



BIG DATA CENTER SUPPORTS RESEARCH AND INNOVATION

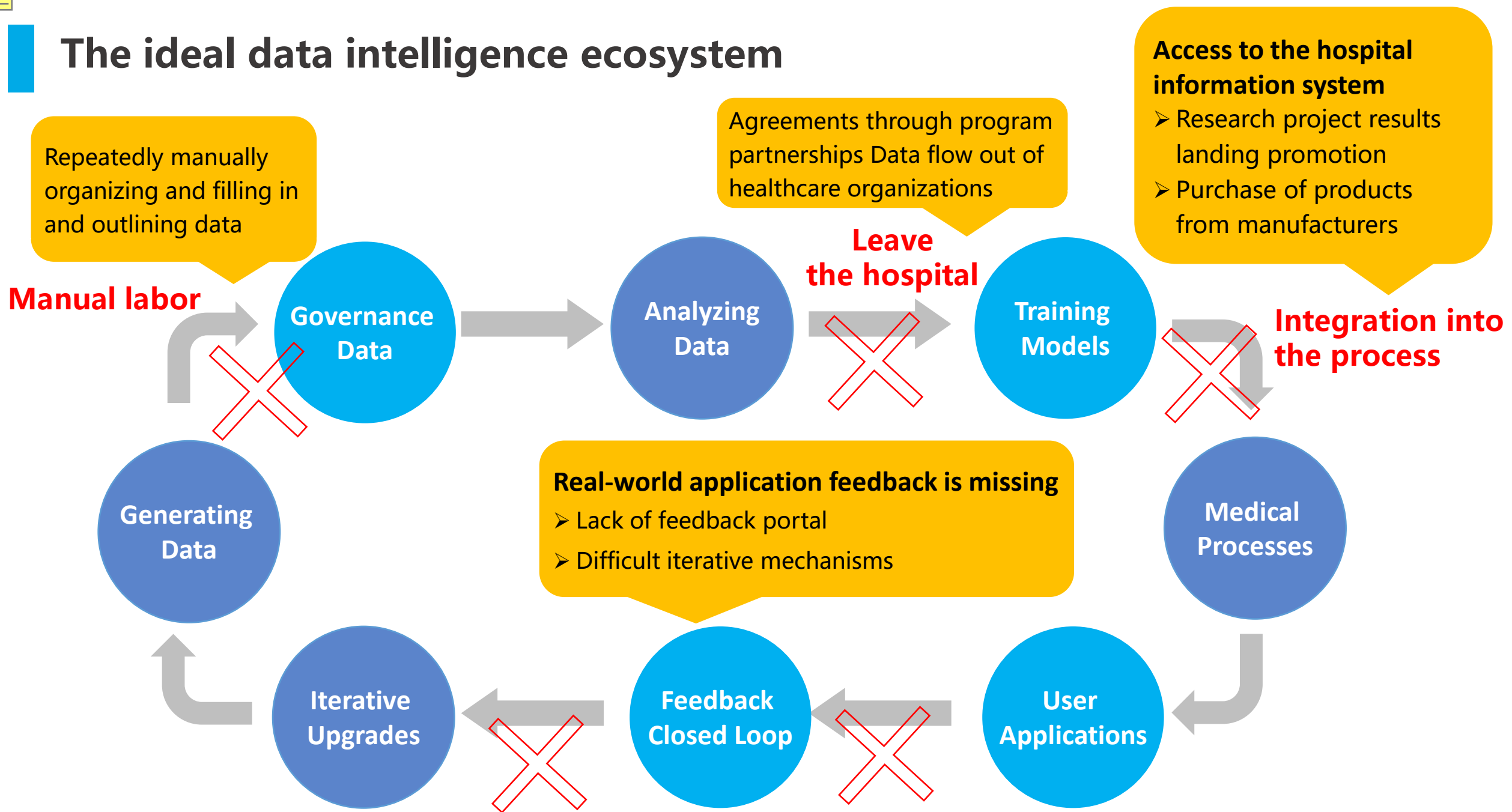
Introduction to Construction Programs and Ideas

Data create the value of medical service

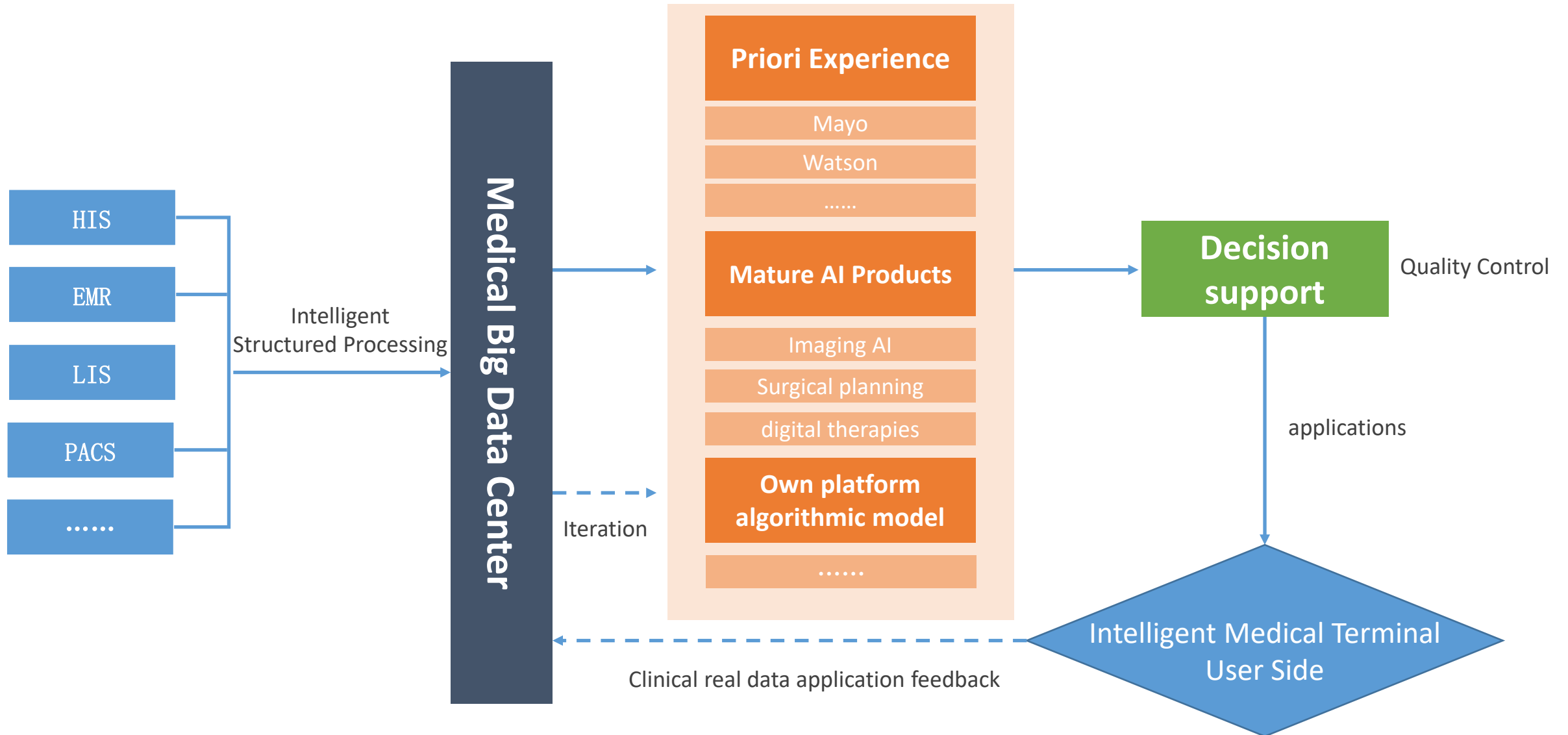
Biohub IT Innovation & Renovation Co Manufacturer



The ideal data intelligence ecosystem



Ideal in-hospital data governance — training — application flow system





01

PART 01

PLATFORM BUILDING PROGRAM

Big Data Center and Data Intelligence Platform Construction Path

1. Establishment of a common hospital-wide data lake platform

Capture;
Splitting;
Cleaning;
Sorting;
Parsing;
Correlation;

.....

4. Forward-looking topics data collection

Missing structured data after the consultation cannot be traced;
Relying on mobile terminals for post-diagnosis structured data governance, overlaying missing data before structured filling;
Supporting prospective project development and in-hospital follow-up data collection;

.....

2. Big data retrieval, export and descriptive analysis

Search and Query;
Funnel filtering;
Conditional correction;
Generating data sets;
Exporting data requests;
Data export approval;

...

5. Scientific research projects or clinical trials conducted

Data statistical mining analysis;
Data quality evaluation;
Data governance and preprocessing;
Data labeling;
Model training and iteration;

.....

3. Establishment of a fine-grained disease-specific database

Adoption of standards, codes, and expert consensus;
Medical record writing habit browsing and validation;
Natural language processing learning and recognition;
Algorithm optimization and validation, calibration;
Specialized disease library parsing;

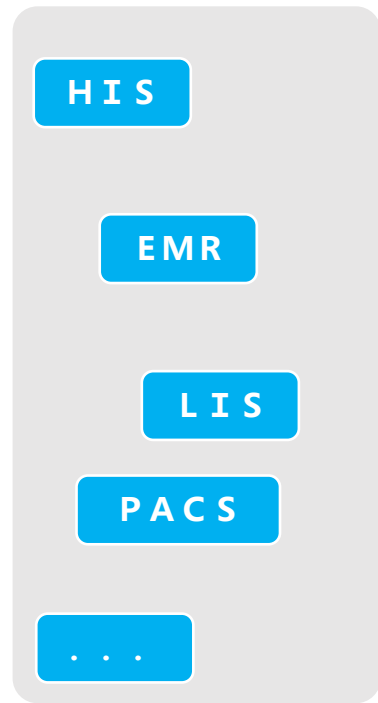
.....

