Bachelor's thesis

Information and Communications Technology

2020

Yang Haorong

THE APPLICATION OF BIG DATA

in electronic medical records

TURKU AMK
TURKU UNIVERSITY OF APPLIED SCIENCES

BACHELOR'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Information and Communications Technology

2020 | 27

Haorong Yang

THE APPLICATION OF BIG DATA IN ELECTRONIC MEDICAL RECORDS

ABSTRACT

With the advent of the era of big data, various industries and fields have begun to involve big data which has fundamentally changed the organization, management, analysis and utilization of data. Among them, medical and health care is one of the most promising fields for applying big data to make changes. In view of the deep influence of Coronavirus disease 2019 and the large number of infectious persons, various medical data and diagnostic data stored in electronic medical records provide direct data support for scientific research. Medical health big data has great potential in improving the treatment effect of patients, predicting the outbreak of epidemics, gaining valuable insights, avoiding preventable diseases, reducing the cost of medical services and comprehensively improving the quality of life. Starting from the basic characteristics of medical big data and the development of electronic medical records, this thesis analyzes the framework and analysis model of a medical big data platform as well as the application, privacy and security of electronic medical records. A literature review was carried out to achieve the objectives of this thesis.

KEYWORDS:

big data, blectronic medical records, security analytics.

CONTENTS

FIGURES

TABLES

1 INTRODUCTION

With the rapid development of the Internet, the Internet of Things, cloud computing, and mobile medical care, big data applications have achieved outstanding development. All these signs indicate that the era of medical big data has finally come. As the population ages and the incidence of various chronic diseases increases, people's demands for medical and health services are also increasing accordingly. Medical institutions also hope to achieve the data-driven clinical research as well as improve the level of clinical research decision-making support via the integration of different hospital internal data resources.(Lee and Yoon 2017, 3.) Furthermore, the application of medical big data analysis has become increasingly evident in the promotion of hospital clinical information construction.

In order to realize the data sharing of different clinical information systems, it is necessary to establish a hospital information integration platform which has become the consensus and direction of hospital clinical informatization construction. In recent years, a great number of general hospitals have established electronic medical record applications based on information integration platforms. With such platforms, hospitals clinical data centers (CDRs) have been built to integrate.(Litjens et al. 2017, 66.) The original data are scattered in each clinical business system, including not only the patient's historical data, the patient's image data produced by medical instruments, electrocardiogram, the waveform data (temporal data) generated by varieties of instruments and equipment.(Allen et al. 2018, 318.) These data can be produced by a data center which is the medical big data. An information integration platform focuses on building a medical big data platform, and medical big data emphasizes data utilization.(Hu et al. 2015, 154.) How to mine valuable knowledge in big data and utilize the knowledge to provide clinical medical services for patients has become an essential challenge in the application of new electronic medical records. Similar chalenges include how this data is structured and how unstructured data is handled and what tools are generally used, how to build data models to analyse data, e.g. which models for structured data, which models for unstructured data, CNN, RNN, how to protect the privacy of such data and how the data are used for clinical services.(Faravelon and Verdier 2010, 204.)

Alibaba Health takes the user as its core and promotes medical e-commerce and new retail business across all channels, providing integrated offline and online solutions for the health industry, realizing the cross-regional shared allocation of existing social medical and health resources, greatly improving the convenience for patients to purchase medicines and meeting consumers' pursuit of a healthy lifestyle on the basis of ensuring professional safety. (Evans 2016,48.)

There is no doubt that electronic medical record applications can greatly reduce the burden on the healthcare system, but to truly implement online diagnostics, the rational use of medical big data and electronic medical records can offer much help. The objectives of this thesis was to discuss the basic characteristics of medical big data and the improvement of the application quality of electronic medical records, analyze the framework and analyze model of the medical big data platform and the application, privacy and security of electronic medical records. A literature review was carried out to achieve the objectives of this thesis.

2 THE MEDICAL BIG DATA PLATFORM ANALYSIS

2.1 Characteristic and benefits of big data

Big data refers to a collection of data that cannot be captured, managed, and processed with conventional software tools within a certain time frame, which has the characteristics of large data volume, rapid growth, complex structure, high value, low density and so forth. Hence, big data needs to be handled with innovative methods and technologies. In fact, the traditional relational database management cannot satisfy the requirements of big data processing. Medical big data not only has all the characteristics of big data, but also includes the features that are unique to the medical field, including temporality, privacy, and incompleteness(Lee and Yoon 2017, 6):

1. Volume: around150 MB for a CT image, about 750 MB for a genome sequence, about 5 GB for a standard medical record, and approximately TB to PB for a community hospital.
2. Variety: The types of clinical information data are complex, including the record-based structured data (EHR/EMR), the unstructured and semi-structured document data in the plain text or PDF format, DICOM-formatted image data as well as a new type of histological data.
3. Velocity: The development of information technology has led to the digitization of an increasing amount of medical information, and much of online or real-time data have continued to increase, such as clinical decision diagnosis, medication, and epidemiological analysis.
4. Value: The effective usage of medical data is conducive to the prevention and control of public diseases, accurate diagnosis and treatment, new drug research and development, medical control fees, intractable diseases, health management, etc.
5. Temporality: There is progress in the patient's treatment and the disease's pathogenesis; and the waveforms and images of medical tests are all functions of time.
6. Privacy: The patient’s medical data are highly private, and the disclosure of information may have severe consequences.
7. Incompleteness: Many records come from manual records, resulting in the incompleteness and deviation of data records. Incomplete collection and processing of medical data make medical databases unable to reflect the disease information fully.

The use of big data analysis and mining technology can help the medical industry to some extent to improve productivity, improve the level of care and enhance competitiveness. For example, comparative effect studies involving big data can improve the efficiency of medical personnel, reduce the cost of patient care and physical damage; in addition, the use of big data to monitor remote patients can also reduce the hospitalization time of patients and realize the optimal allocation of medical resources, in the process of disease prevention using remote monitoring systems can not only reduce the risk of accidents, but also save medical resources and create social and economic value.(Lee and Yoon 2017, 11)

2.2 Platform structure

There are several medical big data platforms built for hospitals and the key task is to build a hospital clinical big data analysis platform, integrate varieties of data modeling and mining assessment techniques, discover clinical diagnosis and treatment rules, as well as support medical decision-making and medical research. (HIStalk, 2018.) Various business data are generated by the hospital

clinical information system which needs to adopt appropriate data preprocessing, conduct data analysis and then establish the related business forecasting models. These models can improve the clinical business process and efficiency.(Litjens et al. 2017, 60) For instance, by utilizing the real-time statistical analysis, prediction and early warning in the application of electronic medical records in hospitals, one can often obtain an unexpected yield result while analyzing the behavior of patients as well as the clinical efficacy of drugs.

