

**Review Article**

# Stroke Care and the Role of Big Data in Healthcare and Stroke

**Lidong Wang<sup>1,\*</sup>, Cheryl Ann Alexander<sup>2</sup>**<sup>1</sup>Department of Engineering Technology, Mississippi Valley State University, Itta Bena, USA<sup>2</sup>Technology and Healthcare Solutions, Incorporated, Itta Bena, USA**Email address:**

lwang22@students.tntech.edu (Lidong Wang)

\*Corresponding author

**To cite this article:**Lidong Wang, Cheryl Ann Alexander. Stroke Care and the Role of Big Data in Healthcare and Stroke. *Rehabilitation Science*.

Vol. 1, No. 1, 2016, pp. 16-24. doi: 10.11648/j.rs.20160101.13

**Received:** September 11, 2016; **Accepted:** September 30, 2016; **Published:** November 25, 2016

---

**Abstract:** Theoretically current electronic health data can now be securely linked on an unprecedented scale, potentially illuminating how diseases manifest and which treatments are best applied in the real world. Increasing volumes of information on real-time, actual patient experiences are now contained on social media and patient portal websites. Innovations and insight into the health care for individuals and entire populations can be gained when information from health monitors, genomic data, and clinical trial data is merged. In other words, we now have the theoretical technology to accumulate, store, convert, access, and evaluate massive amounts of data at a modest cost. Performance and clinical data from health care facilities, including clinics and hospitals, clinical research data by industry, and academic data from patient populations and the general public which may be generated through social media and/or other sources is included in big data. Just as access to sizable datasets evolves and becomes easier, analytical mistakes may occur more often and be easier to make lest rigorous standards and governance controls are employed. Indeed, it is more likely that improved analytics will also introduce us to at least a few more uncomfortable insights into the negligible value of some medicines. It is also noteworthy to mention that one common error is the assumption that the value of big data is within the data itself—its volume, accuracy, accessibility, “linkability,” etc. Unfortunately, despite the importance of the information, or the “bigger” the data, the greater the likelihood that this does not hold true. This review paper examines the relationship of big data to stroke care in a variety of stroke-related issues including: big data in stroke care, big data and visual analytics, big data in telecardiology, and some challenges and indications for future research.

**Keywords:** Stroke, Big Data, Hypertension, Telecardiology, Electronic Medical Record (EMR), Electronic Health Record (EHR), Big Data Analytics, Machine Learning, Data Mining

---

## 1. Introduction

The health care system in the United States is one of the most expensive in the industrialized world yet it continues to deliver some of the worst patient outcomes in health care of other Western countries. A better approach to this problem is the “big data” approach by which specific algorithms mine the massive volumes of electronic health record (EHR), clinical trial, and genomic data to create tailored health care advice to both clinicians and patients [1]. Using this approach has the potential to advance both quality and cost.

Stroke is the fourth-highest cause of death and the primary reason for disability in the US. Even minor improvements in

stroke care can increase economic savings from the long-term care most often required of stroke patients. A stroke is serious, life-threatening; a medical complaint, which arises when the blood supply to a section of the brain is interrupted. Actually two primary reasons of stroke exist: ischemic – when the blood flow is cut off as a result of a blood (85%-90% of all cases); hemorrhagic – when a damaged blood vessel delivering blood to the brain ruptures. In the UK, stroke accounts for the third largest number of deaths and causes the most cases of adult disability. Some risk factors include high blood pressure, atrial fibrillation (irregular heartbeat), age, being overweight, lack of exercise, a poor diet, and smoking. Data entry and the collection of patient data is burdensome

and decreases time spent in patient care [2].

Differences about best practices may take years to settle via clinical studies and then to circulate to clinical practice, especially to independent state-of-the-art stroke facilities, and a data-driven clinical support system (CDSS) has the capacity to assimilate enormous amounts of real-time information to mine for and disseminate best practices. Still, stroke is an exciting domain for clinical decision support systems due to the large number of clinical variables, medications and dosages that may be administered, and surgical possibilities [2]. And when a stroke initially happens, treatment is also required, in addition to long-term management related to the high rate of recurrence. Success has been achieved with machine learning techniques, with respect to both prognosis and diagnosis in a variety of medical domains.

Huge amounts of data have historically been available in the health care industry, driven by keeping, compliance & regulatory requirements, and patient care. Although the majority of data is still kept in a hard copy file, the most current trends are toward rapid digitalization of these huge amounts of data. Actually mandatory requirements and the likelihood of improvement in health care quality and delivery while reducing costs, these extremely large amounts of data (known as “big data”) actually show promise to support a wide range of medical and health care functions, including CDSS, disease surveillance, and population health management. According to US reports, data from the US health care system reached 150 exabytes in 2011. In the health care industry where data is critical, where data is necessary to documents the history and evolution of a patient’s or group of patients’ illness and care, tools are essential for health care providers to make informed treatment decisions. Medical imaging is

growing by 20 to 40% annually so that in 2015, an average facility was generating 665 terabytes of medical data every year. In order to cut health care costs dramatically and accommodate large population sets, large sets of medical data need to be regularly collected and the EHR filled with high resolution X-ray images, mammograms, 3D magnetic resonance imaging (MRIs), etc., which would not only drive increases in efficiency and quality but also cut the costs of health care drastically. According to reports, big data for the US healthcare system will soon reach the zettabyte (1021 gigabytes) scale, and not too much later, the yottabyte (1024 gigabytes) at the current rate [3].