6. Integration of models into business processes to form CDSS (AI-enabled)

Iterated model release;
Integration into corresponding scenario business processes;
CDSS prompts and user confirmation feedback;
Algorithms and data form a closed loop of feedback;

.....

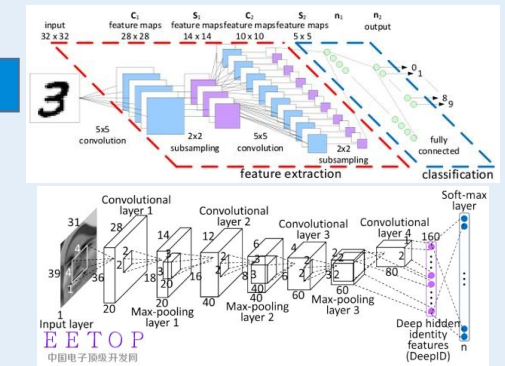
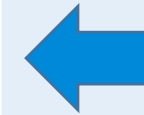
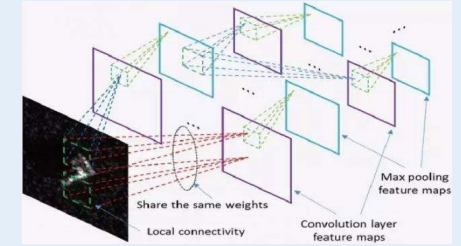
The main idea (multimodal artificial intelligence AI training platform that Intelligent training covering the entire business process)



- Healthcare Data Integration
- Medical Data Governance
- Medical Imaging Training
- Introduction of algorithmic frameworks
- Deep physician engagement
- Clinical Algorithm Requirements Definition
- Hospital-centered building of AI intelligence to assist integration into CDSS



Physicians can be deeply involved



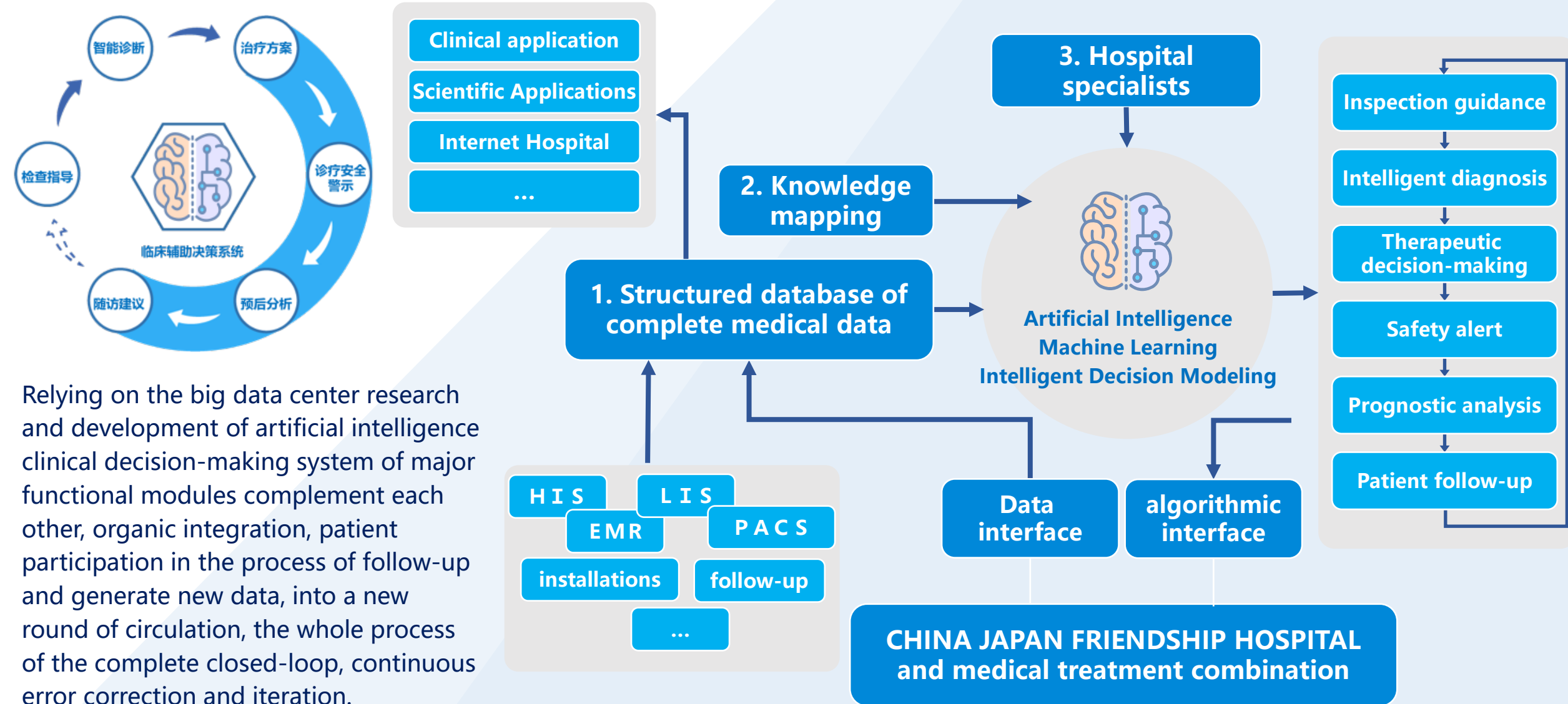
Integration of algorithms, models, etc.

- *Have data sources within the institution
- *Have data application scenarios
- *Have the need for on-the-ground application of AI models
- *Lack of AI algorithms applicable to the organization's own data
- ***Lack of talent and mechanisms to consistently produce algorithms** that meet the clinical and research needs of the organization



Third-party results matching

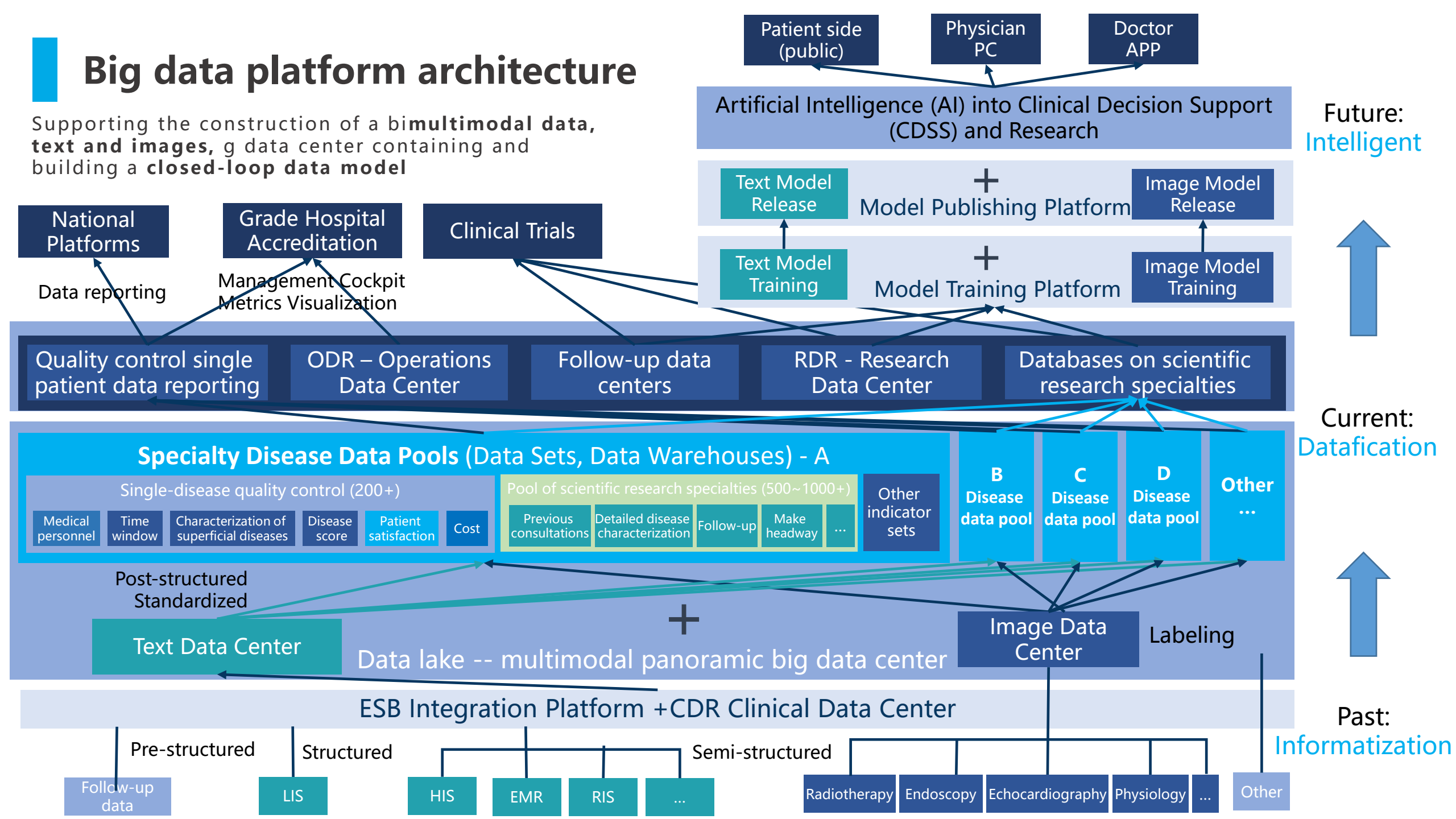
Key concepts (Fine granularity disease-specific database, AI model training platform for multiple medical stages)



Relying on the big data center research and development of artificial intelligence clinical decision-making system of major functional modules complement each other, organic integration, patient participation in the process of follow-up and generate new data, into a new round of circulation, the whole process of the complete closed-loop, continuous error correction and iteration.