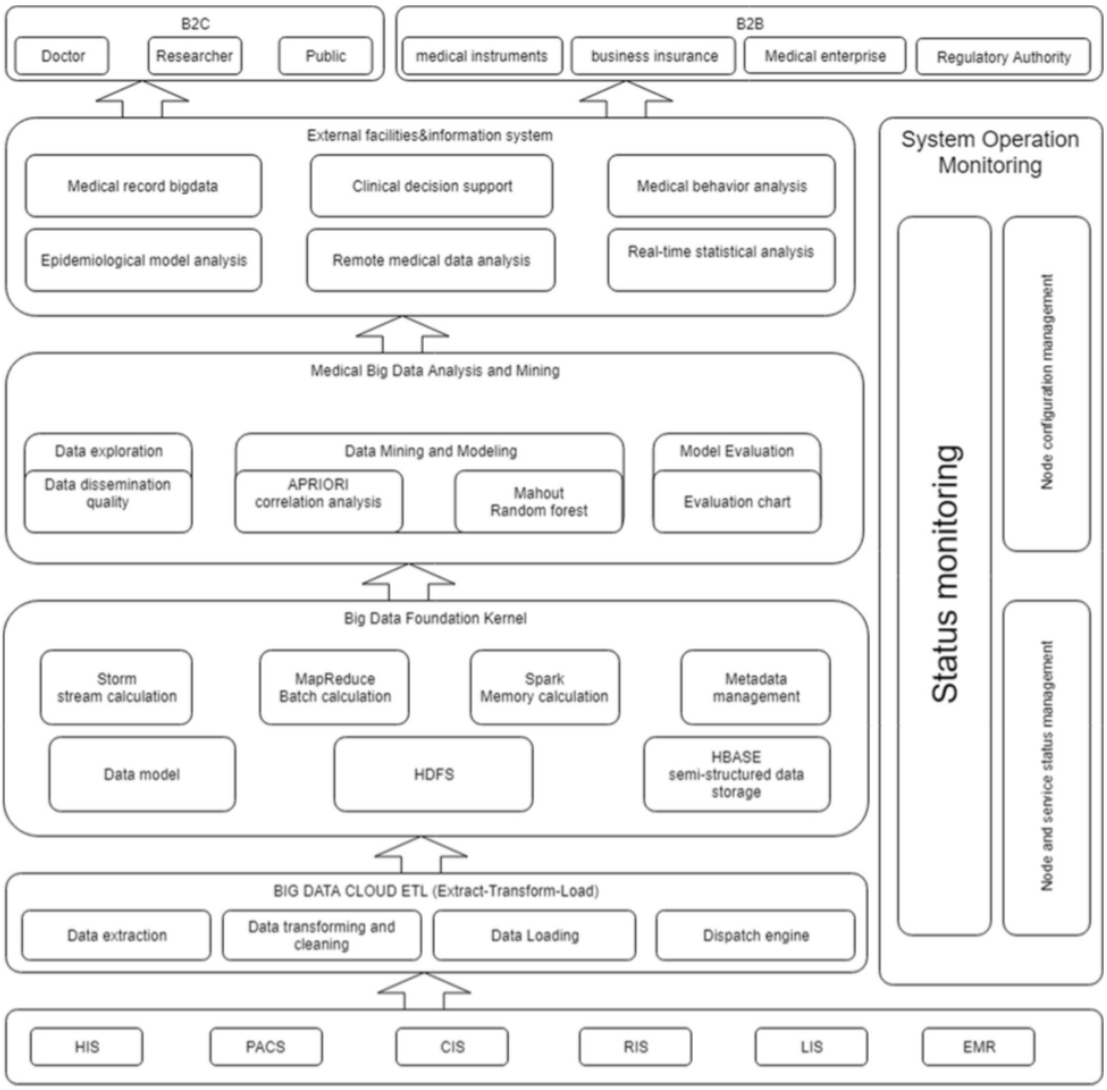

Figure 1. Medical big data analysis platform.(Lee and Yoon 2017, Allen et al. 2018)

Figure 1 indicates the entire structure of the medical big data analysis platform. The bottom layer is the data layer which integrates the data of the entire hospital clinical information system (Allen et al. 2018, 320). Subsequently, cloud computing can clean out some unnecessary data, such as vacancy data, noise data, duplicate data, inconsistent data, and incomplete data. Subsequently, according to the predefined data warehouse model, the data are loaded to the data warehouse. The Big Data Foundation Kernel will process the data stored in the data warehouse. This layer realizes the unstructured data storage architecture.(Bruijn et al. 2011, 557) Subsequently, the data are analyzed by the mining and analyzing model. These results provide real-time dynamic information services to the hospital business system as well as improve the level of hospital information system which can

help clinicians make medical decisions and managerial management decisions. When compared with the traditional clinical data centers, the big data analysis platform adopts the unstructured relational data storage and processing technologies.(Litjens et al. 2017, 68) In data storage, distributed unstructured and semi-structured data storage physical structures, such as Hadoop, are utilized. In data analysis and mining, analysis methods, for instance, correlation analysis and chart evaluation, are adopted. In particular, a series of new innovative models and algorithms are provided in the data visualization and mining algorithm models which are quite suitable for unstructured data processing.

The medical big data analysis platform involves several key technologies. The common technologies include data security patient privacy protection, massive data collection, information fusion technologies, and so on. The key technologies that highlight medical data characteristics include the two aspects as below.

2.3 Image data analysis

The first image data analysis method is image data analysis technology. Since image data are unstructured data, they must be pre-processed to generate an image feature library that can be utilized for high-level mining. Image data analysis solutions mainly include the function-driven models and information-driven models.(Hu et al. 2015, 153)

Function-driven models are organized by varieties of functional modules. Function-driven image data mining aims to design data mining solutions based on the specific requirements of specific applications, including the following(Allen et al. 2018, 318):

1. Image Acquisition Module which extracts image data from the image database;
2. Pre-processing module which extracts image features and store the feature information in the feature database;
3. Search Engine which matches queries utilizing image feature information;
4. Knowledge Discovery Module which performs algorithmic analysis of image data to discover the data's themes, characteristics, as well as relationships.

The information-driven model utilizes the mining algorithms and expertise to meaningfully segment the entire image and then performs the high-level calculations and mining analysis in order to derive an easy-to-use, easy-to-understand model with high-level semantics.(Faravelon and Verdier 2010, 204) The program divides the image information into four main levels:

1. Pixel layer which consists of raw image information and original image features, such as pixels, textures, shapes, and colors.
2. Object Layer which processes object and area information based on the original features of the pixel layer.
3. The semantic layer which combines professional knowledge to generate high-level semantic concepts from the identified objects and regions.
4. Knowledge layer which combines the textual and digital information related to a profession so as to discover the potential domain knowledge and patterns.