A composite term to describe evolving technological capacities to explain complex tasks—Big Data—has been touted by industry authorities, business strategists, and marketing experts as a new frontier for innovation, competition, and productivity. Big data applications in health care are as numerous as they are multifaceted, both in research and practice. Remote patient monitoring, an increasingly prevalent market sector of machine-to-machine communications (M2M), is providing a valuable basis of life-saving information. For example, patients with diabetes are at high risk for long-term complications such as blindness, kidney disease, heart disease, and stroke. Remote tracking in glucometer readers (blood glucose monitors), helps supervise patient compliance with suggested glucose levels [3]. EHRs are populated with data in real time. Patient data can be tracked using a time series, which is also available to identify abnormalities and form the foundation of treatment decisions. In table 1 [1], some factors associated with Big Data in stroke are outlined.

**Table 1.** Big Data and some stakeholders.

Stakeholder	What Patients want from big data	The American Heart Association’ role
Patients	1. Controlled access to portable secure medical information	<i>High priority</i> Facilitate the use of Big Data by patients and professionals
	2. Access to best possible health outcomes at affordable cost	<i>Medium priority</i> Patient-centered advocacy on the use of Big Data
	3. Easy access to medical research/clinical trials	<i>Low priority</i> Facilitate the evaluation of health technology devices
Clinical Investigators	1. High-quality standardized clinical data for secondary use	<i>High priority</i> Training and funding a new generation of Big Data and users (clinicians to developers)
	2. A standardized technology platform with interoperable, feasible, and federated access to a broad range of clinical data	<i>Medium priority</i> Bridge the gap between the theoretical promise of Big Data to potential use (eg, map EMR to “Get With The Guidelines”); provide leadership in data standard, quality, and validity)
	3. Better, novel, more rapid mechanisms of support for analysis of Big Data	<i>Low priority</i> Be the match.com for data owners and data researchers
	4. A mechanism for ongoing discussion of these topics including the clinical investigator community	
Clinicians/ Healthcare System Researchers/ Administrators	1. Engagement across stakeholder domains	
	2. Empowering patients using Big Data	<i>High priority</i> Develop and disseminate accepted clinical standards and benchmarks
	3. Identifying at-risk populations with Big Data decision support	<i>Medium priority</i> Sponsor the development of multidisciplinary tools for data analysis
	4. Assessing and equipping providers with tools for collection, distillation and visualization	<i>Low priority</i> Convene all benchmarking communities and stakeholders
	5. Leveraging Big Data to enrich the practice of medicine—more efficient and more enjoyable	
	6. Teaching providers about Big Data	

Studying large datasets of patient features, outcomes of treatments and their cost can help identify the most clinically effective and cost-efficient treatments to apply. Analyzing large datasets of patient characteristics, outcomes of treatments and their cost can help identify the most clinically effective and cost-efficient treatments to apply. Models, analytics and visualizations of endless amounts of data come together to offer a different perspective of any problem in the framework of other problems, as well as in the contexts of time and geography. The four *Vs* (*velocity, veracity, volume, & variety*) are often used to illustrate diverse characteristics of big data [8].

Chronic disease, an aging population, and new and changing consumer expectations are shifting the way health care is acquired, purchased, and received. Social media and mobile technologies has improved access to care and health care delivery in new ways [5]. Health care systems in most countries are changing payment models from a fee-for-service approach to an outcomes-based or accountable (quality) approach, increasing the need for more accurate data and data tracking, and making documentation requirements more specific.

Lack of a comprehensive view of clinical and operational methods needed for discovery of areas where development was essential and operation based on vague or unfinished cost or care measurements have historically plagued health care organizations. Other industries have led the way in converting how data and analytics are used and the health care industry is finally beginning to follow their lead. A more predictive and progressive data model which can predict high risk populations is called *advanced analytics*. Both structured and unstructured data from clinical, operational, and financial systems; streaming data from monitoring and sensing devices across the spectrum of care delivery as a whole; and data from outside an organization such as social media and public health records are big data sources [5]. An inventory of current and future prospective data sources is the best way to identify likely options for health care organizations.