Big data platform architecture

Supporting the construction of a **bimultimodal data, text and images**, g data center containing and building a **closed-loop data model**





02

PART 02

Platform Functional Modules and Project Experience

Platform Functional Modules

1. Establishment of a common hospital-wide platform



2. Big data retrieval, export and descriptive analysis



3. Establishment of a fine-grained disease-specific database



4. Data collection on prospective subjects

5. Scientific research projects or clinical trials conducted

6. Integration of models into business processes to form CDSS (AI-enabled)



Lessons learned from project construction

Data Reach!

Easily searchable, the time cost of obtaining data is dramatically reduced.

In the face of massive hospital data, after researchers have good ideas and thoughts, they can screen the required target cases in a very short period of time through the big data platform using combined query or funnel query.

Isolated? Disconnected?

Most of the hospital programs, where data is derived from CDRs, have poor data quality availability in the CDRs themselves. Lack of inclusion of multimodal data makes integration of full text this center and big imaging platforms difficult.

Assets? Junk?

Historical data quality is poor and should be targeted to develop specialty databases.

Combined with the high quality of the follow-up system, the whole process of integrating data. Combined with the structured management of data in the treatment, the system will no longer produce garbage since the launch.

Data-> Modeling-> Application-> Closed loop-> Iteration ->

Models are integrated into day-to-day operations, with scientific research focused on data insights and algorithmic evaluation.

Lessons learned from project construction

Data Reach!

Easily searchable, the time cost of obtaining data is dramatically reduced.

In the face of massive hospital data, after researchers have good ideas and thoughts, they can screen the required target cases in a very short period of time through the big data platform using combined query or funnel query.

Isolated? Disconnected?

Most of the hospital programs, where data is derived from CDRs, have poor data quality availability in the CDRs themselves. Lack of inclusion of multimodal data makes integration of full text this center and big imaging platforms difficult.

Assets? Junk?

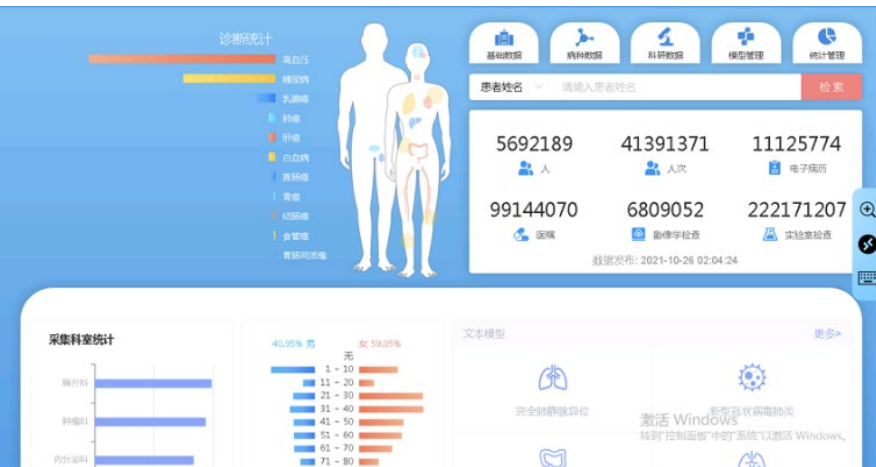
Historical data quality is poor and should be targeted to develop specialty databases.

Combined with the high quality of the follow-up system, the whole process of integrating data. Combined with the structured management of data in the treatment, the system will no longer produce garbage since the launch.

Data-> Modeling-> Application-> Closed loop-> Iteration ->

Models are integrated into day-to-day operations, with scientific research focused on data insights and algorithmic evaluation.

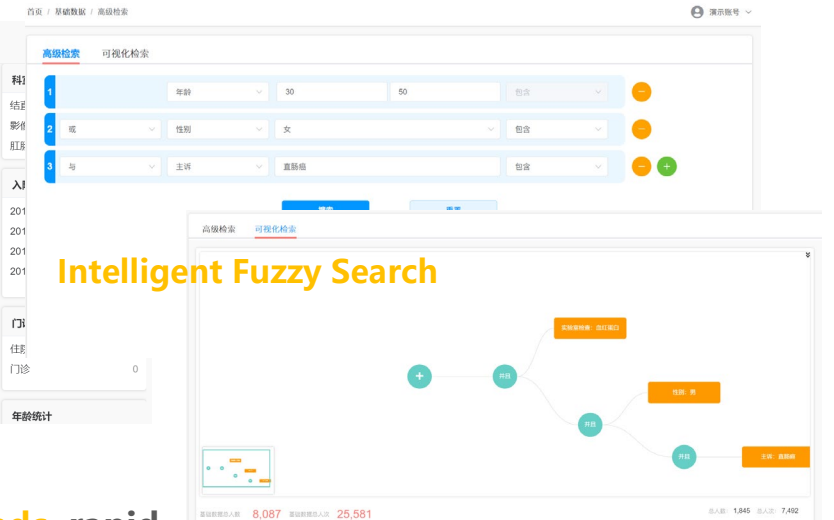
Supporting the Information Center and carrying out the construction of a common big data infrastructure platform for the whole hospital



The dashboard displays various patient statistics and a human body diagram. On the left, there's a '诊断统计' (Diagnosis Statistics) section with a bar chart. In the center, there's a '患者姓名' (Patient Name) search bar and a table of statistics. On the right, there's a '采集科室统计' (Collection Department Statistics) section with a bar chart. The bottom right shows a '文本模型' (Text Model) section with a bar chart.

患者姓名	5692189	41391371	11125774
人	99144070	6809052	222171207

Multi-dimensional precise condition search



Intelligent Fuzzy Search

Visual Search

Dominance : Cross-business system hospital-wide multimodal data query accurately in seconds, rapid verification of mining needs and research ideas

Example: Quick query for patients with “triple positive” (LIS) breast cancer (HIS & EMR), a history of comorbid diabetes or hypertension (EMR), and imaging showing lung metastases (PACS)

Searchable

In the face of the hospital's massive data, after the researchers have good ideas and thoughts, they can screen the required target cases in a very short time through the big data platform using combined query or funnel query.

Easy to fill

Based on the foundation of the previous data governance, it is possible to better carry out the research process on specific topics, and to facilitate the completion of additional medical records, scales, CRF forms and so on.

Easily interpreted

Get maximum data intuition. Use intuitive visualization techniques to turn numerical information into intuitive, graphical image information that is easy for researchers to observe, simulate, and calculate. Discover new associations, new structures between medical data.

Easy-to-navigate

It ensures easy access to in-depth research on relevant departmental topics, greatly reduces the cost of initiating research and improves the efficiency of conducting research, focusing on data insights and algorithmic evaluation.

Lessons learned from project construction

Data Reach!

Easily searchable, the time cost of obtaining data is dramatically reduced.

In the face of massive hospital data, after researchers have good ideas and thoughts, they can screen the required target cases in a very short period of time through the big data platform using combined query or funnel query.

Isolated? Disconnected?

Most of the hospital programs, where data is derived from CDRs, have poor data quality availability in the CDRs themselves. Lack of inclusion of multimodal data makes integration of full text this center and big imaging platforms difficult.

Assets? Junk?

Historical data quality is poor and should be targeted to develop specialty databases.

Combined with the high quality of the follow-up system, the whole process of integrating data. Combined with the structured management of data in the treatment, the system will no longer produce garbage since the launch.


Data-> Modeling-> Application-> Closed loop-> Iteration ->

Models are integrated into day-to-day operations, with scientific research focused on data insights and algorithmic evaluation.

Support for the simultaneous establishment of **big data text centers and multimodal data centers**



