# पेटेंट कार्यालय
# शासकीय जर्नल

# OFFICIAL JOURNAL
# OF
# THE PATENT OFFICE


| | | |
|---|---|---|
| निर्गमन सं. 29/2022 | शुक्रवार | दिनांक: 22/07/2022 |
| **ISSUE NO. 29/2022** | **FRIDAY** | **DATE: 22/07/2022** |


पेटेंट कार्यालय का एक प्रकाशन

PUBLICATION OF THE PATENT OFFICE

(12) PATENT APPLICATION PUBLICATION
(19) INDIA
(22) Date of filing of Application :01/07/2022
(21) Application No.202241038010 A
(43) Publication Date : 22/07/2022

(54) Title of the invention : METHOD AND SYSTEM FOR DEPLOYING DEEP LEARNING TECHNIQUES FOR EFFICIENT BIG DATA ANALYTICS IN HEALTHCARE ENVIRONMENT

| | |
|---|---|
| (51) International classification | :G06F0021620000, G16H0010600000, G06F0016245800, G06N0020000000, G06F0016220000 |
| (86) International Application No<br>Filing Date | :PCT//<br>:01/01/1900 |
| (87) International Publication No | : NA |
| (61) Patent of Addition to Application Number<br>Filing Date | :NA<br>:NA |
| (62) Divisional to Application Number<br>Filing Date | :NA<br>:NA |

(71)**Name of Applicant :**
**1)DR. D. SURESH BABU (Head and BoS Chairperson)**
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. Email: Phone+91-8179108983 Khammam ----------- -----------
**2)DR. K. NAGESWAR RAO (Lecturer)**
**3)M. SRINIVASA RAO (Lecturer)**
**4)J.V.S. ARUNDATHI (Lecturer)**
**5)V. PULLARAO (Lecturer)**
**6)K. USHAPAVANI (Lecturer)**
**7)BABULU (Lecturer)**
**8)D. PAVAN SRIVASTA (Student)**
**Name of Applicant : NA**
**Address of Applicant : NA**
(72)**Name of Inventor :**
**1)DR. D. SURESH BABU (Head and BoS Chairperson)**
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. Email: Phone+91-8179108983 Khammam ----------- -----------
**2)DR. K. NAGESWAR RAO (Lecturer)**
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. Email: Phone+918686206907 Khammam ----------- -----------
**3)M. SRINIVASA RAO (Lecturer)**
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. Email: Phone+916303546939 Khammam ----------- -----------
**4)J.V.S. ARUNDATHI (Lecturer)**
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. Email: Phone+918639168744 Khammam ----------- -----------
**5)V. PULLARAO (Lecturer)**
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. Email : Phone+919398756062 Khammam ----------- -----------
**6)K. USHAPAVANI (Lecturer)**
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. Email: Phone+919154317942 Khammam ----------- -----------
**7)BABULU (Lecturer)**
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. Email: Phone: +919985238180 Phone: +918341022336 Khammam ----------- -----------
**8)D. PAVAN SRIVASTA (Student)**
Address of Applicant :B.TECH. (EEE) III YEAR VASAVI COLLEGE OF ENGINEERING(AUTONOMOUS), HYDERABAD TELANGANA STATE, INDIA. Email: Khammam ----------- -----------

(57) Abstract :
METHOD AND SYSTEM FOR DEPLOYING DEEP LEARNING TECHNIQUES FOR EFFICIENT BIG DATA ANALYTICS IN HEALTHCARE ENVIRONMENT ABSTRACT The present invention provides an approach to deploying deep learning techniques for efficient big data analytics in healthcare environment. The method comprises one or more data stores to store and manage data within a network and one or more servers to facilitate operations using information from the one or more data stores; and a processing system using machine learning and big data analytics. The invention includes a data access interface to receive data associated with a subject, wherein the data is received from a data source comprising an internal data source and an external data source, wherein the internal data source comprises a financial database of a financial institution associated with the subject, and the external data source comprises a public database and a web feed associated with the subject.