The way cardiovascular and stroke research is conducted and clinical care delivered is affected by Big Data. Big Data possesses incredible promise for revolutionizing this research and how clinical care is delivered. “Big Data,” which denotes large and multifaceted datasets—including, for example, biomedicine, genomic, clinical, and environmental data—and basically novel methods for data storage, administration, integration, analysis, and visualization. Uncomplicated cardiovascular research datasets are multidimensional in character with an extensive range of clinical and biomarker outcomes; for example, electrocardiogram, contractile function, molecular imaging, channel activities, genomics, proteomics, metabolomics, and phenotype characterizations [1]. However, this type of data is gathered and reported in variable data formats; they are also unevenly disseminated. Thus, the datasets are widely dispersed and fragmented, making knowledge extraction difficult whether by individual laboratories or organized scientific enterprises via teamwork.

## 2. Big Data in Stroke

Sizable, assorted, complex, longitudinal, and distributed datasets are commonly referred to as “Big Data.” The datasets can be generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital resources accessible today and in the future. US government agencies mark the scientific, biomedical, and engineering research. Large-scale, diverse, and high-resolution datasets allow communities to undergo an intense transformation to foster data-intensive decision-making, including clinical decision-making, never before paralleled. Innovative statistical and mathematical algorithms, prediction formulas, and modeling methods, as well as multidisciplinary methods for data collection, data analysis, and recent break-through technologies for distributing data and information are allowing for a paradigm shift in scientific and biomedical research [1]. Research advancement will increase the chances of, for example, demonstrating social networks and knowledgeable communities, consistent prediction of customer performances and favorites, and the developing of communication designs among unidentified groups at a larger, global scale; removal of meaning from textual data; more efficient correlation of procedures; greater capacity to extract knowledge from large-scale experimental and observational datasets; and mining beneficial material from partial data.

By classification, big data in health care denotes electronic health datasets so great and intricate that they are tough (or impossible) to cope with conventional software and/or hardware; nor can they be simply managed with traditional or ordinary data management means and approaches. Big data in health care is vast not only due to its volume but also due to the diversity of data forms and the speed at which it must be handled. In the health care business, big data is connected to patient health care and wellbeing and make up much of the information in health care. Clinical data is included from clinical decision support systems (physician’s written notes and prescriptions, medical imaging, laboratory, pharmacy, insurance, and other administrative data); patient data in the EHR; machine generated/sensor data, such as from monitoring vital signs; social media posts, including Twitter feeds (so-called tweets), blogs, status updates on Facebook and other platforms, and web pages; and less patient-specific material, including emergency care data, news feeds, and articles in medical journals [8].

Yet discovering the right data is not easy, and although researchers may find the data, it is often contained in disparate locations and databases. Data can also be rendered useless because the merging of information from these sources is always challenging and can lack the necessary identification data. These missing links can often mean the difference from an effective data-mining effort and a failure. Unfortunately, however, those databases are often too unconnected and do not communicate. Fortunately for health care organizations, the

problem of a missing link between databases is a rational problem, which with a little foresight, can be avoided. Another problem that can arise in big data is the connection of analyzed data and the equation being solved. Most of the data in big data is not created from planned experiments or surveys; rather, they are transactional facts or other data accumulated by observations without consideration from either a specific purpose or design. Yet when data is gathered from multiple sources, errors and omissions are always possible. This causes missing values, missing variables, measurement variation, and disputes where data characterization is concerned, as well as links that join the two datasets and foster analysis for application of the variables in both [6].

The first US stroke registry has been created to prospectively gather longitudinal data on each stroke patient from date of event to at least six-months post-stroke event. First national stroke register in the world to prospectively collect longitudinal information on every stroke patient from onset to six months after stroke. This is an effective method of data collection; a solitary source of data which evades replication of data entry, and saves money so that huge amounts of data are turned into timely, consequential, and available data for a wide choice of audiences to foster change and improvement. With the appropriate organization and management of datasets, keeping them accessible, comprehensive, and analyzable is a significant duty for basic science researchers. Transformations of big data analytics has resulted in new digital technologies and informatics systems, and preclinical researchers can utilize them to tackle these challenges. These enabling platforms are intended to support integrated community endeavors and are immediately pertinent in cardiovascular science. Big Data is quickly developing with respect to volume (bytes) but more so in relation to significance and importance to scientific research [1, 4].

A complete big data cyber infrastructure is need for broad communities of scientists and engineers to gain access to diverse data and to the highest quality and most useful inferential and visualization instruments. There are several potential areas for research, including but not limited to: fresh collaboration environments for assorted and distant clusters of investigators and students to organize their work (e.g., through data and model sharing and software reuse, tele-presence capability, crowdsourcing, social networking capabilities) with significantly boosted efficacy and success for the scientific partnership; automation of the discovery development (e.g., through machine learning, data mining, and automated inference). automated modeling instruments to suggest numerous interpretations of massive datasets which are beneficial to various disciplines; new data curation techniques for handling the intricate and large stream of scientific output in a wide variety of disciplines; advance of systems and methods that proficiently include autonomous anomaly and trend detection with human collaboration, response, and reaction; end-to-end systems that enable the advance and usage of scientific workflows and novel applications; innovative tactics for the advance of research

questions that might be followed in lieu of access to heterogeneous, widely varied, big data; new models for cross-disciplinary data fusion and knowledge sharing; new tactics for efficient data, knowledge, and model sharing and collaboration diagonally across numerous domains and disciplines [7].