Support the establishment of multimodal data centers and modeling platforms

**中山大学孙逸仙纪念医院**
SUN YAT-SEN MEMORIAL HOSPITAL, SUN YAT-SEN UNIVERSITY

科研数据中心
RESEARCH DATA REPOSITORY

基础数据

专病数据库

专病质控

数据导出

eCRF

外部数据库

公共数据库

公共库列表

生物医学库

基因数据库

数据库类型：

数据库类型

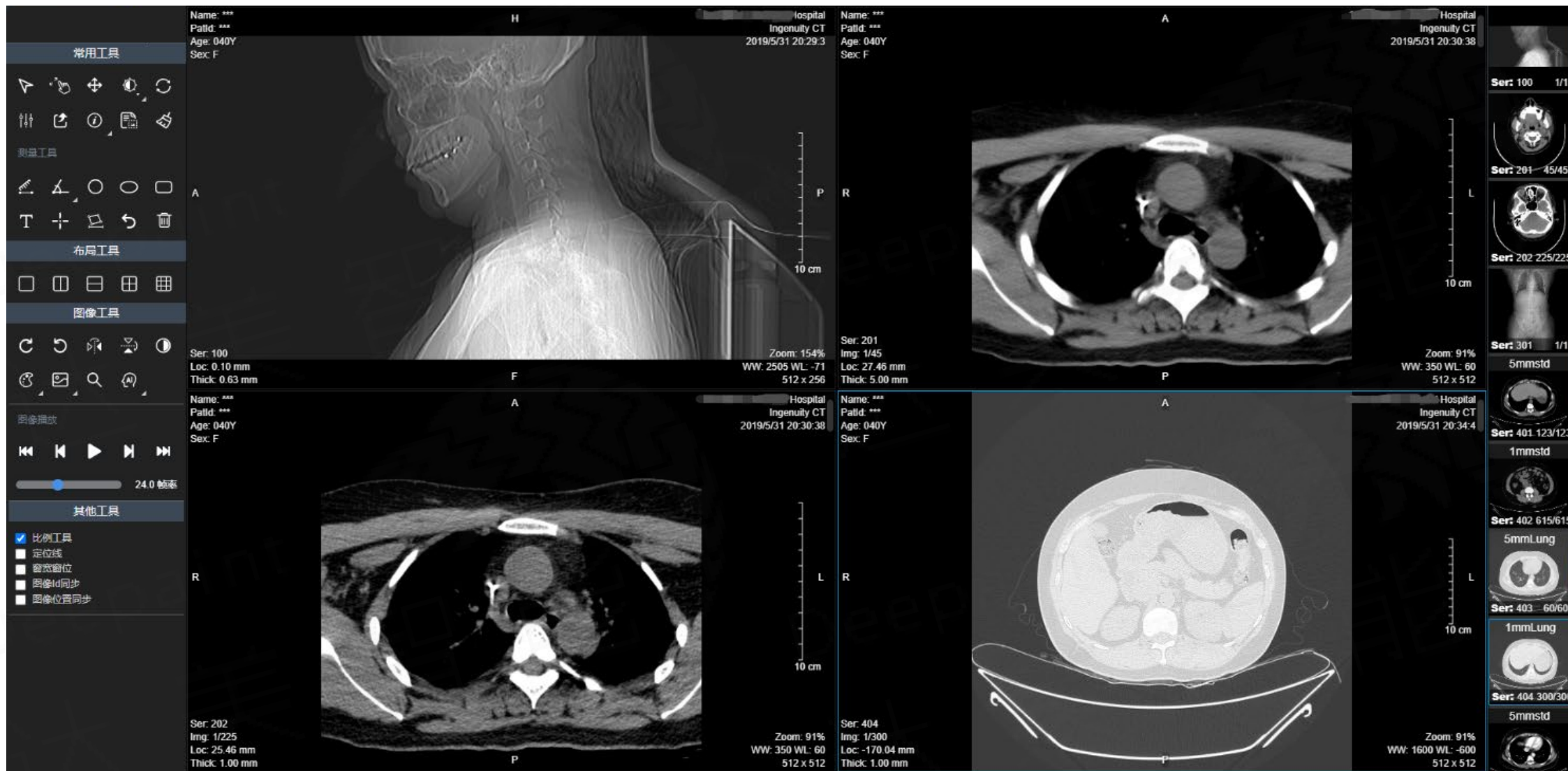
设备类型：

设备类型

查询

数据库	数据集	癌症类型	数据信息	设备类型	参与者	研究	系列	数据格式	数据量(G B)	内网链接 (包括元数据)	操作
TCIA	QIN-Breast	Breast cancer		MR, PET, CT	68	214	530	Dicom	11.286	/home/lihanwei/...	详情
TCIA	ACRIN-FLT-Bre...	Breast cancer		CT, PT, OT	83	363	1480	Dicom	74.2	/home/lihanwei/...	详情
TCIA	BREAST-DIAG...	Breast cancer		MR, MG, CT, PT	88	148	429	Dicom	60.8	/home/lihanwei/...	详情
TCIA	Breast-MRI-NA...	Breast cancer		MR, SEG	64	189	2602	Dicom	19.5	/home/lihanwei/...	详情
TCIA	ISPY1	Breast cancer		MR, SEG	222	847	9032	Dicom	76.2	/home/lihanwei/...	详情
TCIA	RIDER Breast ...	Breast cancer		MR	5	10	40	Dicom	0.39	/home/lihanwei/...	详情
TCIA	TCGA-Breast	Breast cancer		MR, MG	139	164	1,877	Dicom	88.1	/home/lihanwei/...	详情

Support the establishment of **multimodal data centers and modeling platforms**



Lessons learned from project construction

Data Reach!

Easily searchable, the time cost of obtaining data is dramatically reduced.

In the face of massive hospital data, after researchers have good ideas and thoughts, they can screen the required target cases in a very short period of time through the big data platform using combined query or funnel query.

Isolated? Disconnected?

Most of the hospital programs, where data is derived from CDRs, have poor data quality availability in the CDRs themselves. Lack of inclusion of multimodal data makes integration of full text this center and big imaging platforms difficult.

Assets? Junk?

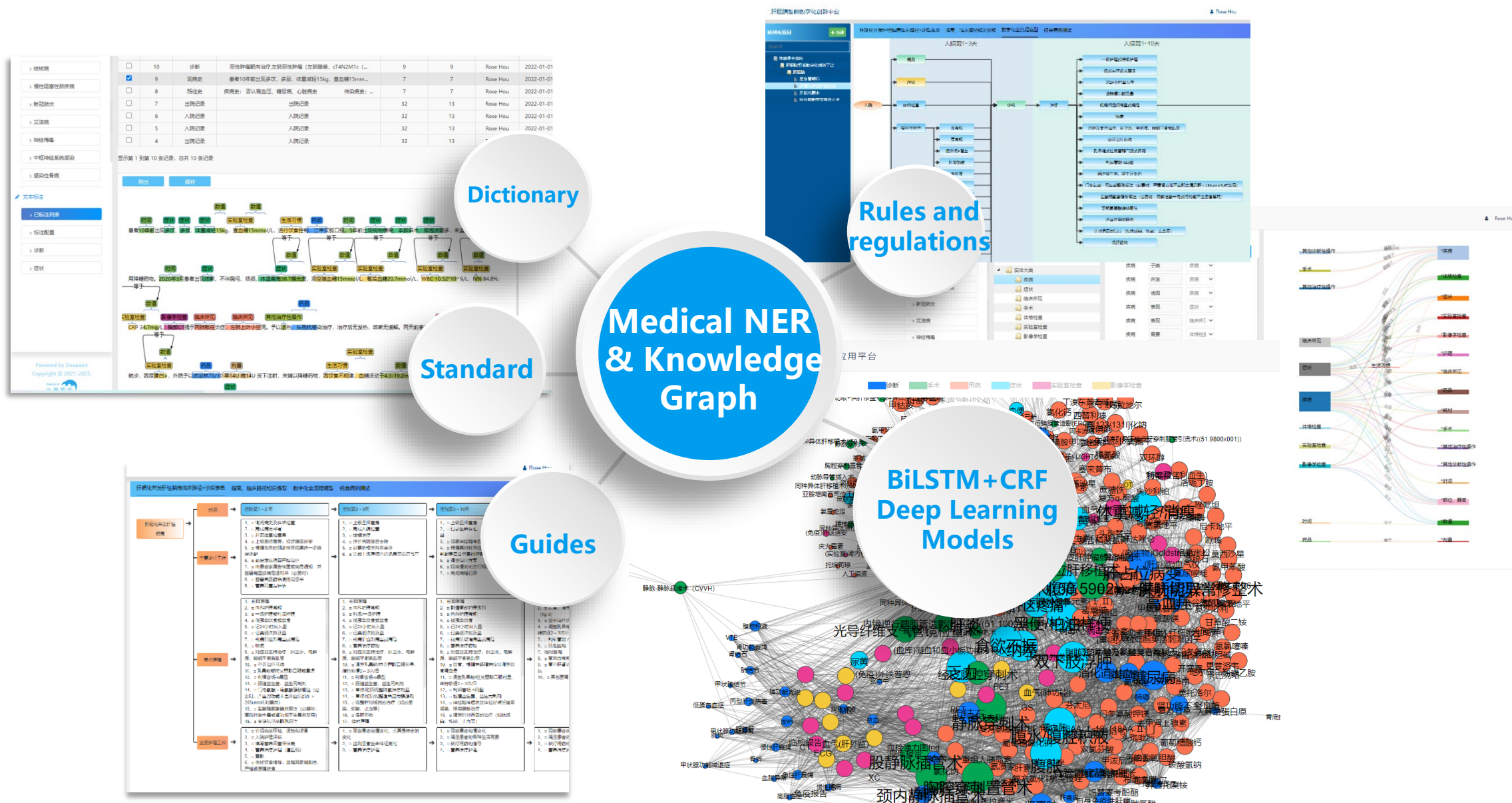
Historical data quality is poor and should be targeted to develop specialty databases.

Combined with the high quality of the follow-up system, the whole process of integrating data. Combined with the structured management of data in the treatment, the system will no longer produce garbage since the launch.

Data -> Modeling -> Application -> Closed loop -> Iteration ->

Models are integrated into day-to-day operations, with scientific research focused on data insights and algorithmic evaluation.

