds, providing right care, providing personalise experience, and many more. It is impossible for insurance companies to analyse big data without any analytic tools, Insurance company works upon mathematical model to predict the outcomes. Although data remains essential, there have been significant changes in the amount of information available and the method use to gather and evaluate it [70]. Hence in today's time not only analytics but rational explanations for the action place the important role providing effective analytics which is a challenging task due to numerous data and several other human factors, but it can provide a remarkable financial and healthcare benefits to the users [71].

2. **Smart Services:** Smart services can be used as a communication tool between healthcare providers and users and is very much connected to a person good lifestyle and personal wealth being. This can be analysed through Genomic Guided Big Data and Biomedical Sensors (Heartrate, Temperature, Breathing Rate, Blood Pressure) etc. All these helps to analyse and predict patient's condition.Smart services can do wonder when combine with proper analytic tools and technologies, can be a promising field of study in near future, not only in healthcare industries but also in business fields, informatics etc. One of the major uses of smart services in healthcare is to detect diseases in early stage and prescribing medicines as well as providing prevention and caution for the future well-being.

3. **Health Policy:** As the healthcare move towards digitalization and it moves towards adopting big data analytics tools and technologies, it become very important for administrators to adopt appropriate healthcare policies to reduce the risk of fraud and to protect privacy, security, confidentiality, combat the potential pressure of data commercialization, quality, transparency, safety etc. In the field of healthcare sectors, it become very important for administrators and healthcare providers to take security and safety in the account. Appropriate healthcare policies should be implemented. For Example: compensations and legal procedures should be enforced in case of security violation. Big data securities should be considered and future policies should be enforced keeping future challenges of big data and other advance technologies in consideration [72].

4. **Healthcare Security:** Forecasting potentially dangerous cases and assessing health related dangers in actual time are crucial and difficult problem in the developing health areas. The evolution of smart health protection services is revolutionizing the healthcare industries as a result of incorporation of several modern technological approaches [73]. Big Data holds great promise for enhancing patients' outcomes, forecasting epidemics breakouts, gaining insight full knowledge, preventing avoidable diseases, lowering delivery cost and general enhancement of quality of life [74]. But the major challenge is to protect and secure patients personal data. Constant bugging towards personal information is one of the main reasons why user does not completely support towards their data. Overcoming the edges of security hazards is a difficult process, rules and regulation should be implemented towards offenders. By precaution and overcoming limitations a good healthcare security system can be develop.

VII. BENEFITS OF BIG DATA IN HEALTH SECTOR

Involving Big data in health sector incorporated many benefits into it thus reducing the time in studying. Big analytics has introduced many benefits in the sectors such as, it has introduced predictive analysis which proved to be a major milestone in real time patient prediction, disease observation, and early screening of diseases. Due to introduction of latest technology, it has also helped in development of technology that are useful for the patient treatment, and has also significantly reduced frauds among the industry.

In this field of illness prevention, medical outcome prediction, minimizing medical errors, and enhancing all facets of healthcare, big data analytics have proved highly beneficial in the healthcare industries. But everything comes with their own limitation so does big data in health sector. Big data application in health sector necessitates specialized knowledge, which is hard to get thus possess as great challenge among the industry. Healthcare industries are very prone to cyber-attacks thus privacy is a big concern among the people. Keeping data safe from the possible attacks is major issue among the healthcare industries. Big data analytics has its own befits and limitation thus industries should focus on improving the benefits and work upon the potential limitation [75].

VIII. CONCLUSION &FUTURE RESEARCH DIRECTION

Big data originates from unstructured, semi-structured, and structured information. The rapid growth of big data can be attributed mainly because of its tools and wireless

technology. This survey provides insights into the uses and advantages of big data analytics in the healthcare sector. Big data analytics plays a crucial role in storing and analysing data for cancer genomics. Cancer is a complex disease that affects millions of people each year, making it essential for healthcare industries to analyse data for the development of new treatment methods. Big data analytics enables real-time data prediction using wireless technology, which can improve cancer care and delivery. The future of cancer research will focus on data sharing and better drug delivery in cancer genomics. Big data will continue to attract researchers and have a significant impact on healthcare industries and the future of cancer research. Accessible resources that provide data will play a vital role in facilitating global research efforts against cancer. In the coming decades, big data using wireless technology is expected to become a central part of many research projects, and numerous industries will focus on leveraging its potential.