No. of Pages : 20 No. of Claims : 5

**FORM 2**

**THE PATENT ACT, 1970**

**(39 OF 1970)**

&

**THE PATENT RULES, 2003**

**COMPLETE SPECIFICATION**

[SEE SECTION 10 AND RULE 13]

| **TITLE**: METHOD AND SYSTEM FOR DEPLOYING DEEP LEARNING TECHNIQUES FOR EFFICIENT BIG DATA ANALYTICS IN HEALTHCARE ENVIRONMENT | |
|---|---|
| **APPLICANTS**: | |
| **DR. D. SURESH BABU** | DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS<br>SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A),<br>KHAMMAM, KHAMMAM DISTRICT,<br>TELANGANA STATE, INDIA-507001. |
| **DR. K. NAGESWAR RAO** | DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS<br>SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A),<br>KHAMMAM, KHAMMAM DISTRICT,<br>TELANGANA STATE, INDIA-507001. |
| **M.SRINIVASA RAO** | DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS<br>SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A),<br>KHAMMAM, KHAMMAM DISTRICT,<br>TELANGANA STATE, INDIA-507001. |
| **J.V.S. ARUNDATHI** | DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS<br>SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A),<br>KHAMMAM, KHAMMAM DISTRICT,<br>TELANGANA STATE, INDIA-507001. |

| | |
|---|---|
| **V. PULLARAO** | DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS<br>SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A),<br>KHAMMAM, KHAMMAM DISTRICT,<br>TELANGANA STATE, INDIA-507001. |
| **K. USHAPAVANI** | DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS<br>SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A),<br>KHAMMAM, KHAMMAM DISTRICT,<br>TELANGANA STATE, INDIA-507001. |
| **BABULU** | DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS<br>SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A),<br>KHAMMAM, KHAMMAM DISTRICT,<br>TELANGANA STATE, INDIA-507001. |
| **D. PAVAN SRIVASTA** | B.TECH. (EEE) III YEAR<br>VASAVI COLLEGE OF ENGINEERING(AUTONOMOUS),<br>HYDERABAD<br>TELANGANA STATE, INDIA. |



**The following specification particularly describes the invention and the manner in which it is to be performed**

# METHOD AND SYSTEM FOR DEPLOYING DEEP LEARNING TECHNIQUES FOR EFFICIENT BIG DATA ANALYTICS IN HEALTHCARE ENVIRONMENT

## FIELD OF THE INVENTION

**[0001]** The present invention generally relates to healthcare environment. More particularly, the invention relates to implementing deep learning techniques for efficient big data analytics in healthcare environment.

## BACKGROUND OF THE INVENTION

**[0002]** Machine learning evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Operations are achieved by building a machine-implemented model from example inputs in order to make data-driven predictions or decisions rather than following strictly static program instructions.

**[0003]** One type of machine learning involves supervised learning based on a training set as part of a classification process. Examples of machine learning algorithms used for classification include the well-known Naïve Bayes and C4.5 algorithms, or a so-called “stacked” combination of two or more such algorithms. The machine learning algorithm examines the input training set, and the computer “learns” or generates a classifier, which is able to classify a new document or another data object under one or more categories. In other words, the machine learns to predict whether a document or another type of data object, usually provided in the form of a vector of predetermined attributes describing the document or data object, belongs to a category. When a classifier is being trained, classifier parameters for classifying objects are determined by examining data objects in the training set that have been assigned labels indicating to which category each object belongs in the training set. After the classifier is trained, the classifier's goal is to predict to which category an object provided to the classifier for classification belongs.

**[0004]** A technical problem associated with machine learning, and in particular with classifiers, is that, in practice, the classifiers that assign objects to categories make mistakes. For example, classifiers may generate false positives, i.e., instances of mistakenly assigning an object to a category, and false negatives, i.e., instances of mistakenly failing to assign an object to a category when the object belongs in the category. These mistakes are often caused by a number of factors, including deficiencies of the training set. For example, typically, the larger the training set, the better the classification accuracy. However, large training sets may be unavailable. And if they are available, these data sets may be clustered and anonymized data, which may lead to other challenges in usage and processing. These and other technical problems may result in machine learning inaccuracies and decreased predictive efficiency and reliability.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0005]** FIG. 1 illustrates a network diagram illustrating a health data system where big data analytics is applied.

**[0006]** FIG. 2 illustrates an alternative embodiment of the health data system of FIG. 1.

**[0007]** FIG. 3 illustrates a flow chart illustrating an embodiment of a method of obtaining and processing healthcare data by implementing big data analytics.