The techniques of image data mining mainly include image data preprocessing techniques, for example, denoising, contrast enhancement, and image segmentation, feature extraction and pattern techniques which contains supervised learning and unsupervised learning, such as classification, rule extraction, prediction, and clustering; (Bruijn et al. 2011, 557)

Classification, as a type of supervised learning, can be considered as the predictive modeling of output vectors or predictive variables that are classified. Classification means building a rule to assign an object to one of a pre-specified set of classes (predictors) based on the measurement vectors made on these objects. The classification techniques include logistic regression, naive Bayesian methods, decision trees, neural networks, Bayesian networks, and support vector machines. The classification technology flow based on the image data can be mainly divided into three steps:

1. Establishing an image representation model, performing feature extraction on the image samples that have already been labeled, as well as establishing attribute descriptions for each image.

2. Obtaining a high accurate considerable classification according to the training and learning of sample data set.

3. According to the classification model, the unmarked image data set is automatically classified and discriminated against.

Clustering refers to the unsupervised learning utilized to find groups in data with the usage of distance metrics. Clustering techniques include k-means clustering, principal component-based clustering, as well as self-organizing maps. Clustering performance can be assessed by its performance in the subsequent supervised learning tasks. Clustering is usually applied to the microarray data analysis or phylogenetic analysis, and it can also be utilized to redefine the diseases based on the pathophysiological mechanisms that provide more specific treatment options.(Litjens et al. 2017, 80) As a statistical analysis tool that quantifies the relationship between the dependent variable and one or more independent variables to describe the trend of the data, the regression supervision learning output variables are continuous. The clustering technology of image data is based on the distribution of image data without prior knowledge as well as classifies the image data without categorizing the marks into different meaningful clusters. Usually, four steps are included:

1. Image feature extraction and selection;

2. Establishing an image similarity model;

3. Trying different clustering algorithms;

4. Assessing the best grouping plan.

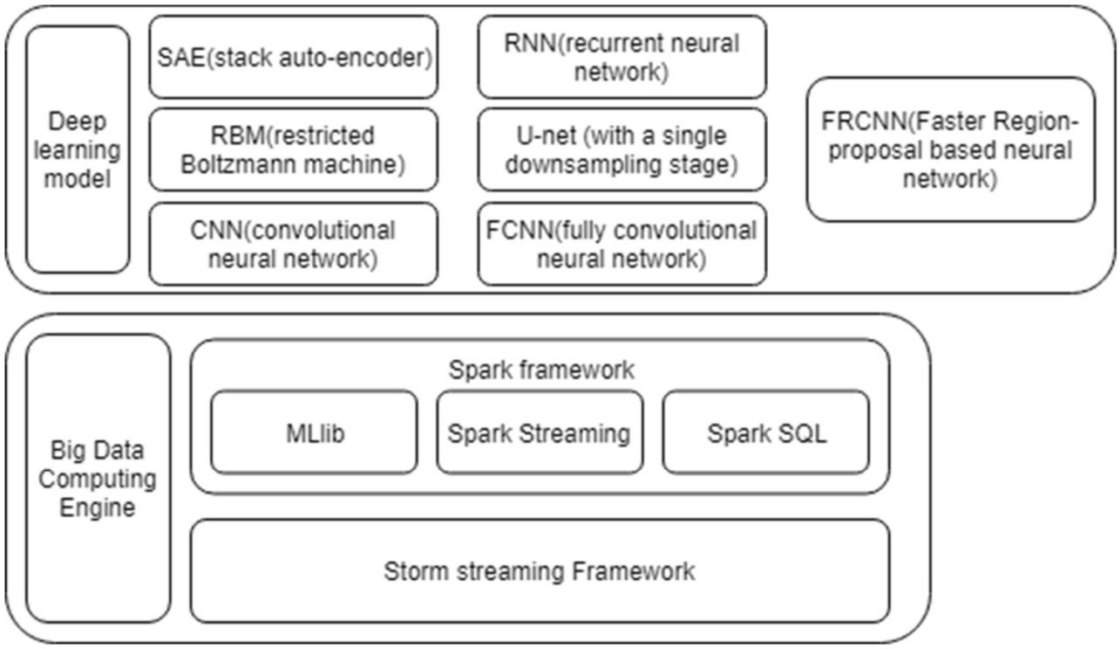

Figure 2. Image analysis structure.(Bruijn et al. 2011, Shen et al. 2017)

Figure2 illustrates one example of the image analysis structure which contains two parts, including the big data computing engine and the deep learning model. Two well-known frameworks can be chosen in the big data computing engine. The spark computing framework is different from MapReduce which will store the mediation data to disk after complete working.(Bruijn et al. 2011, 562) Spark utilizes the in-memory computing technology to analyze operations in memory while the data are not yet written to the hard disk. Spark's ability to run programs in memory can be up to 100 times faster than Hadoop MapReduce. Even when running programs on a hard drive, Spark can be up to 10 times faster. In addition, Spark SQL, Spark Streaming, and MLlib in the Spark kernel can help hospitals process data more efficiently.(Litjens et al. 2017, 88) The storm streaming framework has short delay while compared to the Spark, and it can protect the data loss.

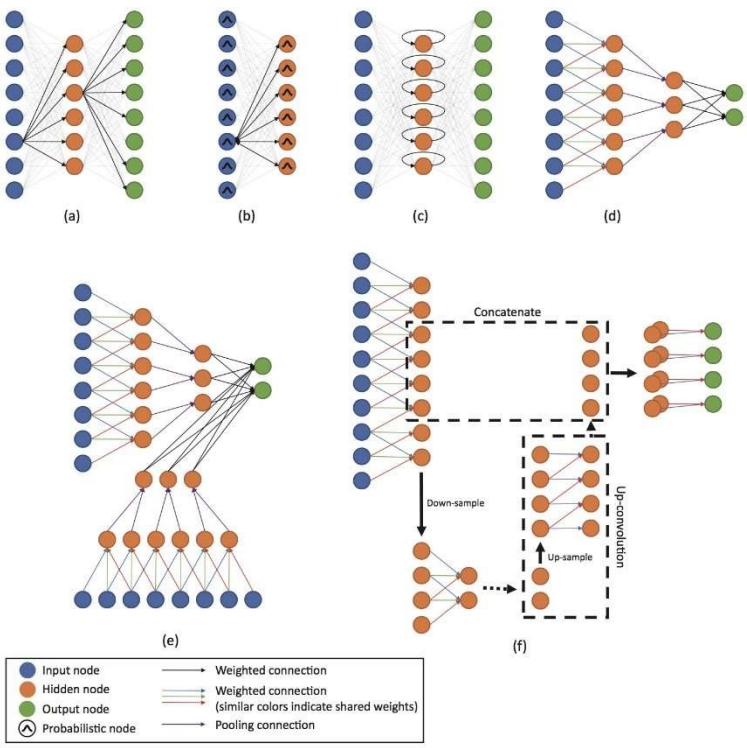

Figure 3. Deep learning model in medical image processing. (Litjens et al. 2017)