Natural semantic processing based on medical noun entities



Text Data - NLP Natural Semantic Processing

入院情况: 宫颈鳞癌。诊断依据: 1.主诉: 阴道流血量多3月, 动脉灌注化疗后15天。 2.本科检查: 外阴: 已产式, 阴道: 通畅, 见少量血液, 宫颈: 宫颈4-5点处见菜花样组织, 范围约2*1cm, 侵及左侧穹隆, 未达阴道壁中1/3, 接触伴有出血, 子宫前位, 稍增大, 无压痛, 质地硬, 双附件区未及包块无压痛。 3.辅助检查: 6.10我院阴超: 子宫大小73*50*56mm, 节育器位置宫腔内, 宫腔分离4mm, 子宫后壁见无回声, 大小14*8mm。血常规: 血红蛋白54g/L; 心电图: 103次/分。 6.11本院住院肝胆B超: 肝内血管瘤14*12mm, 肝内胆管结石: 9*6mm。心超: 左室舒张功能欠佳, 双肾输尿管正常。 6.28门诊复查血肝肾功能正常, 血常规: WBC 3.3×10^9 , HGB 98g/l, 尿常规: 隐血2+, 胸部CT: 1.两肺少许纤维灶, 两肺纹理增强, 建议结合临床及隔期复查。 2.附见: 肝右叶结节状略低密度结节灶, 直径约15mm, 建议增强CT进一步检查, 核酸阴性。 轻度贫血: HGB 98g/l。肝血管瘤: CT: 肝右叶结节状略低密度结节灶, 直径约15mm, 6.11本院住院肝内血管瘤14*12mm。肝内胆管结石: 6.11住院B超9*6mm。 诊疗经过: 患者入院 完善相关检查, 排除手术禁忌症, 于7.8行 腹式广泛全子宫+双侧附件切除术+卵巢动静脉高位结扎术+盆腔淋巴结清扫术, 术中见: 子宫前位, 略显饱满, 双侧卵巢输卵管未见明显异常。宫颈部膨大, 约3.5*4.5cm, 宫旁组织增厚。膀胱反折处组织致密, 界限不清。探查腹膜内表面、肠管表面、大网膜、膈下、肝脏表面未及异常病灶。探查腹主动脉旁, 未及明显肿大淋巴结。遂行腹式广泛全子宫+双侧附件切除术+卵巢动静脉高位结扎术+盆腔淋巴结清扫术。剖视标本, 见宫颈肿瘤侵犯阴道穹隆, 未达阴道下1/3。手术过程顺利, 患者生命体征平稳, 术中出血约150ml, 术中保留导尿管, 尿色清。术后留置腹腔引流管一根, 安返病房。术后恢复好, 7.26-7.28行术后化疗(多西他赛+卡铂方案), 现一般情况可予出院。考虑宫颈癌IIA2期, 建议同时放疗。

Value、Unit、Time、Diagnosis/Lesion、检验、检查、Anatomical site、Seen/Described、Surgery、operation、chemotherapy、drugs

Research Big Data Center, Natural Semantic Processing

Strengths: Core field parsing completeness and correctness can be debugged to **98%** or more within **24 hours**

Medical Natural Semantic Recognition Technology

Simulating the logic of human brain to recognize language, intelligent parsing of text, the subject, predicate, object and complement of text statements can be parsed into structured data of "dimension = or ≠ value", and the data will be automatically entered into the database.

Combining the hospital's own case templates and the department's customized clinical knowledge points to add a proprietary clinical dictionary, it performs intelligent semantic analysis of **all text reports: admission records, medical records, surgical records, medical orders, image reports, pathology reports, etc.**, mines all the clinical information, and ultimately produces structured as well as quantitative data. It can realize "Medical Name Entity Recognition", "Name Entity Automatic Coding", "Name Entity Modifier Recognition", "Time Information Extraction" etc.

Base layer

Basic information such as gender, age, department, diagnosis, time, etc.

Time layer

Integrate data from patients' previous consultations to create individual big data in a timeline

Logic layer

Categorize medical data by disease type and build logical links around key events of the disease type such as surgery, treatment, etc.

Detail layer

Key fields can be extracted from different data sources to realize accurate and deep mining of complex information.

Efficiently

It takes about 2 hours to parse 1,000 past EHRs ;
New electronic medical records can be analyzed in real time.

Approximate

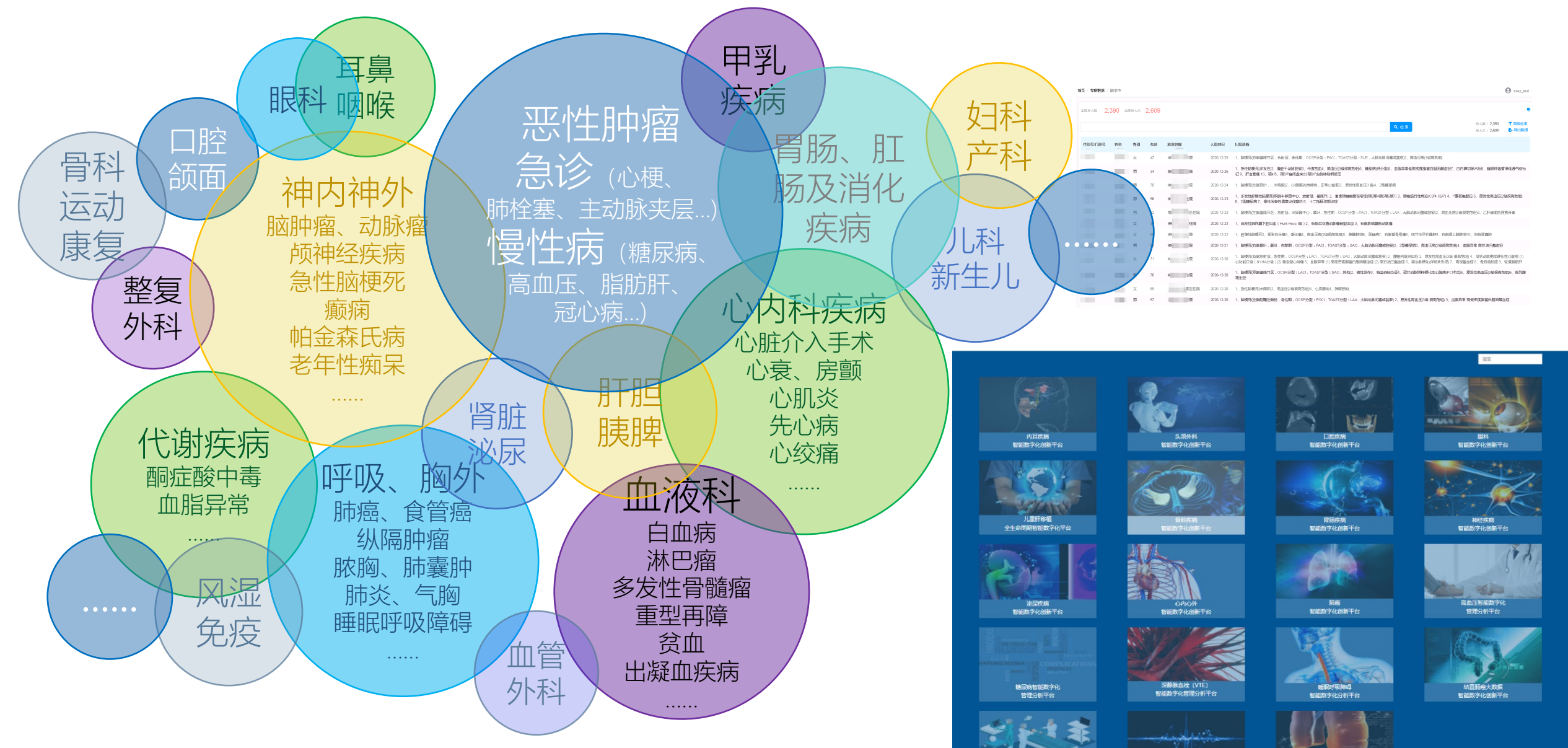
Automatic machine acquisition ensures high accuracy ;
Combined with convenient verification methods and intelligent detection ;
99% accuracy in raw data analysis.

Comprehensively

Data points (dimensions) set ~4000 per medical record
Average valid data points (dimensions) per medical record ~500



Support for **clinical specialties** and development of **database on specialized diseases**



Lessons learned from project construction

Data Reach!

Easily searchable, the time cost of obtaining data is dramatically reduced.

In the face of massive hospital data, after researchers have good ideas and thoughts, they can screen the required target cases in a very short period of time through the big data platform using combined query or funnel query.

Isolated? Disconnected?

Most of the hospital programs, where data is derived from CDRs, have poor data quality availability in the CDRs themselves. Lack of inclusion of multimodal data makes integration of full text this center and big imaging platforms difficult.

Assets? Junk?

Historical data quality is poor and should be targeted to develop specialty databases.

Combined with the high quality of the follow-up system, the whole process of integrating data. Combined with the structured management of data in the treatment, the system will no longer produce garbage since the launch.

Data-> Modeling-> Application-> Closed loop-> Iteration ->

Models are integrated into day-to-day operations, with scientific research focused on data insights and algorithmic evaluation.

Modeling Path

Big Data Platform Data Lake Construction

- Data collection, data into the lake
- Data Governance, Intelligent Search
- Multi-dimensional combination filtering
- Fast extraction of target data
- Export application and management

Disease-specific data platforms

Sedimentation and expansion of specialty disease databases to provide adequate data support for hospital research.

Model Platform

Mining, analysis, modeling based on various types of data

■ Provide **modeling algorithm development services based on text data** to support hospitals to publish high-level scientific research papers.