Hospital readmission is costly and for the most part, avoidable. Diminishing unnecessary readmission is thought to be a vital quality of care parameter which can be considered measurable. In the past it has been difficult to store, manage, and mine huge volumes of structured and semi-structured health datasets. However, recent strides in big data infrastructure has made handling that data more efficient. One of the growing capabilities of new shared nothing, distributed, and parallel computing infrastructure is the capacity for performing comparable operations on great quantities (petabytes) of data. Infrastructures have evolved into the capacity to handle such large volumes, high velocity, and diverse forms of data (variety of data) because of its intrinsic nature of bringing computation nearer to where the data is, which is in contrast to the preceding paradigm of moving data around for huge computations to occur [10].

This capacity to process increased amounts of diverse unstructured, semi-structured, and structured data within the health care informatics setting, permits clinical informatics to develop new visions and uncover original knowledge by uniting data from several sources. These sources can be internal as well as external to the EHR and may consist of millions of rows and hundreds of traits that can be leveraged for predictive modeling. Apache Hadoop is one example of this type of distributed framework which executes the computational paradigm Map Reduce, where the application is separated into many small pieces of work, which may be executed or re-executed individually into many compute nodes in a cluster of data intensive distributed applications [10]. Figure 1 shows the role of Big Data in healthcare information management.

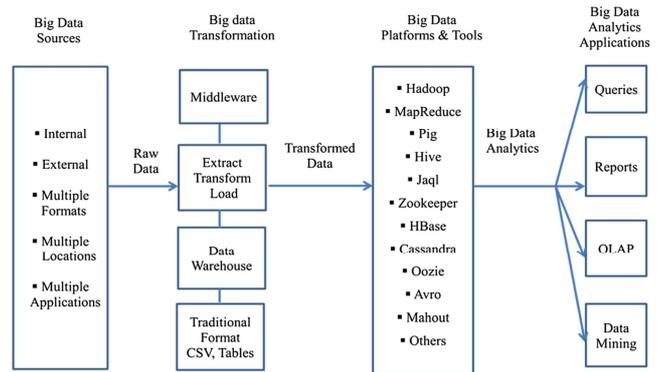


Figure 1. Shows the concept of Big Data analytics in health care management [8].

Sensors are one method to monitor activity level, daily weight, or other relevant health markers (eg, smartphone “apps”). Patients can connect with health care providers via telemedicine, email, or other electronic resources and may even join in clinical research via smartphones. And as the

population continues to engage in creating PGHD (personal-generated health data), these datasets are starting to assume characteristics commonly ascribed to big data: volume, velocity, and variety [1]. However, big data is the vital holistic and interpretive lens through which data are streamed and by which real information is then received lacks context, which is unlike PGHD [2]. The cloud supports secure sharing of data, at both technical and economic levels.

Electrophysiological signal data plays a crucial role in patient care and clinical research; it is the rapidly rising volume of multimodal data across multiple disease domains. Furthermore, the Cloudwave platform performs many functions: (a) delineates parallelized algorithms for calculating cardiac measures using the MapReduce parallel encoding framework, (b) reinforces real-time interface with huge volumes of electrophysiological signals, and (c) characterizes signal visualization and querying functionalities employing an ontology-driven web-based interface. Comparative calculations of Cloudwave with traditional desktop methods to calculate cardiac measures (eg, QRS complexes, RR intervals, and instantaneous heart rate) demonstrates the necessity for big data technologies in appraising health care data. However, one crucial problem in utilizing cloud infrastructure and platforms is data privacy [9]. The defining trait of these datasets is increased volume, and the lingo “big data,” is most frequently used to define both the data and the distinctive features of their administration. End users are often concerned with volume, in addition to velocity of big data, which is defined as the extreme rate of data generation and requirement for quickly analyzing data for critical decision-making tasks.

The growing requirement for adopting a comparable paradigm of “health care intelligence” through almost real-time processing of health care information to sustain preventive care, personalized medicine, and improved treatment outcomes. For example, there can now be routine analysis of continuous EEG, ECG, blood oxygen levels, and video data [9]. Current computational methods for processing signal data are restricted in the ability to sustain a collaborative multi-center research study in this domain and are distinguished by both volume (eg, terabytes (TB) of data per year) as well as velocity (eg, gigabytes (GB) of data per month). Data is, more often than not, required to fit into the memory of a local desktop and also do not have the capacity to efficiently leverage the increasing capacities of distributed computing methods (eg, cloud computing and multicore processing) [9].