Figure3 presents the deep learning model framework for usage in medical image processing:

(a) SAE(stackauto-encoder) Unsupervised learning program, layer-by-layer training, and feature-based description
(b) RBM (restricted Boltzmann machine)
Unsupervised learning program, similar to SAE
(c) CNN(convolutionalneuralnetwork)
Convolutional neural networks are the most widely utilized to extract image features or perform tasks directly, such as classification detection.
(d) RNN (recurrent neural network)
Recurrent neural network can be applied in progressive scan images, such as CT, in order to obtain timing information.
(e) U-net(withasingledownsamplingstage) Similar to a full-web network with a short-cut, features utilized to fuse images at different scales.
(f) FCNN (fully convolutional neural network) Full-volume machine network can obtain the same resolution as the original picture, often utilized for tasks, such as segmentation.

(g) FRCNN (Faster Region-proposal based neural network) A fast-deep learning detection network framework can be divided into two layers, including RPN and RCNN, to detect varieties of objects in an image

Table 1. Examples in processing medical images.(Shen et al. 2017)

Type of image	Dataset	Goals	TOP1 method	TOP1 result
Brain CT	BRATS	Detection of segmented brain tumors	DL CNN+U-net+OriginalCT+ CNN[^doc6]	CT=0.87 core=0.81Enh ance=0.72
Eye CT	Diabetic retinopathy detection competition	Detect diabetic retinopathy	Fractional max-pooling	Score=0.8495 8
Chest CT	LUNA16	Detecting nodes in the chest radiograph	ZNET[^zfnet]	CPM=0.811
Pathology and micrograph processing	CAMELYO N16	determine whether a pathological	Patch normalize+inceptionv 3 network + tumor	AUC=0.9935
		section is a tumor	heatmap + Unicom domain method + RF classification	
Mammogram processing	DREAM challenge	X-ray film to judge breast cancer	Modified VGG +data augment +random crop	AUC= 0.8735
Heart CT	Second Annual data science Bowl	Calculate cardiac volume	U-net +deep learning	Score=0.0039 59
Liver segmentation	Sliver07	Liver location on CT images	Vascular-based semi-automatic segmentation algorithm	Score=85.7
Segmentation of the prostate	PROMISE12	Prostate location on CT images	CNN + U-NET+ RestBlock [^doc4]	Score =86.85

Table 1 shows some examples of different types of medical images which is supported in different deep learning model.

Thereby, the medical image big data mining analysis not only makes large-scale computer screening of specific diseases possible, which is beneficial to the improvement of clinical research level in hospitals, but also diagnoses the diseases through training models.

2.4 Text data analysis

The second key technolygy is the analysis of the unstructured free text data. A hospital electronic medical record contains a large amount of unstructured free text information. Natural language processing (NLP) is a core technique employed in the analysis of unstructured data.(Faravelon and Verdier 2010, 204) Utilizing NLP technology to extract and construct this narrative textual information is an essential step in the study of the secondary use of many EMR data. Natural language processing can also help address some aspects of incomplete data by increasing the amount of computational data.

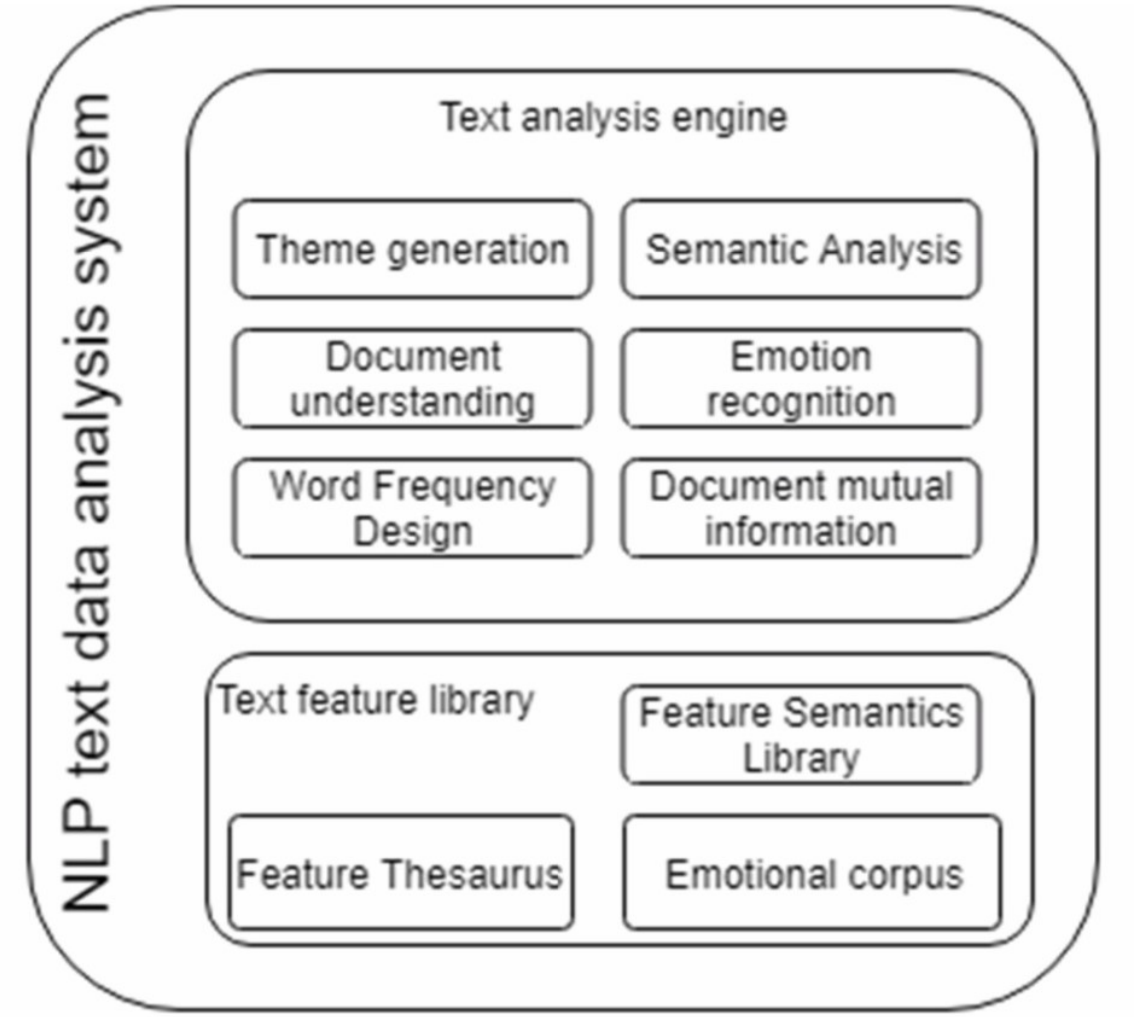

Figure 4. NLP text data analysis system. (Hu et al. 2015, 156)