REFERENCES

[1] Gani, A. Siddiqa, S. Shamshirband, and F. Hanum, ‘‘A survey on indexing techniques for big data: Taxonomy and performance evaluation,’’ Knowl. Inf. Syst., vol. 46, no. 2, pp. 241–284, 2016.
[2] Kasu P, Kim T, Um JH, Park K, Atchley S, Kim Y. FTLADS: Object-Logging BasedFault-Tolerant Big Data Transfer System Using Layout Aware Data Scheduling. IEEE Access 2019;7:37448–62. doi: https://doi.org/10.1109/ACCESS.2019.2905158.
[3] R. Hermon and P. A. Williams, “Big data in healthcare: What is it used for?” 2014.
[4] Cortada, J. W., Gordon, D., &Lenihan, B. (2012). The value of analytics in healthcare: from insights to outcomes.IBM Global Business Services, Life Sciences and Healthcare, Executive Report.
[5] Marr, B. (2015). How big data is changing healthcare. http://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is- changing healthcare/print/. Accessed 1 Dec 2015.
[6] Winters-Miner, L. A. (2014). Seven ways predictive analytics can improve healthcare. Elsevier Connect. https://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve-healthcare. Accessed 1 Dec 2015.
[7] Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. 2018;15:e1002686.
[8] Mehanna H, Paleri V, West CM, et al. Head and neck cancer – Part 1: Epidemiology, presentation, and prevention. BMJ 2010;341:c4684.
[9] Marcu, L. G., Boyd, C., & Bezak, E. (2019). Feeding the data monster: Data science in head and neck cancer for personalized therapy. Journal of the American College of Radiology, 16(12), 1695–1701. https://doi.org/10.1016/j.jacr.2019.05.045
[10] Attar E, Dey S, Hablas A, Ramadan M, Rozek LS, Soliman AS. Head and neck cancer in a developing country: a population- based perspective across 8 years. Oral Oncol 2010;46(8):591–6.
[11] Lo Nigro C, Denaro N, Merlotti A, Merlano M. Head and neck cancer: improving outcomes with a multidisciplinary approach. Cancer Manag Res. 2017;9:363-371. https://doi.org/10.2147/ CMAR.S115761.
[12] Mehra R, Ang KK, Burtness B. Management of human papillomavirus-positive and human papillomavirus-negative head and neck cancer. SeminRadiatOncol. 2012;22(3):194-197. https://doi.org/10.1016/j.semradonc.2012.03.003.
[13] Pignon J-P, le Maître A, Maillard E, Bourhis J. MACH-NC collaborative group. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. RadiotherOncol. 2009;92(1):4-14. https://doi.org/10.1016/j.radonc.2009.04.014.
[14] GloecklerRies LA, Reichman ME, Lewis DR, Hankey BF, Edwards BK. Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program. Oncologist 2003;8(6):541–52.
[15] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424. https://doi.org/10.3322/caac.21492.
[16] Big Data Analytics, IBM, New York, NY, USA, Nov. 2017.
[17] Zhang, Z. Yi, Z. Yan, G. Min, W. Wang, A. Elmokashfi, S. Maharjan, and Y. Zhang, ‘‘Social computing for mobile big data,’’ Computer, vol. 49, no. 9, pp. 86–90, Sep. 2016.