Some datasets can be obtained with a simplified retrieval process and some factors influence how that process proceeds: a lack of summary of approvals for diverse uses of accessible data, who ‘owns’ it, and how to retrieve it; linkage and interoperability boundaries within and between health care, academia and private sector datasets; incomplete data faculties and an immature intermediary market to distribute swift business intelligence results in the innovation space. The following are some concerns about the extent and depth of data coverage and availability across primary and secondary

care and associated services such as care homes, mental health and ambulance services; the requirement for encompassing the governance framework that additionally shields patient confidentiality and attends to issues such as intellectual property, heightened data security, and data quality assurance for innovative big data resources; partial experience among our industrial members and across health care commonly for the extent of big data opportunities and a necessity for a stronger formulation. Altering big data into insight in order to advance health care is not primarily about gathering even bigger amounts of data. Rather there is a necessity for blending technical, analytical, and clinical skills and knowledge and a necessity for a dynamic market for data resolutions and facilities, and partnership across the ecosystem [9]. Big data also needs much more than just data – in addition, a supportive ecosystem also embraces skills, services, technical platforms, standards, legal and governance frameworks, and financing mechanisms is necessary for computation [1].

### **2.1. Big Data and Visual Analytics**

A vital difficulty with Big Data is understanding the data quickly, so by employing a Visual Analytics method, the originally overpowering scale of Big Data develops into a valued asset. Interactive visualization allows researchers to recognize data which may be described as encompassing, multisource, variable, and time varying data. Fortunately, visual analytics reinforces big data by affording interactive visualizations which permit scientists to navigate these complex datasets. Defined as “the science of analytical reasoning facilitated by interactive visual interfaces,” visual analytics is more than simply visualization of the information; rather, it is an approach which blends visualization, human aspects, and data analysis. When combined with visual analytics, big data become increasingly more prevailing because the synthesized data which is provided is identified more rapidly. The utilization of big data is therefore, increased for the decision-making process, which may be limited to time in real-time. In addition, big data and visual analytics can contribute much more significantly to clinical research with the use of de-identified health data and pragmatic clinical trials. The highest level of evidence during the transformation of clinical research into clinical practice however, is provided by randomized controlled trials (RCTs). There does seem to be an increasing number of multi-institution, industrial, global, and multi-center alliances, which quite frequently come together to extract and benefit from the great potential of the use of big data in health care [12]. The combination of big data with visual analytics will provide increased promise for the improvement of health care services delivery and change how clinical research is conducted. Furthermore, in the face of the numerous challenges due to multiple data sources as well as the mystery of data quality, examples of big data registries for refining health outcomes and clinical trials have been very successful.

MapReduce is a widespread programming framework presented by Google to tackle computational and storage

challenges for web-scale data. Apache Hadoop is an open-source implementation of the MapReduce framework, which can be utilized for storage of great volumes of information on the Hadoop Distributed File System (HDFS) and which can effectively manage data by repeating the two steps of 'Map' and 'Reduce' on thousands of computing nodes. And electrophysiological signal data are more and more categorized by both massive volume and high velocity, and are playing a more influential role in maintaining patient care and clinical research [9]. Cloudwave is an adaptable and scalable platform for supporting scientific clinical research studies using massive-scale signal data in diverse disease spheres. Machine learning procedures utilized for neuroscientific "big data" sets are key to understanding these multifaceted brain-behavior associations [9].

One method of machine learning technique is partial least squares regression (PLSR), which has applications in the arena of neuroimaging. During prior studies of brain-behavior relationships, primarily in the investigation of functional MRI, the well-established notion that the location of tissue damage is a vital component for deciding the attendant functional deficit, or rather, sign or symptom. A moderately ranged dataset of both neuroimaging and wide-ranging behavioral information was applied to a machine learning method to recognize potential brain structural connectome-behavior relationships. Models of eloquent functions, in other words, language and motor, provided validity for using this method, while models of more complex behavioral measures gave new insights into the relationships between brain and behaviors. Robustness for this technique was confirmed by the replication of connectome-behavior relationships for a specific function across pathologies. In addition to identifying the potential for applying this to the improvement of post stroke prognoses, the existing analysis suggest that the opportunity to gain further insight into the neural substrates underlying complex behaviors such as those associated with activities of daily living and specific areas of cognition exists [12].

## **2.2. The Role of Carotid Artery Stenting in Stroke**

One effective treatment for the population with ischemic strokes who have moderate-to-severe carotid artery stenosis is carotid artery stenting. However, the midterm result for patients going through with this procedure contrasts significantly with baseline characteristics. The most solid predictive reason influencing post stenting outcome is low-density lipoprotein. In addition, further evidence exists that carotid artery stenting can be beneficial for the patient population of first-time ischemic strokes with baseline mRS scores. An important fact for big data scientists is that stroke is a principal cause of morbidity and mortality in modern society [11]. Stroke causes significant irreversible neurological deficiencies, and research to identify the underlying mechanism is key to reducing the likelihood of recurrence.