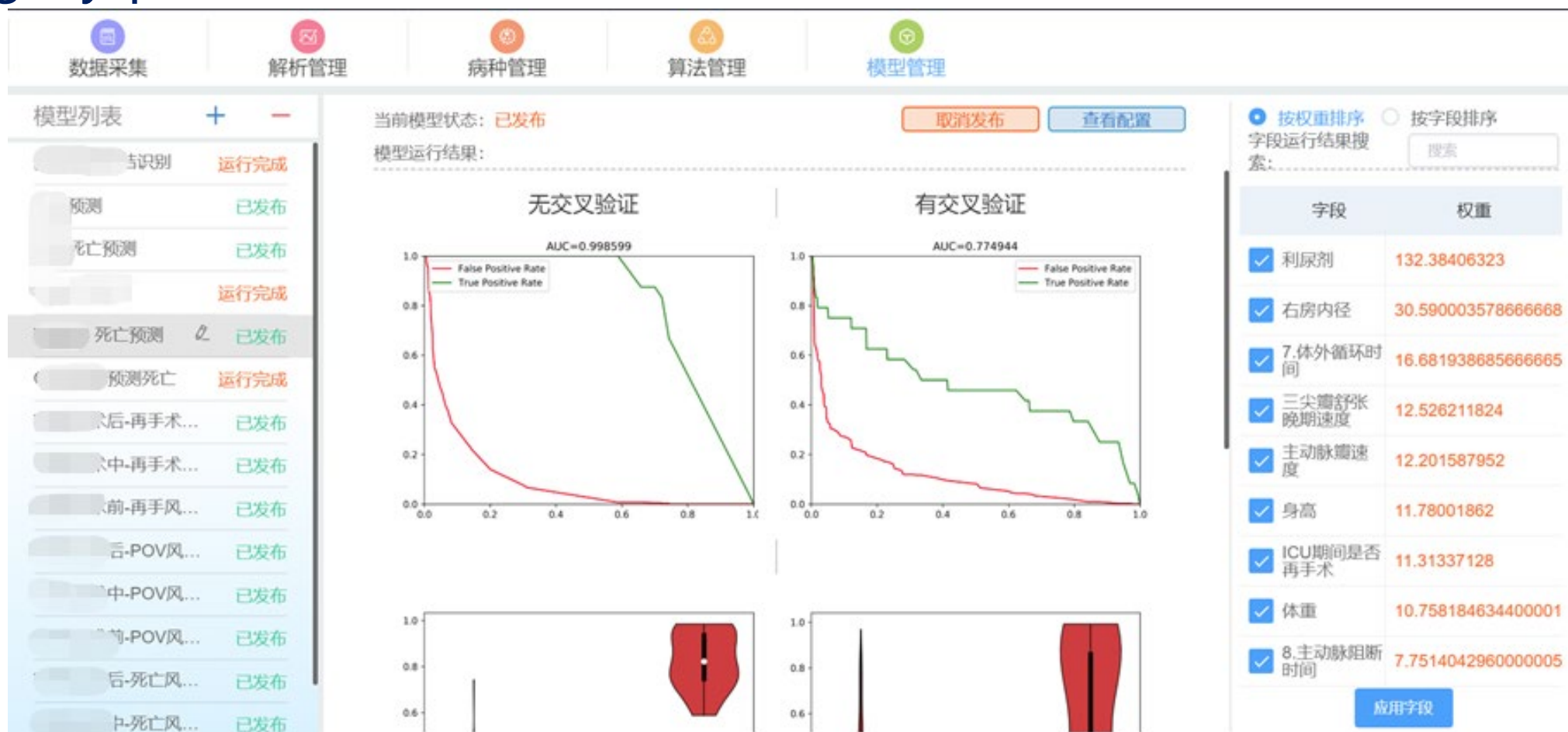
e.g.: Diseases Patient death, postoperative complications, secondary surgery prediction

The screenshot displays a software interface for medical modeling. A modal window titled '编辑模型任务' (Edit Model Task) is open, showing configuration options for a model named 'TAPVC术后-再手术风险评估' (TAPVC Postoperative - Reoperation Risk Assessment). The interface includes a sidebar with a list of models, a top navigation bar with buttons like '取消发布' (Cancel Release) and '设置风险区间' (Set Risk Interval), and a main area with three columns: '模型' (Model), '自变量[X]维度/指标' (Independent Variable [X] Dimension/Indicator), and '终点事件[Y=f(x)]' (Endpoint Event [Y=f(x)]). The '模型' column contains fields for model name, data source, and algorithm. The '自变量[X]维度/指标' column shows a list of variables with checkboxes for selection. The '终点事件[Y=f(x)]' column shows a list of events with checkboxes for selection. The '数据源' (Data Source) is set to '来自数据库' (From Database). The '数据源类型' (Data Source Type) is set to '文本' (Text). The '选择算法' (Select Algorithm) is set to 'Gradient Boost Tree(梯度提升决策树)' (Gradient Boost Tree (Gradient Boosting Decision Tree)). The '当前数据量' (Current Data Volume) is 488. The '指定空值率' (Specify Missing Value Rate) is 0.0%. The '训练集' (Training Set) is 0.0%. The '验证集' (Validation Set) is 0.0%. The '权重' (Weight) column on the right shows a list of weights for various features.

模型	自变量[X]维度/指标	终点事件[Y=f(x)]	权重
烧伤面积30~60			46.6626327702
补液与其他指标关系			7.71949417316
补液与休克关系			3.86219927766
烧伤休克-svc			3.458947
烧伤休克 (lasso)			2.9916658839999997
烧伤休克 (svc)			2.4368023978
TEST			2.1276985283
p&tAVSD-术后转归			1.7300369260000001
p&tAVSD			1.12982092
cAVSD_1			1.083769226
cAVSD_2			0.9893243800000001
TGA术后-死亡风险评估1			0.923690892
TGA术中-死亡风险评估			0.6587734700000001
TGA术前-死亡风险评估			0.30335247600000004
TAPVC术后-再手术风...			
TAPVC术中-再手术风...			
TAPVC术前-再手术风...			
TAPVC术后-POV风险...			

Provide **modeling algorithm development services based on text data** to support hospitals to publish high-level scientific research papers.

e.g.: Diseases Patient death, postoperative complications, secondary surgery prediction

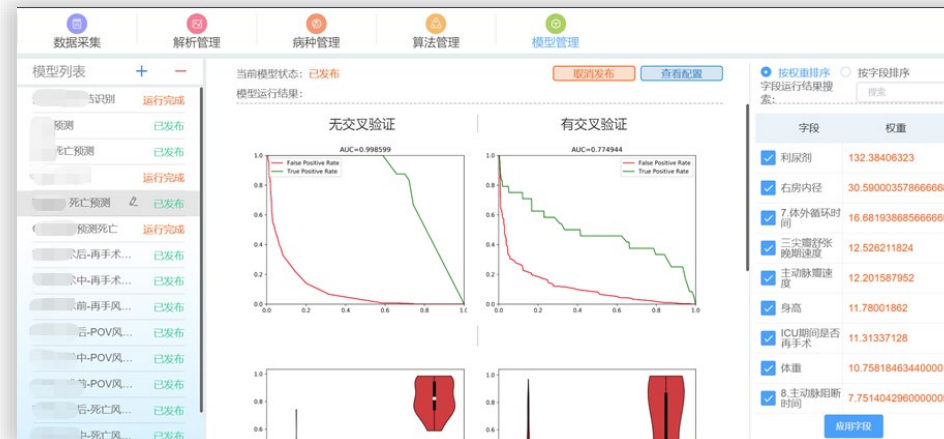


GUANGDONG PROVINCIAL PEOPLE'S HOSPITAL



- Cardiac specialty library incorporates **more than 20,000** visits
- Imaging center currently captures **millions of** exam records
- Storage of more than **250T** of image data (lossless compression)
- Image data is archived in real time, and the **average daily** volume of newly generated data is over **100G** (lossless compression).
- Providing examination reports and image distribution services to the **entire hospital**.
- Provide examination data, image control and original image service to **the third party**.
- Providing examination data and image services for the hospital's **research** programs.

Through the heterogeneous information integration platform to build Guangdong Provincial People's Hospital Hospital **big image platform, cardiovascular disease specialties data platform**, instant, comprehensive and accurate for the diagnosis and treatment of various clinicians and scientific research services. The establishment of **several key specialized disease databases and their model training platforms** has supported the hospital to carry out a large number of prospective and retrospective experiments, and its prediction model for death, postoperative complications, and secondary surgery of patients with complete pulmonary venous ectasia has been incorporated into **the daily diagnosis and treatment process of doctors**, so as to carry out **the whole process of prediction and management** of patients with prevalent heart disease.



■ Provide **modeling algorithm development services based on imaging data** to support hospitals to publish high-level scientific research papers