Figure 4 shows the structure of the free text data analysis technology of the medical big data analysis platform. The text analysis engine which establishes a rich and professional feature database as well as the continued training and iterative update makes it more consistent with actual usage habits. In addition, it also defines reasonable semantic library rules as well as refines it automatically. Furthermore, according to the gradually established emotional corpora, the text sentiment is

analyzed.

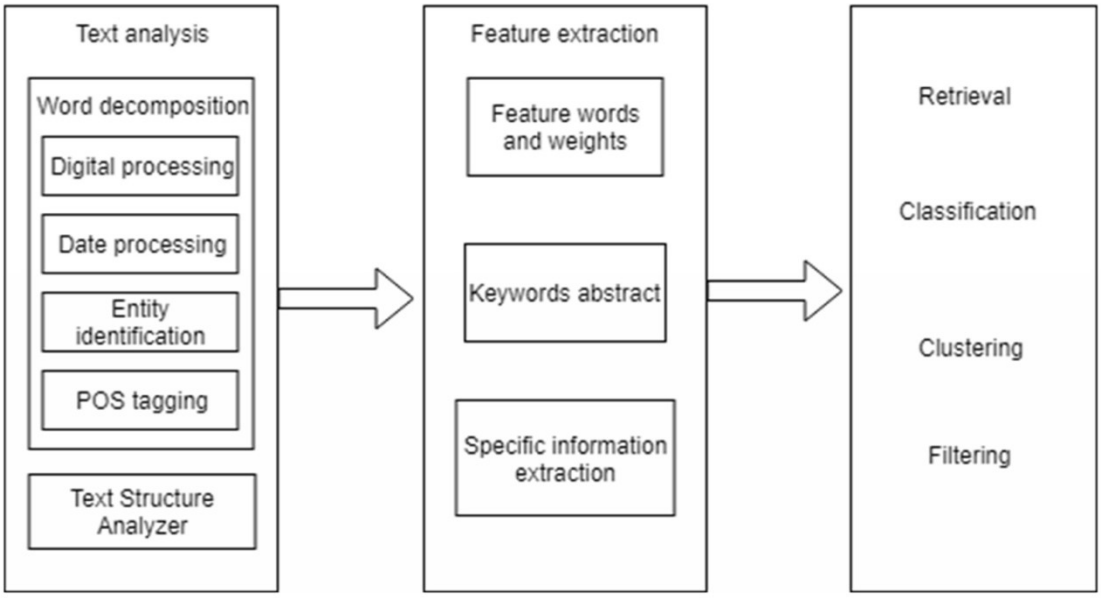

Figure 5. The process of the text analysis.

Figure 5 illustrates the entire process of the text analysis. The first step is the text analysis which performs an in-depth analysis of the text data, including word, grammar, semantics, and emotions. Subsequently, the feature extraction provides a learning framework for knowledge reasoning and language logic analysis. Last but not least, the data can be used for retrieval, classification or other analyses.(Faravelon and Verdier 2010, 208)

For the text feature library, it is not just a simple record and query, but deep learning for text through the text analysis algorithms. In addition to this, the text information of this feature library can also assist in the analysis of other systems.

Thereby, big data text analysis can help doctors to control the quality of clinical electronic medical records as well as improve the quality of electronic medical records in hospitals.

2.5 Security techniques

With the popularity of electronic medical record applications in hospitals, the security of electronic medical records is becoming more and more important. Medical personnel think the permanent retention of medical information in the hospital helps medical personnel to review previous patient visits, understand the condition and make more accurate diagnoses. And some patients sometimes forget to bring medical records or lose part of the information, with electronic medical records, not only the doctor is convenient, after seeing the disease themselves, but also can ask the doctor to

print the medical records, so that can clearly understand the contents of the medical records, without worrying about some doctors write medical records like "heavenly book" can not read. However, other people consider that these computers, as long as they have access to the Internet, can chat, play games, speculate on stocks, watch movies, shop, etc., in addition to doing their normal work, which creates a major loophole in the security of medical records. Despite the accuracy, efficiency and versatility of electronic medical records, some potential problems remain unavoidable. Privacy security is probably the number one threat to the security of electronic medical records, with a large number of data breaches containing personal privacy.(Evans 2016, 48) In recent times, Amazon has been exposed to a breach of about 47GB of medical data, which accidentally exposed at least 150,000 patients' blood test results, medical records management records, etc. In addition to hacking, the daily usage habits of medical staff pose security risks to patients' medical data. So what are the security hazards of electronic medical records? Such as:

1. Hacking of electronic medical record systems can result in the alteration of patient data or the destruction of clinical systems.
2. Misuse of health information files by authorized users of the electronic medical record system.
3. The electronic medical record system faces long-term data management problems.
4. Illegal government or corporate involvement in private health care.

In order to avoid these safety hazards, measures should be taken to prevent them. To protect patients' privacy and avoid tampering with their medical records, the authentication method of electronic signature is adopted; medical and nursing staff are required to insert a key when filling out and accessing electronic medical records; and blocking is not allowed to use web applications, such as P2P, IM, stock trading, games and other web applications.(Yang et al. 2018, 16) The security of electronic medical records is a matter of public importance, and through continuous improvement and refinement, it is possible to facilitate the work of health care workers and to keep patient information confidential.

3 ELECTRONIC MEDICAL RECORD APPLICATION

3.1 Intelligence inquiry application system

Due to the advent of the information age, medical records have been transformed from traditional paper-written medical records into a template to document the editing electronic medical records. The e-health, clinical pathways, antibiotic management, and mobile healthcare can be commonly found in the hospital. Nonetheless, the obvious lack of clinical knowledge library construction has led to insufficient personalization of hospital electronic medical records, which means that there are not sufficient decisions to support the doctor.(Kupwade Patil and Seshadri 2014, 762) The medical big data analysis platform lays the foundation for the dynamic and real-time utilization and the improvement of the clinical knowledge base as well as creates favorable conditions for the construction of intelligent electronic medical records.