## DETAILED DESCRIPTION OF THE INVENTION

**[0008]** Emerging technology solutions are poised to transform health care delivery. The health data system 100 as illustrated in FIGS. 1 and 2 can be used along with other systems to implement a suite of technology-enabled data-driven solutions designed to augment and accelerate effective disease management and care. For example, the health data system 100 can be part of a scalable technology core that can be integrated into local healthcare infrastructure to create a care management framework for delivering patient-centric and value-based care in a community, setting the stage for scaling to a broader set of communities.

**[0009]** In various embodiments it can be beneficial to configure a system that is consumer-centric with mobility and access, self-service enabled and designed with data fluidity and continuity. It can also be beneficial to have a personalized system that is configured to treat a person, not a diagnosis (e.g., by using a contextualized health profile and by leveraging longitudinal physiological plus behavioral, social and environmental data). It can be further beneficial to have an outcome-based system wherein value is driven by health outcomes; where quality is defined by safe and evidence-based care; where efficiency is achieved through optimized allocation of capacity, capability, availability and cost; and where effectiveness is personalized based on personal preference and ability, impacted by social and environmental factors. A desirable system can also include proactive health management that extends beyond reactive episodic care, includes population segmentation and stratification, includes a chronic disease care coordination plan, includes a long term health management plan, and includes a consumer education plan.

**[0010]** A health data system 100 that provides for obtaining, storing, curating, analyzing and/or providing access to health data can prove desirable and provide a basis for a wide range of applications as described in detail herein. This result can be achieved, according to one embodiment disclosed herein, by a health data system 100 as illustrated in FIG. 1.

**[0011]** Turning to FIG. 1, the health data system 100 is shown as comprising a plurality of data-source devices 110, a health data server 120, and a plurality of user devices 130. In some embodiments, the user devices 130 represent user services available from the health data system 100. The data-source devices 110 and user devices 130 are connected to the health data server 120 by a secure network 140 that can consist of any combination of wireless and wireline links. The data-source devices 110 are shown as comprising a smart phone data-source 110A, a laptop data-source 110B and a server data-source 110C, but in further embodiments, any suitable device can comprise a data-source 110, including a desktop computer, a tablet computer, a gaming device, a smart-television, a headset computer, a smartwatch, a body monitor device, or the like. Additionally, various embodiments can include any suitable number of any such data-source devices 110.

**[0012]** Similarly, although the user devices 130 are shown as being invoked from a smart phone user device 130A and a laptop user device 130B, in further embodiments, a user device 130 can comprise any suitable device including a server, a desktop computer, a tablet computer, a gaming device, a smart-television, a headset computer, a smartwatch, a body monitor device, or the like. Additionally, various embodiments can include any suitable plurality of any such user devices 130.

**[0013]** The server 120 can include one or more server systems, which can include any suitable plurality of devices and/or a cloud-based system. Additionally, the server 120 can comprise a plurality of modules, databases, or the like. For example, FIG. 2 illustrates one embodiment of the health data system 100 that comprises one or more agent modules 205, a security module 309, a market rules module and/or database 310, a switch module 311, a common info model module and/or database 312, a big data store 313 and an API 314, which are part of a cloud-based server system 120. The cloud-based server system 120 shown in FIG. 2 is implemented as a private cloud for illustration purposes only.

**[0014]** In accordance with various embodiments, the server 120 is configured to receive, process and store data obtained from data source devices 110 (see, e.g., FIG. 3). The server 120 can also be configured to process and/or retrieve stored data and provide it to the one or more user devices 130 in response to various queries or data requests that the user devices 130 may provide (see, e.g., FIG. 4).

**[0015]** Referring to FIGS. 1 and/or 2, one or more data source device 110 can be associated with one or more data source 201. For example, data sources 201 can include a variety of potential stakeholders and their associated data sources that might approve data for sharing via the health data system 100 in a targeted health ecosystem. For example, data sources 201 can include healthcare providers (e.g., data can be electronic medical records, lab data), health insurers/payers (e.g., data can be claims records), pharmaceutical and medical device companies (e.g., data can be clinical trial records, adverse event data), research (e.g., data can be a genomic profile), government/community health programs (e.g., data can be population health statistics) partnership databases), and/or individual patients (e.g., data can be biometrics, activity/behavior). Each data source 201 can be independently owned with its own set of unique data access control rules. Accordingly, the health data system 100 advantageously provides access to disparately held data (e.g., across the data sources 201) for use by independently developed health data services (e.g., via the user devices 130).