**Artificial Intelligence Assisted
Endoscopy in Real Time**

**Early diagnosis of upper
gastrointestinal tumors**

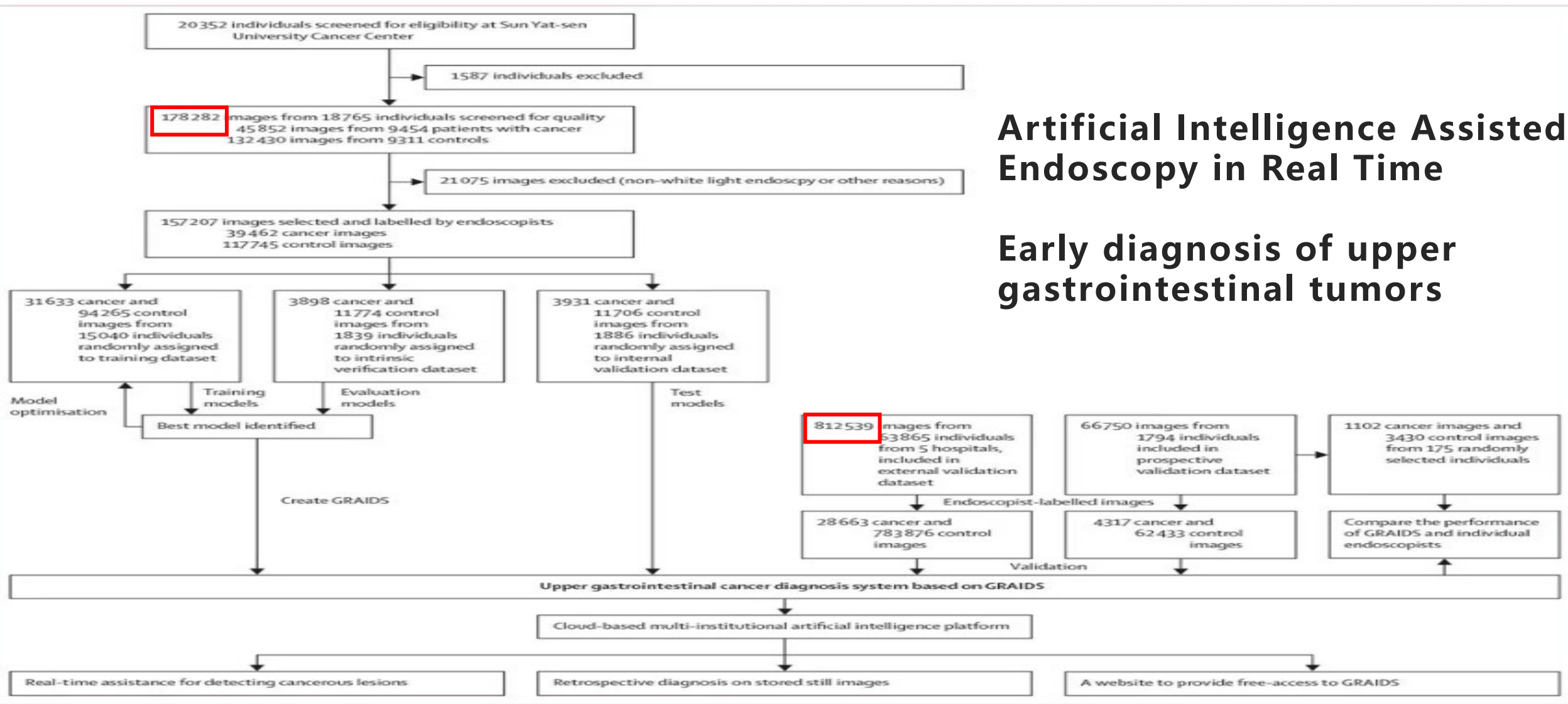
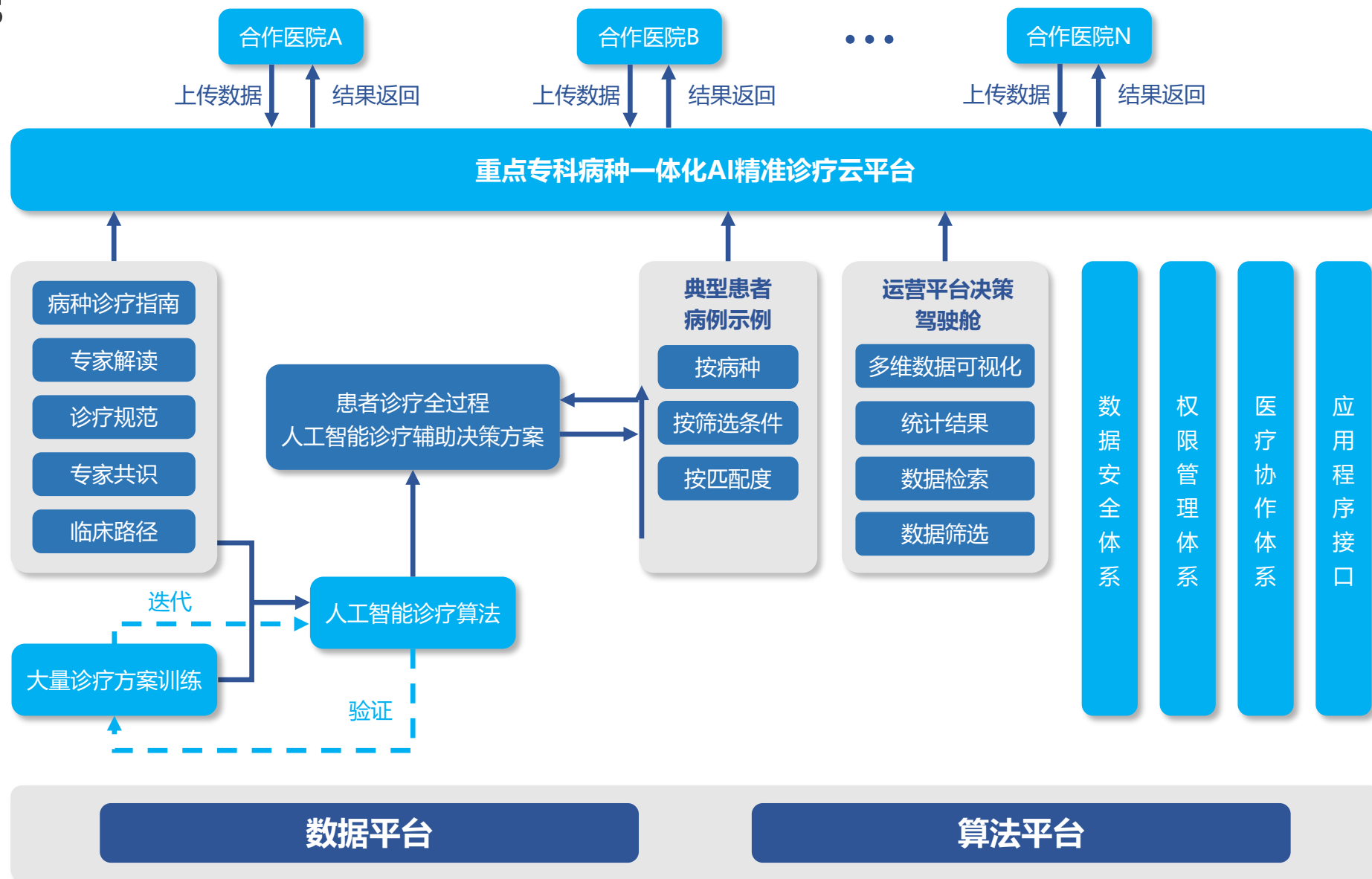
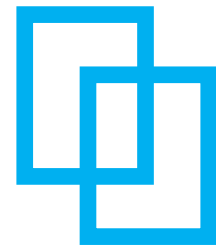


图1. 上消化道癌内镜AI智能诊断系统（GRAIDS）的开发与评价流程图

Translational output of scientific research: establishment of an integrated AI precision diagnosis and treatment cloud platform for key specialized diseases





03

PART 03

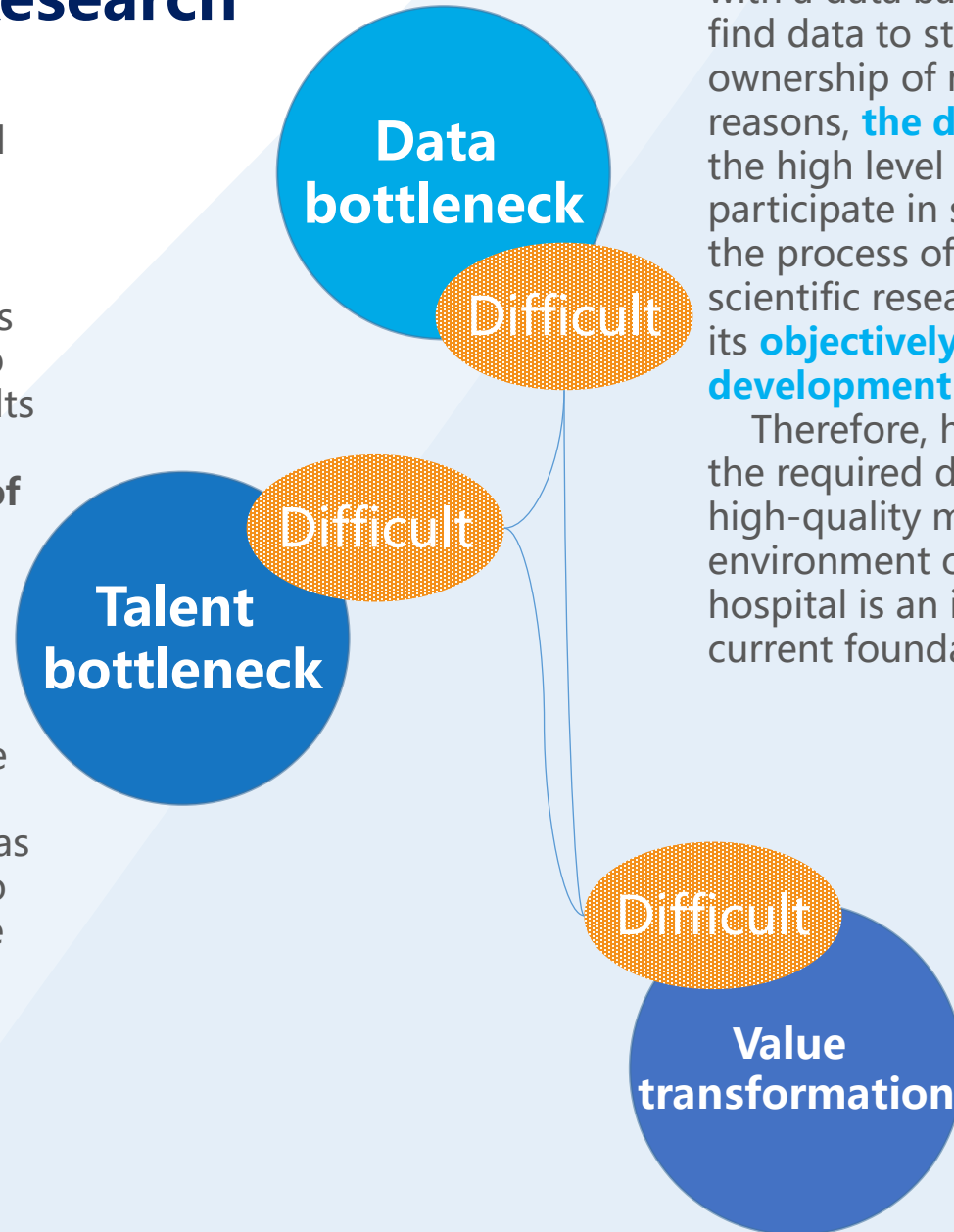
Summary of construction ideas

Conduct of Scientific Research

Due to the cross-cutting nature of medical research, it requires the cooperation of medical, math, statistics, computer science, and the most current focus on big data and artificial intelligence machine learning talents with different knowledge systems in order to facilitate the production of high-quality results based on high-quality data.