The incidence of extracranial carotid stenosis with a reduction of 50% in the lumen is over 5% in people aged 65 and older. When stenosis is 50% or greater, the risk of ipsilateral stroke grows, an estimated 1%-5% per year, while it

increases with degree of stenosis. There may also be evidence that the incidence of stroke is higher in Taiwanese and Chinese populations compared to European populations. A high risk of recurrence exists following initial ischemic stroke. Several factors have demonstrated a prediction for recurrence: stroke severity, age, and degree of stenosis; evidence shows that the of recurrence during the first 30 days after initial stroke is 4% and in the first year is 12%. The cumulative rate of recurrence is 10.5% based on the findings of one Taiwanese study, which included a 2.5-year follow-up duration. Medical treatment, lifestyle changes, and surgical interventions such as carotid artery stenting and carotid endarterectomy are several options for treating carotid artery occlusion in stroke patients [11].

Based on the initial findings, the degree of stenosis, degree of stroke, age of the patient, and other presenting baseline characteristics determine what risks and benefits the physician may decide. Carotid artery stenting can successfully stop secondary stroke, ameliorate atherosclerosis, and reestablish blood supply to the brain parenchyma. When a reduction of the diameter of the lumen is 70%, current guidelines recommend carotid artery stenting in patients with stenosis, that is when assessed by noninvasive imaging, and 50% if measured by catheter-based imaging. For patients needing stenting, a high low density lipid (LDL) level has been connected with poor outcomes. The proinflammatory nature of carotid artery stenosis may be validated by the fact that gouty arthritis is associated with poor outcome. A well-known risk factor for gouty arthritis is a high serum uric acid concentration is a known risk factor for gout, and one researcher has discovered that in the population of patients with ischemic stroke, the rate of good clinical outcomes increases significantly with an increased serum uric acid concentration [11].

## **2.3. The Role of Hypertension in Stroke Care & Big Data Analytics**

A major public health issue, hypertension and diabetes quite often exist simultaneously with the development of demographic aging, rapid urbanization, and the globalization of unhealthy lifestyles and are a foremost public health issue. For example, hypertension already affects one billion people globally, and leads to heart attacks and strokes, as well as higher blood pressures and presently kills nine million individual each year. Big data predicative analytics is used to generate models to forecast future conclusions or events founded on current big data. Big data prescriptive analytics is prescriptive analytics for big data, which answers such questions and complicated equations such as what we should do, why we should do it, and what should happen with the best result under any doubt. For example, it is an optimum marketing approach for any e-commerce corporation [13].

One key platform for big data analytics is Apache Hadoop. Hadoop can easily scale up to hundreds or possibly thousands of nodes as an open source platform for storing and processing large datasets utilizing clusters and commodity hardware. Apache Spark, one of the most common big data analytics services, has mobilized from existing as a component of the

Hadoop system, to the big data analytics platform for quite a few enterprises [13].

#### ***2.4. Telecardiology Expanding the Role of Big Data in Stroke Care***

Telecardiology allows experienced cardiologists to diagnose timely and offer efficient therapeutic treatments for the population of rural areas where there are fewer or no trained cardiologists. As a tool to reduce the cost of transportation from home to hospital in addition to unnecessary hospital transfers, it is lowering the mortality rate of patients with heart attacks. As a crucial tool for telecardiology, wireless telecommunication brings timely and reliable services with fewer interruption errors when compared to a traditional telephone line. Mobile computing and cloud computing has allowed groundbreaking advancement in telecardiology in the past ten years. The distribution of cloud computing has economically facilitated the collaborative application of telecardiology between hospitals and expanded services from regional to international. However, the conveyance of telecardiology services to an experienced cardiologist for well-timed and effective interpretation remains a great trial [14].

Availability, accessibility, and scalability are the crucial features of cloud computing technology. More importantly, many cloud computing providers can regularly back up data and store the duplicated data at the variable datacenters of the cloud providers. In comparison to traditional data analysis, big data computing needs large-scale data capacity [14].

### **3. Challenges of Big Data in Stroke**

Big data may also increase and further intensify disparities in health care outcomes, creating another public health concern. In order to improve health and disease management, there has been increased enthusiasm for harnessing the use of big data in health care from cell phones, geospatial location, and biological real-time monitoring of health conditions; however, restricted access to smartphones and health literacy are irregularly distributed by age, race, socioeconomic status, and rurality [1].

Emerging use of wearable sensors and connected devices is a chief new source of big data, and permit continuous personal health data acquisition. A considerable percentage of cardiovascular disease (CVD) and stroke events across primordial, primary, and secondary prevention, is one of most significant reasons to integrate sensor data, many of which largely go unmeasured [1].

Privacy has become a challenging issue for the big data era because data that are de-identifiable by Health Insurance Portability and Accountability Act (HIPAA) standards may become identified when further data are come to light [10]. Organizations who share data may learn from experience in collecting data on the results of their model of data sharing, publicizing this information and the lessons learned, and continuously refining the data-sharing process to amplify the advantages of data sharing and reduce the risks [1].