[18] Amalina, F., Targio Hashem, I. A., Azizul, Z. H., Fong, A. T., Firdaus, A., Imran, M., &Anuar, N. B. (2020). Blending Big Data Analytics: Review on challenges and a recent study. IEEE Access, 8, 3629–3645. https://doi.org/10.1109/access.2019.2923270
[19] Nazir, S., Khan, S., Khan, H. U., Ali, S., Garcia-Magarino, I., Atan, R. B., & Nawaz, M.a (2020). A comprehensive analysis of healthcare big data management, analytics and Scientific Programming. IEEE Access, 8, 95714–95733. https://doi.org/10.1109/access.2020.2995572
[20] Hassani, Sahar, and Ulrike Dackermann. "A Systematic Review of Advanced Sensor Technologies for Non-Destructive Testing and Structural Health Monitoring." Sensors 23.4 (2023): 2204.
[21] Admin. (2022, July 3). Big Data in healthcare: Examples, advantages and disadvantages. Business Compiler. https://www.businesscompilerng.com/2022/05/big-data-analytics-in-healthcare.html
[22] Cavalieri, S., De Cecco, L., Brakenhoff, R. H., Serafini, M. S., Canevari, S., Rossi, S., Lanfranco, D., Hoebers, F. J., Wesseling,
[23] F. W., Keek, S., Scheckenbach, K., Mattavelli, D., Hoffmann, T., López Pérez, L., Fico, G., Bologna, M., Nauta, I., Leemans, C. R., Trama, A., … Licitra, L. (2020). Development of a multiomics database for personalized prognostic forecasting in head and neck cancer: The Big Data to decide eu project. Head & Neck, 43(2), 601–612. https://doi.org/10.1002/hed.
[24] Bogowicz, Marta, et al. "Privacy-preserving distributed learning of radiomics to predict overall survival and HPV status in head and neck cancer." Scientific reports 10.1 (2020):454
[25] Pramanik, Md Ileas, et al. "Healthcare informatics and analytics in big data." Expert Systems with Applications 152 (2020): 113388.
[26] Kumar, Sunil, and Maninder Singh. "Big data analytics for healthcare industry: impact, applications, and tools." Big data mining and analytics 2.1 (2018): 48-57.
[27] Gandomi and M. Haider, ‘‘Beyond the hype: Big data concepts, methods, and analytics,’’ Int. J. Inf. Manage., vol. 35, no. 2,
[28] pp. 137–144, 2015.
[29] Tsai, Chiaojung Jillian, Nadeem Riaz, and Scarlett Lin Gomez. "Big data in cancer research: real-world resources for precision oncology to improve cancer care delivery." Seminars in radiation oncology. Vol. 29. No. 4. WB Saunders, 2019.
[30] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, ‘‘Big data analytics: A survey,’’ J. Big Data, vol. 2, no. 1, p. 21, 2015.
[31] Marcu, L.G., Boyd, C. and Bezak, E., 2019. Feeding the data monster: data science in head and neck cancer for personalized therapy. Journal Of The American College Of Radiology, 16(12), pp.1695-1701
[32] Willems SM, Abeln S, Feenstra KA, de Bree R, van der Poel EF, de Jong RJ, Heringa J, van den Brekel MW. The potential use of big data in oncology. Oral Oncology. 2019 Nov 1;98:8-12.
[33] Shvachko K, Kuang H, Radia S, Chansler R. HDFS - The Hadoop distributed file system. In: 2010 IEEE 26th Symp Mass Storage SystTechnol MSST2010. p. 1–10.
[34] Swift - OpenStack n.d. https://wiki.openstack.org/wiki/Swift (accessed March 10, 2020).
[35] Ghemawat S, Gobioff H, Leung ST. The google file system.OperSyst Rev 2003;37:29–43.doi: https://doi.org/10.1145/1165389.945450.
[36] What is data storage? IBM. (n.d.). Retrieved May 7, 2022, from https://www.ibm.com/topics/data-storage
[37] Nachiappan R, Javadi B, Calheiros RN, Matawie KM. Cloud storage reliability for Big Data applications: A state of the art survey. J NetwComputAppl 2017;97:35–47. doi: https://doi.org/10.1016/j.jnca.2017.08.011.
[38] Apache SparkTM - Unified Analytics Engine for Big Data n.d. https://spark. apache.org/ (accessed March 10, 2020).
[39] Samzan.d. http://samza.apache.org/ (accessed March 10, 2020).
[40] Apache Hadoop n.d. https://hadoop.apache.org/ (accessed March 10, 2020).
[41] Apache Flink: Stateful Computations over Data Streamsn.d. https://flink. apache.org/ (accessed March 10, 2020).
[42] Apache Storm n.d. https://storm.apache.org/ (accessed March 10, 2020).
[43] The Apache Software Foundation. Welcome to Apache Flume — Apache Flume. Apache Softw Found 2012. https://flume.apache.org/ (accessed June 19, 2019).
[44] Soualhia M, Khomh F, Tahar S. Task Scheduling in Big Data Platforms: A Systematic Literature Review. J SystSoftw 2017;134:170–89. doi: https://doi.org/10.1016/j.jss.2017.09.001.
[45] Big Data Processing. Big Data Processing - an overview | ScienceDirect Topics. (n.d.). Retrieved May 13, 2022, from https://www.sciencedirect.com/topics/computer-science/big-data-processing