Hospitals can deploy a certain amount of high-level interdisciplinary talent in very focused disciplines, but it is impossible to deploy a full complement of interdisciplinary talent for all disciplines. Even if high level interdisciplinary talent is staffed for key disciplines, much of their time and work is spent helping the specialty validate some of the most basic or initial ideas (with a significant degree of trial and error to failure ratio), and thus the utility of staff time utilization is low.

It is only by being freed from low-level data organization and validation of ideas that the institution's valuable high-level interdisciplinary talent can contribute greater value.



The hospital has built a good business system, with a data base, easy for doctors to retrieve and find data to start scientific research, but due to the ownership of medical data and security reasons, **the data does not go out of the hospital**, the high level of talent outside the hospital can not participate in scientific research, the complexity of the process of data use review and approval of the scientific research work is seriously prolonged, and its **objectively become a huge bottleneck in the development of high-quality scientific research.**

Therefore, how to retrieve and accurately obtain the required data, and then immediately carry out high-quality modeling work under the environment of intranet without leaving the hospital is an important enhancement under the current foundation.

It is **difficult to translate** the scientific output of high-level papers **into clinical use** in the form of models and systems, which are then integrated into the daily work of hospitals, and **there is a lack of continuous system iteration and use of the feedback loop, failing to integrate AI into the CDSS clinical decision-making assistance.**

Conduct of Scientific Research

Due to the cross-cutting nature of medical research, it requires the cooperation of medical, math, statistics, computer science, and the most current focus on big data and artificial intelligence machine learning talents with different knowledge systems in order to facilitate the production of high-quality data based on high-quality data.

Hospitals can deploy a certain amount of high-level interdisciplinary talent in focused disciplines, but it is impossible to deploy a full complement of interdisciplinary talent for all disciplines. Even if high level interdisciplinary talent is staffed for key disciplines, much of the time and work is spent helping the specialists validate some of the most basic or initial ideas (with a significant degree of trial and error to failure ratio), and thus the utility of staff time utilization is low.

It is only by being freed from low-level data organization and validation of ideas that the institution's valuable high-level interdisciplinary talent can contribute greater value.

Data
bottleneck

Difficult

The hospital has built a good business system, with a data base, easy for doctors to retrieve and find data to start scientific research, but due to the ownership of medical data and security reasons, **the data does not go out of the hospital**, the high level of talent outside the hospital can not participate in scientific research, the complexity of the process of data use review and approval of the scientific research work is seriously prolonged, and its **objectively become a huge bottleneck in the**

Continuous
innovation

Big Data Platform + Future Trends

1. Advanced architecture and strong expandability
2. Text + image à multimodal
3. Accurate data governance
4. Integration process (AI+Workstation/APP)

Difficult

Value
transformation

papers **into clinical use** in the form of models and systems, which are then integrated into the daily work of hospitals, and **there is a lack of continuous system iteration and use of the feedback loop, failing to integrate AI into the CDSS clinical decision-making assistance.**

Advancements and Advantages

1. Establishment of an industry-complete horizontal interoperability medical data platform (including text and images)

For large tertiary hospitals, even though almost all of the patients' medical data is generated and stored in the hospital, big data vendors in the industry tend to build only for text or only for medical imaging systems. A high-quality structured text and image-based big data platform for full clinical diagnosis and treatment is still a gap. With the multiplier effect of complete data aggregation, together with artificial intelligence algorithms and data mining tools, this big data platform can become a clinical research platform with unlimited value. This project will rely on our existing technical foundation of medical text and medical images to substantially help large medical institutions to build a complete medical data platform for the first time.

2. Establish an industry-leading intelligent medical system integrated into the entire medical process.

Various types of AI-assisted medical clinical decision-making system in the medical institutions in the process of landing operation are faced with two major problems, either the lack of hospital's own raw data to support the training and application, or the lack of reasonable display scenes and terminals, and thus a variety of different but similar intelligent AI algorithms are difficult to integrate into the specific landing medical institutions to optimize and improve the adaptability and robustness of the data, and difficult to AI algorithm output results in a timely and convenient way to the medical institutions, the results of AI algorithm outputs. The results of the algorithms are not only difficult to integrate into the data of specific medical institutions, but also difficult to give the results of the AI algorithms in a timely and convenient manner to doctors who need AI decision-making assistance.

This project will integrate the big data center, integrate and push the relevant data to the intelligent AI decision-making assistance algorithm center, and distribute the AI results to the fixed or mobile terminals of the clinical departments, medical technology departments, nursing departments and other healthcare personnel, and prepare the relevant algorithmic opinions in real time on the terminals of the healthcare users who need to see such information, so as to empower the healthcare personnel to make medical decisions accurately and efficiently in a timely manner.

Advancements and Advantages

3. Establishing a self-training platform for medical data intelligence algorithms in top hospitals to enable doctors to take the initiative in scientific research

Under the premise that core medical data does not leave the hospital, the platform is used to establish a high-quality and high-efficiency flow mechanism for data within the hospital, and to support the normalization of scientific research topics through the integration of high-level data governance tools and modeling and mining analysis tools. The intellectual property rights of the platform are owned by the hospital, so experts can get rid of process constraints and intellectual property rights concerns, carry out high-level scientific research practices on the platform, and efficiently transform their clinical knowledge experience and medical decision-making experience into intelligent algorithms. At the same time, every spark of knowledge and every flash of inspiration will not be extinguished or die out in the cumbersome approval process. On the training platform where data is constantly updated and aggregated, without the need to bring in a third party, researchers can retrieve, select, and prepare data at any time, carry out model training and validation, preview and validate the inspiration and decide on the direction of the research.

4. Establish a self-training platform for medical data intelligence algorithms in top hospitals, so that doctors can take the initiative in scientific research Industry-leading Intelligent Medical Complete Closed Loop

The ideal data intelligence ecosystem is to generate data, manage data, analyze data, train models, integrate into processes, apply to users, close the loop with feedback, iterate and upgrade, and so on and so forth. This has been widely realized in the Internet domain, especially in the TMT industry, and the fundamental reason is that all the links to complete the above process are often completed in the product system of a single enterprise. However, in the field of medical informatization, due to the complexity of medical business, diverse needs, complicated systems, a large number of vendors, and a huge construction time span, the function carried by each system is an aspect of its specific needs, and the utilization of data in a complete and continuous closed loop cannot be carried out.

Therefore, this project will seize the opportunity of construction, combined with the advantages of specific medical institutions specialties specific single disease, leading to the establishment of high-quality flow of medical data within the hospital:

- Doctors through the mobile terminal to confirm, new, local governance data -

- Data center training, computing, and timely pre-distribution of data computing results--

- Closed-loop data application ecology in medical institutions with doctors' access, evaluation and feedback.

System Architecture and Resource Requirements

网关服务器
CPU: 16核 / 内存: 32 / 硬盘: 1T / 操作系统: CentOS/Windows

第三方模型接入管理服务

CPU: 4核
内存: 16G
硬盘: 1T
操作系统: Windows Server 2016 R2

模型训练服务

CPU: 32核
GPU: Nvidia RTX 3080
内存: 64G
硬盘: 2T
操作系统: Ubuntu 16.04.5 LTS

CPU: 12核 / 内存: 64G / 硬盘: 1T
操作系统: Windows

结构化专病种数据库服务器

结构化文本数据
A病种专病数据库 B病种专病数据库
标签化影像数据
X病种影像标记 Y病种影像标记

专病数据解析服务

CPU: 8核
内存: 32G
硬盘: 1T
操作系统: Windows Server 2016 R2

根据病种
解析服务

原始文本信息数据库

检查报告 电子病历 医嘱资料 检验数据

CPU: 8核 / 内存: 64G / 硬盘: 1T
操作系统: Windows Server 2016 R2

调阅及标注服务

CPU: 8核
内存: 32G
硬盘: 1T
操作系统: Windows Server 2016 R2

根据病种
标记服务

影像文件

存储池

按需按照冷温热存储进行分层级规划

患者索引数据库

基本资料 诊疗事件

CPU: 12核 / 内存: 128G / 硬盘: 1T
操作系统: Windows

医院 数据源

文本及影像数据

原始数据文件 DICOM
原始文本文件 HELP....
病理 XXXX
描述文件 CSV/EXCEL
内镜 JPG

既有数据平台
定义接口对接

合作机构 数据接入

Q/R
LAN
FTP

CDS影像归档服务

CPU: 8核
内存: 32G
硬盘: 2T
操作系统: Windows Server 2016 R2

MPI病人主索引服务

CPU: 8核
内存: 64G
硬盘: 2T
操作系统: Windows Server 2016 R2

CDS临床数据服务

CPU: 8核
内存: 32G
硬盘: 1T
操作系统: Windows Server 2016 R2

所有机器需支持线性扩展