Many big data science methods and methodologies such as data mining, machine learning algorithms, crowdsourcing annotation platforms, cloud computing infrastructure, and Bayesian network algorithms are recent in the basic cardiovascular society. But despite the source, these approaches are practicable in the world of basic cardiovascular science. Big data theories of integrating numerous forms of data would allow basic scientists to potentially develop and recognize novel targets that may not be identified by traditional approaches. New data mining and analysis practices would also permit researchers to query for genes and proteins linked to CVD and stroke. Multiple genes and/or proteins that together cause CVD and stroke could be identified using a systems approach. Advances such as these could foster new prognostic markers and, theoretically, therapeutic targets [1].

Combined with a distributed system such as Hadoop, the Apache Mahout framework delivers a beneficial set of machine learning libraries for applying modeling tasks such as classification and clustering although there is substantial necessity to discover advanced domain specific implementations of these algorithms. Hadoop can be leveraged as a big data framework to archive performance, scalability, and fault tolerance for the task at hand. Hadoop is a common open source map-reduce implementation, and is currently being utilized as an alternate to store and process extremely large datasets on commodity hardware. Avoiding hospitalization is a foremost factor for reducing patient morbidity, improving patient outcomes, and reducing health care costs [14]. There is a need for ongoing work to leverage big data infrastructure for designed risk calculation instruments, designing much more complex predictive modeling and feature extraction techniques, and prolonging proposed solutions to predict other clinical risks.

Most telemedicine services are offered only within a hospital, between a clinic and a specific hospital, or between under-resourced clinics and a metropolitan hospital, which do not have a wide enough service area. Furthermore, the length of the services is more often than not, constrained by the accessibility of experienced cardiologists. In an attempt to achieve defeat over these restrictions, a common telecardiology platform with shared access nationwide, or even globally is in high demand [14]. A matter for note is that cloud computing users are given instantaneous access to computing resources as utility computing at a low cost.

There are quite a few advantages to taking a privacy law tactic to challenging big data surveillance: the laws are principle-based, which gives them the flexibility to tackle the developing challenges of data processing and predictive analytics. Privacy law, a well-known framework for residents and governments – hands them a lens through which to evaluate the opportunities of big data surveillance and the legal uses of that data for compliance. Opportunity for independent oversight and enforcement and the chance to develop a language to describe the different effects and wider societal problems of global, creeping surveillance is huge [15]. However, big data is unlike data linking: with data-linking the

emphasis is on individual privacy problems, whereas big data analysis is centered on the collective. There is a necessity to convey how to catch the broader societal evils of insistent and universal big data surveillance and shape that into the oversight model. And there may be indirect penalties or troubles related to big data that are not the case with data-linking (i.e. being in a predictive category vs. targeted as a distinct individual). There is a necessity to move toward academia and officials in the human rights arena. This interchange must contain what forms of government-held knowledge we are prepared to control in the name of big data, and what types we are not.

It is difficult to aggregate and analyze unstructured data [16]. Efficiently handling large volumes of medical imaging data and understanding unstructured clinical notes are challenges [17]. The capture, indexing and processing of continuously streaming, fine-grained, and temporal data is a challenge [18]. Data hackers have become more damaging in big data. Data leakage can be costly [19] Lack of infrastructure and policies, standards and practices that make the most of big data in healthcare were also cited as a concern [20].

## 4. Conclusion

Further study is necessary for understanding exactly how personal health data can be optimized so that we can enhance what is known currently as big data. Further study and understanding in the ability to interpret data from the point of care and from devices and “wearables” driven by the “Internet of Things,” in addition to environmental data (social, financial), provides the basis for potentially actionable advances in care delivery. By coupling big data with analytics and machine learning, researchers can gain understanding to establish the groundwork for a cloud-based interoperable ecosystem. Technology and treatment elasticity that can effortlessly trail the patient through the evolutions of care are vital to operationalizing the advantages of big data.

Big data analytics is an evolving science and technology which involves the multidisciplinary state-of-the-art data and communication technology (ICT), mathematics, operations research (OR), machine learning (ML), and decision sciences for big data. The core elements of big data analytics consist of big data descriptive analytics, big data predictive analytics, and big data prescriptive analytics. By definition, big data descriptive analytics are descriptive analytics for big data and is utilized in discovery for new, nontrivial information, and to explain the features of objects and relationships between entities within the current use of big data [13]. Some of the problems focused on include what occurred and when, in addition to what is happening. Big data predicative analytics is predicative analytics for big data, which emphasizes predicting trends by focusing on questions such as what will occur, what’s going to transpire, what is likely to happen, and why it will occur.

In this current day of big data, it is necessary to both benefit from what these data can deliver in terms of our decisions, yet appreciate their limitations. The swiftness of decision making

– the time removed from data input to decision output – is a key element in the big data conversation. Big data systems must also be capable of management of data and linking data flows coming in at multiple frequencies. Manipulated by the three Vs, big data has a tendency toward holding too much uncertainty which can be ascribed to data inconsistency, incompleteness, ambiguities, and latency. The big data core technologies are distributed file systems, programming models, and scalable high-performance databases. The year 2013 was an informative year in the sphere of big data analysis and visualization in mass surveillance. For example, EHRs and real-time self-quantification could represent an enormous leap toward streamlining the prescriptions of drugs or diet and fitness plans. It will be necessary to examine threats and risks and reassess procedures in view of big data, and adapt technical solutions in response.