[46] Duggal, N. (2022, March 3). What is data processing: Definition, cycle, types & methods [updated]: Simplilearn. Simplilearn.com. Retrieved May 13, 2022, from https://www.simplilearn.com/what-is-data-processing-article
[47] Z. Lv, H. Song, P. Basanta-Val, A. Steed, and M. Jo, ‘‘Next-generation big data analytics: State of the art, challenges, and future research topics,’’ IEEE Trans. Ind. Informat., vol. 13, no. 4, pp. 1891–1899, Aug. 2017.
[48] Says:, B., says:, A. A., says:, E. S., & says:, D. S. (2021, November 10). Real time big data applications in various domains. Edureka. Retrieved May 16, 2022, from https://www.edureka.co/blog/big-data-applications-revolutionizing-various-domains/
[49] Amalina, Fairuz, Ibrahim AbakerTargio Hashem, Zati Hakim Azizul, Ang Tan Fong, Ahmad Firdaus, Muhammad Imran, and Nor BadrulAnuar. "Blending big data analytics: Review on challenges and a recent study." Ieee Access 8 (2019): 3629-3645.
[50] Schlick, C.J.R., Castle, J.P. and Bentrem, D.J., 2018. Utilizing big data in cancer care. Surgical Oncology Clinics, 27(4), pp.641- 652.
[51] Sabharwal, S., Gupta, S., &Thirunavukkarasu, K. (2016, April). Insight of big data analytics in healthcare industry. In Computing, Communication and Automation (ICCCA), 2016 International Conference on (pp. 95-100). IEEE.
[52] Pramanik, M. I., Lau, R. Y., Demirkan, H., & Azad, M. A. K. (2017). Smart health: big data enabled health paradigm within smart cities. Expert Systems with Applications, 87, 370-383.
[53] M. Viceconti, P. J. Hunter, and R. D. Hose, Big data, big knowledge: Big data for personalized healthcare, IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, pp. 1209–1215, 2015.
[54] Y. Sun, H. Song, A. J. Jara, and R. Bie, Internet of things and big data analytics for smart and connected communities, IEEE Access, vol. 4, pp. 766–773, 2016.
[55] Mike, W. Hoover, T. Strome, and S. Kanwal. Transforming health care through big data strategies for leveraging big data in the health care industry, http://ihealthtran.com/iHT2 BigData 2013.pdf, 2013.
[56] J. Anuradha, A brief introduction on big data 5Vs characteristics and Hadoop technology, Procedia Computer Science, vol. 48,
[57] pp. 319–324, 2015.
[58] HealthITAnalytics. (2020, November 30). Big Data Analytics may lead to more precise cancer treatments. HealthITAnalytics. Retrieved May 29, 2022, from https://healthitanalytics.com/news/big-data-analytics-may-lead-to-more-precise-cancer-treatments
[59] How is big data helping fight cancer. How is Big Data helping fight Cancer. (n.d.). Retrieved May 29, 2022, from https://www.polestarllp.com/big-data-analytics-in-cancer
[60] 56. Flatiron Health. 2019. https://flatiron.com/oncology/ (Accessed 1 May 2019).
[61] Kulaylat AS, Schaefer EW, Messaris E, et al: Truven health analytics marketscan databases for clinical research in colon and rectal surgery. Clin Colon Rectal Surg 32:54-60, 2019
[62] Surveillance, Epidemiology, and End Results (SEER) Program. 2019. https://seer.cancer.gov/ztml(Accessed 1 May 2019)
[63] RuppinE.. Cancer Data Science Laboratory. 2019. https://ccr.cancer. gov/cancer-data-science-laboratory (Accessed 1 May 2019).
[64] Wang Z, Jensen MA, Zenklusen JC: A practical guide to The Cancer Genome Atlas (TCGA). Methods MolBiol 1418:111-141, 2016
[65] TARGET: Therapeutically Applicable Research To Generate Effective Treatments. 2019.
[66] Head and neck cancer - statistics. Cancer.Net. (2022, April 1). Retrieved June 10, 2022, from https://www.cancer.net/cancer- types/head-and-neck-cancer/statistics
[67] Resteghini, C., Trama, A., Borgonovi, E., Hosni, H., Corrao, G., Orlandi, E., ...&Licitra, L. (2018). Big data in head and neck cancer. Current Treatment Options in Oncology, 19, 1-15
[68] Tiwana, M. S., et al. "25 year survival outcomes for squamous cell carcinomas of the head and neck: population-based outcomes from a Canadian province." Oral oncology 50.7 (2014): 651-656.
[69] Mazur T, Mansour TR, Mugge L, Medhkour A. Virtual reality–based simulators for cranial tumor surgery: a systematic review. World Neurosurg. 2018;110:414–22. https://doi.org/10.1016/j.wneu.2017.11.132.
[70] Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol. 2015;5:272. https://doi.org/10.3389/fonc.2015.00272. (old 89)
[71] Grégoire V. Tumor control probability (TCP) and normal tissue complication probability (NTCP) in head and neck cancer. Rays. 30(2):105–8 http://www.ncbi.nlm.nih.gov/pubmed/16294902.