At the beginning, the construction of the electronic medical record emphasizes on the process operation information system only and realizes the medical-driven information-based application based on the process. Nonetheless, in some other fields, such as stock and finance, they have their decision system, which can help people to make the decision automatically. The clinical electronic medical records and the electronic collection of data make data analysis applications increasingly popular, and the medical procedures can be tracked through analysis to improve the medical defects.(Yang et al. 2018, 18) The medical big data platform and the electronic collection of data make the applications of data analysis increasingly popular, and the medical procedures can be tracked through analysis to improve the medical decision system. Data analysis turns electronic data into meaningful information that improves the quality of patient care.

It is a significant goal of medical big data and electronic medical record application to conduct a hierarchical analysis of medical big data, form clinical knowledge rules, serve electronic medical record application, as well as achieve the clinical application intelligence. It is also the direction and inevitable trend of electronic medical record application in the future.

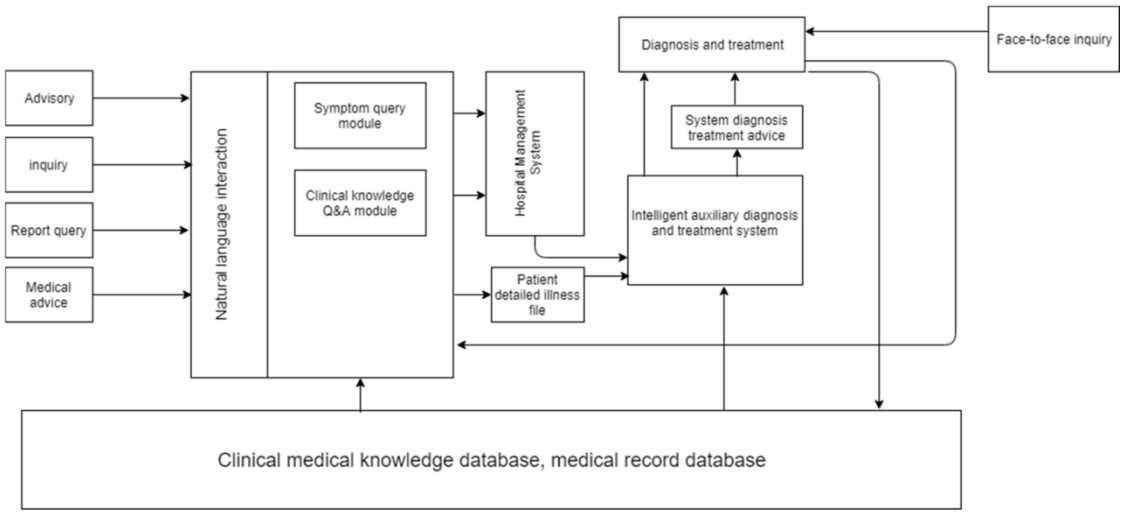

Figure 6. Intelligent inquiry application system structure diagram.(Yang et al. 2018)

Figure 6 indicates that based on the intelligent inquiry application system structure, the Big Data knowledge base and the patient's electronic medical records are utilized to implement patient consultations and consultations, and meanwhile to provide supplementary advice to doctors for diagnosis and treatment.

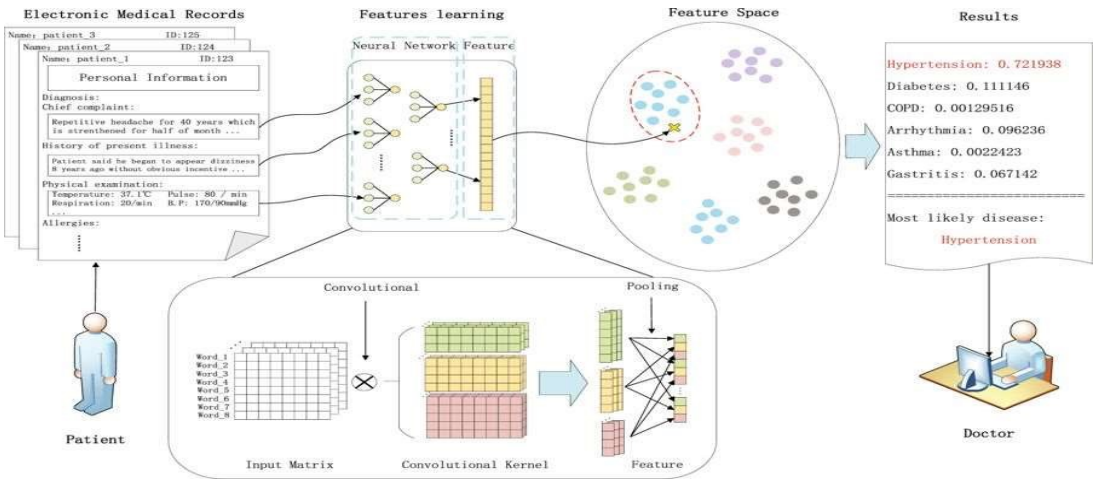

Figure 7. The whole process for advising. (Yang et al. 2018, 22)

Figure 7 shows one example of whole process that electrical medical records provide diagnosis advice for the doctor using CNN model. With the rapid development of medical artificial intelligence technology, similar clinical information applications will become increasingly common.

3.2 Using Big Data to Patient Health Status

In 2009, only 1.5 percent of the more than 3,000 U.S. hospitals with complete electronic medical records and health information technology (HIT) systems were surveyed, and these hospitals are largely large teaching hospitals located in large cities. Only 4% of them have complete electronic medical records in their clinics and doctors' offices. At the time, the United States was ranked 8th in the ranking of countries with access to health information archives, followed by New Zealand, Australia, the United Kingdom, Italy, the Netherlands, Sweden and Germany.(Evans 2016, 66) The greatest benefit of an electronic medical record is that it can be accessed frequently and can be readily documented in surgical records, the basis for each examination, hospital discharge reports, medical personnel contacted, etc. It's not just about storing medical files, it can directly avoid a lot of problems, like allergies. The information collected by the medical information technology system not only allows patients to log in and retrieve their own test data and test results, but also allows them to monitor the early signs and symptoms of an impending epidemic of infections and monitor them, as well as the side effects of newly marketed prescription drugs. So the role of an integrated national, interregional and even global health IT system is self-evident.