**[0016]** Each stakeholder data source 201 can choose the specific fields and elements, or subsets of data, which they approve to share, and the system 100 can manage the approvals of identified data (e.g., through consent/data use agreement) and/or identifiable data (e.g., through consent/Business Associate Agreements) as described in more detail herein and as illustrated in FIG. 3. Although examples of data associated with a given data source are described above, data sources can provide or be the source of any suitable type of data without limitation.

**[0017]** Agents A-C 306, 307, and 308 of FIG. 2 represent an example architecture of the distributed agent modules 205 that are configured to obtain, receive and/or access data on a manual and/or automated basis from the data sources 201. The agent modules 205 can be associated with the one or more data sources 201 and/or the data source devices 110, and a given data source 201 or data source device 110 can be associated with one or more agent modules 205. These agent modules 205 ensure that any needed metadata/supporting elements (e.g., consents, access rights, source information) are transmitted alongside data that is obtained, received and/or accessed from the data sources 201.

**[0018]** In some embodiments, manual interactions can be conducted via a web-enabled portal within the data source 201 (e.g., via one or more data source device 110), in which an owner of the data source 201 is responsible for deciding what data is transmitted to the server 120 from the data source 201 by agent modules 205. Automated interactions can be via script programs that are configured (e.g., with business rules, or the like) and then scheduled to run on a particular frequency, and/or in real time based on desired monitoring criteria to specify what data is transmitted to the server 120 from the data source 201 by the agent modules 205. The data in the data source 201 that can be transmitted to the server 120 by the agent modules 205 is designated by an identifier. An identifier is a string of alpha-numeric characters that uniquely identify a patient record in the data source 201.

**[0019]** Business rules can be configured in the agent modules 205 and can include naming conventions, data lineage tracking, permissible and non-permissible fields based on the ability to share identifiable data, sharing restrictions, and other stakeholder organization-specific rules for the agent to follow when managing data. In some embodiments, as illustrated in FIG. 2, agent modules 205 can encrypt data that will be transmitted to the server 120 prior to leaving environment and/or firewall of data source 201.

**[0020]** In various embodiments, data that is not in accordance with the sharing business rules defined by the data source 201 will remain in the environment of the data source 201 and will not be brought into the health data system 100 by the agent modules 205. In some embodiments, this can include identifiable data that does not contain a flag indicating that a patient's consent for sharing was obtained and a notice of privacy practices was given, such that the health data system 100 must infer that the data source 201 does not have permission from an individual to share the individual's data outside of the environment of the data source 201 (e.g., blocks 403, 406 and 409 of FIG. 3).

**[0021]** 3. To validate a request for data by user device 130 the market rules module 310 checks if the specific record requested by the user device 130 from data source 201 is allowed by the market rules specified by data source 201. A market rule can be specified to inform the switch 311 to not persist data from 201 collected by 205 in the common information model 312. This market rule can have the following specification: (<source id><record id>on-demand): this rule states that a record identified by <record id> from the data source 201 identified by <source id> should not be stored in the common information model 312. For records from the data source 201 identified by this rule, the switch module 311 stores a query in the common information model 312 for this record and executes this query at the time a request is made for this record by the API 314.

**[0022]** In various embodiments, business and/or market rules can facilitate the application of algorithms and logic such as data consistency checking and cleansing, reference data standardization, and master patient indexing, in order to facilitate reuse and avoid duplication or mis-representation of data. In various embodiments, this can include operations such as identification and adjusting of null values and inconsistencies in units of measure, and the usage of demographic values to align multiple stakeholders' records for the same individual under a single global identifier within the system 100. Accordingly, in some embodiments, business and/or market rules can enable the use of data across a plurality of services to derive novel insights while also protecting information privacy.

