|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-A-015 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | N/A | Geneva, 25-27 September 2018​ | |
| **DOCUMENT** | | | | |
| **Source:** | | National Institute of Metrology (China) | | |
| **Title:** | | Suggestions on data quality evaluation of AI for health | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Wenli Liu National Institute of Metrology China | | Tel: +86-010-64525051 Fax: +86- 010-64525057 Email: [liuwl@nim.ac.cn](mailto:liuwl@nim.ac.cn) |
| **Contact:** | | Shan Xu Ministry of Industry and Information Technology (MIIT) China | | Tel: +86-010-62300338 E-mail: [xushan@](mailto:xushan@)caict.ac.cn |
| **Contact:** | | Chengwei Li National Institute of Metrology China | | Tel: +86-010-64525053 Fax: +86- 010-64525057 Email: [lichw@nim.ac.cn](mailto:lichw@nim.ac.cn) |

|  |  |
| --- | --- |
| **Abstract:** | AI for health is inseparable from medical big data. Only when medical health data is accurate and reliable can we ensure the validity and reliability of diagnostic results or services obtained from AI technology. We recommend that data quality evaluation for health artificial intelligence should focus on three areas, including quality control of medical devices, standardization of medical big data, rationality and effective use of medical big data. Equipment control should be carried out through first verification and daily quality control. With the effectiveness of medical big data, we began to establish and improve the standard system and strengthen the training of data collectors. Developing comprehensive, uniform and reasonable international standards for different types of AI medical services and supervision of medical big data is the final and also key part in the chain. |

**I. Background**

The 21st century is the era of artificial intelligence. With the development of information technology, artificial intelligence and machine learning technology, abbreviated AI and ML, is shorted for is gradually applied to many aspects in our daily life, among which medicine and health is one of the most important application areas. AI technology in the field of medicine and health is inseparable from medical big data, and technologies such as machine learning require medical big data as its calculation objects and basement. Various medical institutions, medical information platforms, digital diagnosis and treatment equipment, smart health electronic products, wearable personal health equipment, and health information platforms are the main sources of medical big data. Medical big data also includes various data types such as physiological parameters, medical images, and personal health information.

Recently, many research institutes and enterprises in China have invested a lot of energy in the research of AI technology for medicine and health, and several kinds of AI service have been supplied to the public. These kinds of service include electronic medical record management systems, medical device intelligent management systems, medical image cloud, and personal health management and so on. All the services are based on medical big data. Only when medical health data is accurate and reliable can we ensure the validity and reliability of diagnostic results or services obtained from AI technology.

In China, AI for medical purpose must get the third type of medical device registration certificate from China Food and Drug Administration (CFDA) before they can be used in hospital. The third type of medical device is the one which has the lowest safety risk. The corresponding AI service provider needs to submit necessary technical documents to prove its safety and effectiveness according to the requirements of CFDA. However, how to evaluate the safety and effectiveness of software? How to regulate the legitimacy and accuracy of these services used in hospitals? All these problem require corresponding standards.

**II. Suggestion**

We believe that the quality evaluation of medical big data in AI technology needs should be carried out focused on three aspects.

1. **Quality control of medical equipment**

In order to ensure that the medical big data is accurate and reliable, the measurement accuracy of the medical equipment used to obtain the data must be guaranteed first. The quality control of these equipment should be carried out by two steps.

1. **First verification or calibration**

Any medical equipment used for acquiring big data should be verified or calibrated before being used. In this circumstance, the verification or calibration for a certain equipment should be made by professional metrological technician referring to domestic or international verification, calibration specifications or testing standards. The parameters which do not meet the requirement must be adjusted and calibrated.

1. **Quality control in daily use**

Medical big data are cumulative results, so the medical equipment used must show consistent performance during the period of use. The medical equipment operators need to undertake relevant training, and should have certain skills related to calibration, equipment adjustment to monitor the operation status of the equipment. Once the performance of medical equipment does not meet the requirement for acquiring accurate medical big data, verification or calibration in the first step should be performed immediately.

It is believed that quality control of medical equipment can be started from the following aspects.

1. Pilot study on quality control of medical equipment for the most important physiological parameters and medical images in medical big data

* Physiological parameter: China has performed some research on calibration techniques of measurement equipment or quality control device for several physiological parameters such as blood oxygen saturation, blood pressure and blood sugar, and has developed corresponding standard instrument. The next stage in our research is to integrate calibration methods of these important physiological parameters for the purpose of realizing fast, real-time, on-line monitoring and calibrating, and ensuring the accuracy of measurement results from these equipment.
* Medical image: For CT, MRI, PET/CT and DEXA equipment, calibration of this machines and corresponding domestic verification regulation or calibration specifications has been finished. The verification regulations or calibration specifications for these equipment contain evaluation criteria for patient safety and imaging performance, but there is no specific operation mode for calibration. , the international standard mostly adopts subjective and visual methods are commonly adopted in international standards for quality evaluation, which makes the evaluation result random and introduce human error into it. In order to evaluate the performance of medical image equipment objectively and accurately, we intends to solve the problem in two ways.
* Develop automatic and objective evaluation software for different medical image equipment. After obtaining the quality control images, they should be clarified and evaluated automatically through some certain image analysis algorithm.
* Cooperate with medical image equipment manufacturers to develop new products with self-calibration capabilities. We plan to research on self-calibration technologies such as by using built-in phantoms to let equipment adjust and calibrate by itself, and reduce human error introducing by the procedure of calibration.