In 2003, 87-year-old Florence Rothman underwent conventional valve replacement for cardiac, aortic stenosis as well as showed good health at discharge. She was sent to the emergency department four days later, and then was declared dead after a series of complicated operations. Her two sons, Michael Rothman and Steven Rothman, closely worked with medical staff and hospital administrators

to learn how to avoid an unstable state like their mother after the patient left the hospital. They were very surprised to find that the hospital had a complete electronic medical record system with a large amount of data accumulation. All these data covered the patient's information on the hospital's medical treatment process, which was sufficient to identify the patient's health and remind the doctor of the patient's discharge. With the strong support of the hospital, they focused on the electronic medical record data analysis and hoped to find an algorithm model utilizing the current EMR data to track the health status of patients. Subsequently, it could provide doctors with meaningful information indicators in the future treatment of patients. After several years of hard work, they proposed the Rothman Index (RI), and establishing a patient readmission risk prediction model is not a new concept. Most models relied on the retrospective analysis of medical quality evaluation in which the data could be obtained after the patient was discharged from the hospital. They utilized the data of the patient during the hospital to establish a model with the purpose of guiding the patient's clinical care in real time.(HIStalk 2018) This work can benefit from the development of two technologies, including electronic medical records and data mining. The electronic medical record application extends the patient's clinical information and the patient's basic information range as well as makes the data more real-time. Herein, data mining is a general method to calculate and analyze a great number of discrete data models.

Developing clinical risk prediction models faces many challenges and difficulties. There are many types of data in hospital EMRs, such as patient test reports and free text records in the electronic disease records, which constitutes the variables needed to construct a risk prediction model. A good patient health prediction model requires the usage of objective data from clinical electronic medical records in order to determine the patient's health risk factor as well as provide a reference for medical personnel. Meanwhile, the hospital should classify the risk factors and classify the clinical information of high-risk patients to help the hospital react quickly in the emergency.

An accurate risk prediction evaluation algorithm can help improve the improvement of products related to electronic medical records. The model building process itself makes the electronic medical record system more efficient. And the data mining analysis will help identify various risks of the patient during the hospital through the electronic medical records system, which is difficult to achieve in the traditional paper records.

3.3 Application Scenario

1. Hospital management decision support: By mining the clinical data, operational data and material data of the hospital, it solves various problems in hospital management, improves the efficiency of equipment use and reduces the operating costs of the hospital. Through data analysis, it realizes the management of performance, medical insurance, pharmacy, outpatient, inpatient, surgery, etc., and monitors the operation status of the hospital in real time, and provides the basis for decision support for the development direction and operation of the hospital. The big data analysis summarizes the problems of the hospital and gives solutions to reduce hospital costs and increase hospital revenue.
2. Health management: Achieving the management of healthy people through the analysis of data, so that people do not get sick and less sick, is the ultimate direction of medical big data applications. With the help of the Internet of Things, smart medical devices, smart wearable devices, real-time collection of residents' health data, through the monitoring of physical signs data, health management is realized.
3. Pharmaceutical research and development:Through the medical and pharmaceutical big data,

using the algorithm system of artificial intelligence deep learning ability, the analysis of different compounds and chemical substances in the development of drugs, predict the safety, effectiveness, side effects in the drug development process, can effectively reduce the drug development costs, shorten the development cycle and reduce drug prices.

4. Chronic disease management: The management of chronic diseases usually takes place outside the hospital, through intelligent terminals, data management systems, mobile medical devices and medical health applications, to achieve network access to a number of test data, as well as intelligent monitoring and tracking of patient behavior and medication records. Through data monitoring, it is possible to know the patient's current physical condition and whether he/she is taking his/her medication as prescribed. The chronic disease management type of medical big data enterprise, its data may come from clinical care organizations, or it may come from the accumulation of smart devices used by patients.
5. Smart old age: There is a combination of the areas of smart aging and slow disease management, but smart aging also focuses on healthy older adults. Companies in the field of aging are still relatively shallow in the application of big data, most of them collect the physical data and status of the elderly through smart wearable devices or other sensors, and then evaluate and monitor the physical condition of the elderly through data. (Shakyawar et al. 2019, 58)

In the medical field, big data has a wide range of applications, including disease prevention, clinical applications, Internet medicine and other aspects. It can be said that medical big data is the future development trend of the medical field. At present, in the application of big data in the medical industry, China is still in the initial stage, the government, hospitals and data mining technicians need to work together to make big data in the medical field.

4 SECURITY IN ELETRONIC MEDICAL RECORDS

In modern hospitals, electronic medical records are widely used, which has completely changed the traditional medical treatment model, realized the transformation from paper medical records to electronic medical records, and the hospitals keep these medical records uniformly. The various medical order data and diagnosis data stored in these electronic medical records provide direct data support for scientific research and teaching. In addition, as healthcare professionals seek various possible ways to reduce costs while improving healthcare processes, delivery and management, big data emerges as a viable solution and is expected to transform the healthcare industry. This transition from passive medical care to active medical care may lead to an overall decline in medical costs and ultimately economic growth. While the healthcare industry is harnessing the power of big data, as emerging threats and vulnerabilities continue to grow, security and privacy issues become the focus.

4.1 Security in Healthcare

Adoption of big data in healthcare significantly increases security and patient privacy concerns.

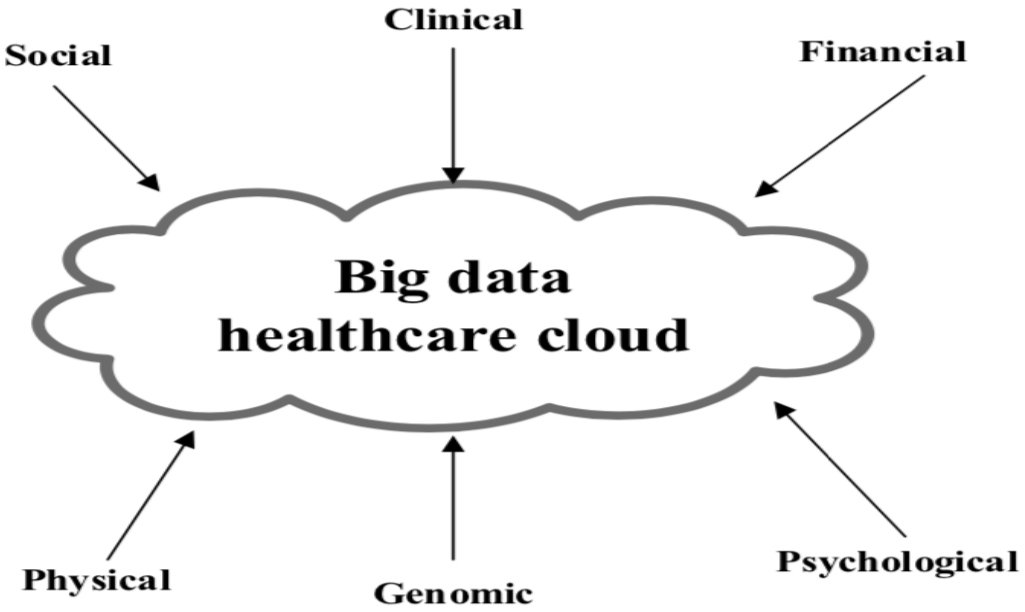

Figure 8. Big data healthcare cloud. (Kupwade Patil and Seshadri 2014)

Figure 8 depicts a big data medical cloud that hosts clinical, financial, social, genomic, physical, and psychological data related to patients.