---

## References

- [1] Antman, E. M., Benjamin, E. J., Harrington, R. A., Houser, S. R., Peterson, E. D., Bauman, M. A., ... & Daugherty, A. (2015). Acquisition, Analysis, and Sharing of Data in 2015 and Beyond: A Survey of the Landscape: A Conference Report From the American Heart Association Data Summit 2015. *Journal of the American Heart Association*, 4 (11), e002810.
- [2] Coroian, D., & Hauser, K. (2015, April). Learning stroke treatment progression models for an MDP clinical decision support system. In *SIAM Int'l Conference on Data Mining*.
- [3] Adolph, M. (2013). Big Data: Big today normal tomorrow, ITU-T Technology Watch Report, November 2013. <http://itu.int/techwatch>.
- [4] Kavanagh, M. (2015). Bringing Big Data to Quality Improvement in the Sentinel Stroke National Audit Program, Royal College of Physicians, Presentation, 16 June.
- [5] IBM Software (2013). Data-driven healthcare organizations use big data analytics for big gains, White Paper, IBM Corporation, USA, February.
- [6] De Veaux, R. D., Hoerl, R. W., & Snee, R. D. (2016). Big data and the missing links. *Statistical Analysis and Data Mining: The ASA Data Science Journal*.
- [7] NSF-NIH Interagency Initiative. (2012). Core techniques and technologies for advancing big data science and engineering (BIGDATA).
- [8] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2 (1), 1.
- [9] Sahoo, S. S., Jayapandian, C., Garg, G., Kaffashi, F., Chung, S., Bozorgi, A., ... & Zhang, G. Q. (2014). Heart beats in the cloud: distributed analysis of electrophysiological ‘Big Data’ using cloud computing for epilepsy clinical research. *Journal of the American Medical Informatics Association*, 21 (2), 263-271.
- [10] Zolfaghar, K., Meadem, N., Teredesai, A., Roy, S. B., Chin, S. C., & Muckian, B. (2013, October). Big data solutions for predicting risk-of-readmission for congestive heart failure patients. In *Big Data, 2013 IEEE International Conference on* (pp. 64-71). IEEE.

- [11] Yu, C. S., Lin, C. M., Liu, C. K., & Lu, H. H. S. (2016). Impact of baseline characteristics on outcomes of carotid artery stenting in acute ischemic stroke patients. *Therapeutics and clinical risk management*, 12, 495.
- [12] Kamal, N., Wiebe, S., Engbers, J. D., & Hill, M. D. (2014). Big data and visual analytics in health and medicine: From pipe dream to reality. *Journal of Health & Medical Informatics*, 2014.
- [13] Sun, Z. Natural Treatment of Hypertension and Diabetes Based on Big Data Analytics.
- [14] Hsieh, J. C., Li, A. H., & Yang, C. C. (2013). Mobile, cloud, and big data computing: contributions, challenges, and new directions in telecardiology. *International journal of environmental research and public health*, 10 (11), 6131-6153.
- [15] Denham, E. (2016). Speech to the big data surveillance plenary research workshop, information and privacy commissioner for BC, May 12.
- [16] White, S. E. (2014). A review of big data in health care: challenges and opportunities. *Open Access Bioinform.*, 6: 13-18. DOI: 10.2147/OAB.S50519.
- [17] Priyanka, K. and Kulennavar, N. (2014). A survey on big data analytics in health care. *Int. J. Comput. Sci. Inform. Technologies*, 5: 5865-5868. <http://www.ijcsit.com/docs/Volume%205/vol5issue04/ijcsit20140504224.pdf>
- [18] Schultz, T. (2013). Turning healthcare challenges into big data opportunities: A use-case review across the pharmaceutical development lifecycle. *Bull. Association Inform. Sci. Technol.*, 39: 34-40. DOI: 10.1002/bult.2013.1720390508
- [19] Schmitt, C., Shoffner M. and Owen, P. (2013). Security and privacy in the era of big data: The SMW, a technological solution to the challenge of data leakage. RENC1, University of North Carolina at Chapel Hill.
- [20] Bulletin Board, (2014). Industry assesses potential, challenges of big data. *J. AHIMA*.

## Biography



**Ms. Cheryl Ann Alexander** is a research scientist and CEO of Technology & Healthcare Solutions, Inc., a non-profit research and consultant firm located in Mississippi, USA. She received her Bachelor's degree from University of Mississippi and her Master's degree from University of Phoenix in USA. She is currently a doctoral candidate and will graduate soon. She is a member of both healthcare and engineering organizations. She has published over 30 papers in the areas of medical applications and healthcare in professional journals.