1. Improve medical staff’s awareness of quality control for medical big data. The acquisition of medical big data is inseparable from the medical staff who actually operate the medical equipment.

In order to improve the awareness of calibration and quality control for medical equipment in different levels of hospitals, we proposed a variety of ways to enhance the cognitive level of medical staff.

* Participate in or organize conference about medical big data quality control. We intend to clarify the basic concepts of quality control for medical big data, and preach about the importance of measurement and quality control technology in ensuring the accuracy of medical big data.
* Cooperate with domestic and international medical management departments. By cooperating with government, we intend to draft quality control standard of medical big data such as real-time online detection of physiological parameters and objective automatic evaluation of medical image.
* Train domestic metrological staff. China relevant institutes will preach about verifications and quality control standard to domestic metrological staff of all levels as well as organize some training programs for them. These domestic metrological staff will help their local hospitals with medical equipment calibration and quality control for medical big data by using metrological technologies.

1. **The consistency and effectiveness of medical big data**

The accuracy and reliability of medical big data is the basis for realizing AI technology, but the consistency and effectiveness of medical big data directly determine the pros and cons of AI calculation results.

Consistency contains multiple levels of meanings. For example, the same type of medical data should have the same or similar data format. It also means that the same type of medical data should be obtained under the same or similar conditions and parameters.

Effectiveness also includes multiple levels of meaning. It means that medical data should meet the requirements from AI technology. And it also indicate that the information contained in the data can be useful in AI technology.

In order to explain how to ensure the consistency and effectiveness of medical big data more thoroughly, we proposes to start the research from the following aspects.

1. **Establish and improve a standard system for medical big data**

A unified standard system of medical big data should be proposed and recognized by majority of the committee members. These standards should include data collection equipment standards, data collection process standards and data transmission and preservation standards.

1. **Strengthen the training for medical big data collectors**

In addition to setting up operation procedures and standards, we believe that training for health care big data collectors should be strengthened. The data collectors may come from hospitals, universities and other scientific research institutions as well as AI technology companies. Even using the same equipment to acquire medical data, different people may get slightly different data in the parameters and so on. This kind of slight difference may result in less accuracy of AI results. In order to reduce the differences caused by human factors, we suggest that there should be more trainings for the medical big data collectors, and we should verify the quality of every certain collector’s data regularly.

1. **Standardization, reasonability and effectiveness in usage of medical big data**

After obtaining accurate, consistent and effective medical big data, how to use these data standardized, reasonably and effectively is the last but not the least important part in the chain of AI technology. Scientific and unified evaluation criteria should be developed according to different application scenarios.

In this situation, we believe that group can start the work from the following aspects.

1. **Develop comprehensive, uniform and reasonable international standards for different types of Ai medical services**

Currently, a variety of AI medical services have appeared in the Chinese market, such as medical image AI diagnosis, personal health management, and accurate diabetes prediction. Among them, medical image AI diagnosis is the most widely used AI technology. Medical image AI diagnosis is an auxiliary diagnostic method designed to help doctors screen and identify possible lesions. According to some statistics, the diagnostic accuracy of these various AI products is claimed to be more than 95%. Among all these products, the AI diagnosis of lung nodules in CT image is the most typical one. As to the AI diagnostic results for lung nodules, the false positive rate of is high and there is also a certain ratio false negative results.

To summarize, we believe that there are three main problems in medical image AI diagnosis.

1. The accuracy of the image data. To solve this problem, it requires training and quality control for medical image equipment, data acquisition process, and data collector.
2. The accuracy of the doctor's diagnosis. The results of AI technology is based on the comparison of image feature recognition technology and doctor's diagnostic results. So the correctness of the doctor's diagnosis must be guaranteed. Different doctors should have a unified evaluation standard for the same image as well as the same doctor has a unified evaluation standard for different images.
3. The accuracy of the diagnostic algorithm. 95% accuracy does not mean that AI diagnosis can replace the doctor. According to the feedback from many doctors on the use of AI diagnostic technology, it is generally believed that the accuracy of this technology must reach 100% before it can be considered as a usable diagnostic tool. The accuracy rate of 100% here means that all suspected lesions must be identified, but false positives are allowed. However, no single technology can achieve 100% accuracy. How to establish the availability criteria for AI diagnosis needs to be determined by participants from different fields.

Combined with the existing experience, we believes that the focus group can consider the following suggestion in the future standard setting.

* Conduct extensive contacts and cooperation with hospitals and AI service providers to develop quality control standards for different types of medical big data collection;
* Conduct pilot research and work with a number of hospitals in a certain region to develop domestic diagnostic criteria for different medical data;
* Publish data validation sets for different types of medical health AI services, and use the data sets to verify the diagnostic accuracy of AI service.

1. **Supervision of medical big data**

Medical big data contains patient's personal information, health information and so on. It can directly reflect the health condition of a specific population in a country. Medical big data should be regarded as an important data of strategic significance.

In this situation, we believes that the focus group should propose an effective regulatory program for the use of medical big data.

1. Privacy in use of medical big data. In the process of using medical big data, it should be paid attention to protect patient’s privacy at first. Unnecessary personal information should be excluded during data collection and use. At the same time, each country should set up medical big data management regulations according to their national conditions and strictly manage the data types and purpose for using it.
2. Security in use of medical big data. AI services based on medical big data are mainly divided into medical one and commercial one.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_