Traditional security solutions cannot be directly applied to large and inherently diverse data sets. With the increase in popularity of healthcare cloud solutions, complexity in securing a mass of distributed Software as a Service (SaaS) solutions increases with varying data sources and formats. So, big data governance is necessary prior to exposing data to analytics.

4.2 Research status of privacy protection

1. Privacy anonymity protection: In the specific patient diagnosis and treatment file, the patient ’s name and certificate are usually used numbers, etc. serve as unique identifiers, but the information itself should be protected, so the correct approach should be to protect the information anonymously without affecting the accuracy of the information. In response to this idea, some researchers further proposed an anonymous method to protect identity during the publication of privacy-protected data, that is, first delete the identity mark in the data to be published, and then anonymize the identification data. In this way, the accuracy of information can be further improved on the basis of protecting privacy, and in the implementation of this method, two methods such as generalization and lossy connection can be used.(Abouelmehdi et al. 2018, 42.)
2. Medical data classification protection system: Different information has different weights in privacy protection. Therefore, if all information is generalized and high-level protection methods are adopted, the actual application efficiency will be affected and resources will be wasted. However, if only the core information is protected, it may also cause hidden information leakage. Therefore, it is necessary to build a relatively complete data grading system, and adopt different protection measures for different levels of personal information and data.(Shakyawar et al. 2019, 54-68.)
3. Privacy protection based on access control: In the information system of the medical field, the difficulty of privacy protection is mainly due to the large number of participants, which leads to a corresponding increase in potential leakage points. Using access control technology, you can set different access rights for different people, and then access different data and content, which can also effectively solve the problem of data classification. For example, for personnel in the finance department, they should only be able to access patient-related information and not view the doctor's diagnosis information. The access control technology widely used nowadays is mainly role-based access control, which can control the access content and the operation authority. However, in the implementation of rule setting and authority grading, the specific process and means are relatively complicated, and it is difficult to achieve unified authorization through unified rule setting. In many cases, it is necessary to set up separately for special situations, and it is difficult to achieve overall management and adjustment. Therefore, further research on the specific application of the rule engine in the medical field is needed. (Abouelmehdi et al. 2018, 24)

We mainly reviewed the privacy preservation methods that have been used recently in healthcare and discussed how encryption and anonymization methods have been used for health care data protection and presented their limitations.

5 CONCLUSION

This thesis started with introducing the basic characteristics of medical big data as well as the development of electronic medical records, then analyzed the framework and analysis model of the medical health big data platform, and finally analyzed the application and security of electronic medical records.

The application of electronic medical records is facing new tasks, such as medical material sharing, regional sharing, mobile closed loop and personal privacy leakage. It is necessary to utilize a big data platform and related technologies to enhance the intelligence level of electronic medical record application. The developmental direction of the electronic medical records is the construction of clinical knowledge base which realizes the intelligent and precise medical treatment. Big data analysis is an auxiliary means to enhance the quality of electronic medical record application. The significant direction of clinical integration platform (data center) construction aims to establish a medical big data analysis application platform.

The establishment of a hospital data clinical center requires the usage of big data technology and infrastructure to build an inclusive and shared patient clinical information platform with the purpose of achieving hospital clinical information system data collection and information services. The application of electronic medical records is a long-term task, and intelligence is an essential direction of clinical information.

REFERENCES

Abouelmehdi, K. Beni-Hessane, A. and Khaloufi, H. (2018), "Big healthcare data: preserving security and privacy", Jounal of Big Data, vol.5, pp. 20-42.

Allen, W. Gabr, R. Tefera, G. Pednekar, A. Vaughn, M. and Narayana, P. (2018), "Platform for Automated Real-Time High Performance Analytics on Medical Image Data", IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 2, pp. 318-324.

Dash, S. Shakyawar, S. Kumar Sharma, S. Sharma, M. and Kaushik, S.(2019), "Big data in healthcare: management, analysis and future prospects", Journal of Big Data, vol.6**,** pp 54-58.

de Bruijn, B. Cherry, B. Kiritchenko, S. Martin, J. and Zhu, X. (2011), "Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010", Journal of the American Medical Informatics Association, vol. 18, no. 5, pp. 557-562.

Evans, R. (2016), "Electronic Health Records: Then, Now, and in the Future", Yearbook of Medical Informatics, vol. 25, no. 1, pp. S48-S61.

Faravelon, A. and Verdier, C.(2011), "Towards a Framework for Privacy Preserving Medical Data Mining Based on Standard Medical Classifications", in E-Health, Bˆatiment IMAG C 220 rue de la chimie, 38400 Saint Martin d'H`eres, pp. 204-211.

Hu, Y. Lin, W. Tsai, C. Ke, S. and Chen, C.(2015), "An efficient data preprocessing approach for large scale medical data mining", Technology and Health Care, vol. 23, no. 2, pp. 153-160.

Kupwade Patil, H. and Seshadri, R.(2014), "Big Data Security and Privacy Issues in Healthcare," 2014 IEEE International Congress on Big Data, Anchorage, AK, pp. 762-765.

Lee, C. and Yoon, C. (2017), "Medical big data: promise and challenges", Kidney Research and Clinical Practice, vol. 36, no. 1, pp. 3-11.

Litjens, G. Kooi, T. Bejnordi, B. Setio, A. Ciompi, F. Ghafoorian, M. van der Laak, J. van Ginneken, B. and Sánchez, C.(2017), "A survey on deep learning in medical image analysis", Medical Image Analysis, vol. 42, pp. 60-88,.

Rothman Healthcare Corporation and HIStalk (2018), "HIStalk Interviews Michael Rothman ", Histalk2.com. [Online]. Available: https://histalk2.com/2010/10/25/histalk-interviews-michael-rothman-president-rothman-healthcare-corporation/.

Shen, D. Wu, G. and Suk, H.(2017), "Deep Learning in Medical Image Analysis", Annual Review of Biomedical Engineering, vol. 19, no. 1, pp. 221-248.

Yang, Z. Huang, Y. Jiang, Y. Sun, Y. Zhang, Y. and Luo, P.(2018), "Clinical Assistant Diagnosis for Electronic Medical Record Based on Convolutional Neural Network", Scientific Reports, vol. 8, no. 1, pp.16-28.