[72] Lambin P, van Stiphout RGPM, Starmans MHW, et al. Predicting outcomes in radiationoncology—multifactorial decision support systems. Nat Rev ClinOncol. 2013;10(1):27–40. https://doi.org/10.1038/nrclinonc.2012.196.
[73] Elhalawani H, Kanwar A, Mohamed ASR, et al. Investigation of radiomic signatures for local recurrence using primary tumor texture analysis in oropharyngeal head and neck cancer patients. Sci Rep. 2018;8(1):1524. https://doi.org/10.1038/s41598-017- 14687-0.
[74] ElhalawaniH,MohamedASR,White AL, et al.Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. Sci data. 2017;4:170077. https://doi.org/10.1038/sdata.2017.77.
[75] Ranjbar S, Ning S, Zwart CM, et al. Computed tomography-based texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinoma. J Comput Assist Tomogr. 2017;42(2):1. https://doi.org/10.1097/RCT.0000000000000682.
[76] Orlandi E, Licitra L. Personalized medicine and the contradictions and limits of first-generation deescalation trials in patients with human papillomavirus-positive oropharyngeal cancer. JAMA Otolaryngol Neck Surg. 2018;144(2):99. https://doi.org/10.1001/jamaoto.2017.2308.
[77] Gatta G, Capocaccia R, Botta L, et al. Burden and centralised treatment in Europe of rare tumours: results of RARECAREnet—a population-based study. Lancet Oncol. 2017;18(8):1022–39. https://doi.org/10.1016/S1470-2045(17)30445-X.
[78] Coveney PV, Dougherty ER, Highfield RR. Big data need big theory too. Philos Trans A Math Phys Eng Sci. 2016;374(2080):20160153. https://doi.org/10.1098/rsta.2016.0153
[79] Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517. https://doi.org/10.1001/jama.2017. 7797
[80] Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Affairs, 33(7), 1123–1131.