**[0023]** The health data server 120 can also include one or more switch modules 311. In one embodiment the implementation of a switch module 311 includes:

1. A selected switch module 311 is set up with an input message queue. The input message queue receives messages from the agent module 205. The format of each message can include a list of (<attribute>, <value>) pairs.

2. The selected switch module 311 is set up with a data quality queue to log messages identifying records obtained from the data source 201 that have been found to have data quality issues by the data quality engine (DQE). DQE includes:

a. Data quality rules for each data source 201 connecting into system 100

b. Data quality rule can include the following format: <attribute><type><possible values>

c. Application of data quality rules corresponding to the data source 201 to messages placed into the input message queue of the switch module 311 by the agent module 205 to flag a data quality issue if any attribute of the message in the input message queue does not have a value contained in the set of possible values for that attribute in the data quality rules for that data source 201.

3. The health data server 120 sends to the data source 201 data quality messages from the data quality queue for resolution by the data source 201

4. If no data quality issue is detected with the message in the input message queue then the switch module 311 loads up the data in the message into the corresponding table in the common information model 312. For example, if the message is about the medication administered to a given patient at a hospital then the contents of this message are loaded into the medication table in common information model 312 tagged with the identifier for that patient along-with the identifier of the data source 201 associated with the hospital from where the agent module 205 collected the data for that patient. At the time of loading the data into the common information model 312, one or more coding dictionaries can be used to map terms in the message into standard terms. For example, blood pressure can be mapped to hypertension.

5. The switch module 311 periodically executes the match engine. Match engine detects when data from two different data sources 201 belongs to the same patient.

6. The switch module 311 is set up with an output message queue. Output message queue receives messages from the API module 314.

7. The switch module 311 invokes the market rules module 310 for each message in its output message queue. If the market rules module 310 validates the message for access to requested data from 312 then the switch module 311 maps the message in the output message queue into a query for common information model 312. The query format is (<source id>, <record id>) where <source id> identifies a data source 201 and <record id> identifies a record from source 201. This query is then executed against common information model 312 by switch component 311. The data obtained from common information model 312 by switch 311 is returned to the user device 130 by the API 314.

8. The switch module 311 applies a natural language processing (NLP) engine to unstructured attributes in data from the data sources 201 collected by the agent modules 205 (unstructured attributes are identified in 312 as attributes that allow arbitrary length character strings as values). The resulting data includes (<attribute>, <value>) lists and these lists are used to add data to the corresponding tables in the common information model 312. In this way the switch component 311 reduces unstructured data from 201 into structured data in the common information model 312. By doing so, the switch component 311 integrates structured data from the data sources 201 with unstructured data from the data sources 201.

9. The switch module 311 tags each data stored in the common information model 312 with any data used in the generation of the data being stored in the common information model 312. This way complete data traceability is maintained in the common information model 312. For example, if there two sets of medication records on a patient in 312 and a user service 130 generates a reconciled medication list then prior to storing the reconciled medication list in the common information model 312, the switch module 311 tags the reconciled medication list with the identifiers for the two medication list that have been reconciled.

10. The switch module 311 implements a data exclusion service to de-identify data from the data sources 201. This service is provided with a list attributes that need to be de-identified. Upon execution of the data exclusion service on specified data the values of the identified attributes are masked to make it impossible to obtain the original values of these attributes by replacing each byte of storage allowed for holding a value for an attribute with the null byte ‘0’. Simultaneously,

the data is tagged with the names of attributes that are masked. In this way the input data is rendered de-identified by the data exclusion service. In an alternative embodiment of data exclusion service the value of a de-identified attribute can be replaced with a unique tag and the correspondence between the original value of the attribute and the replacement tag can be added to a tokenization table in the common information model 312.

**[0024]** The switch component 311 can implement various services to manage the lifecycle of data from onboarding to termination, including:

**[0025]** Registration service to onboard & register the data source 201 and the data service 130. The registration of the data source 201 with system 100 requires the data source 201 to provide an interface to system 100. This interface consists of an IP address and a set of instructions on how to obtain data from that IP address. The data source 201 provides a set of data access rules to health data server 120 that govern the use of the data provided by the data source 201 by the user devices 130 participating in system 100. In health data server 120, these rules are called market rules configured in the market rules module 310. Registration by data service with the system 100 requires system 100 to provide to the data service 130 an IP address and a set of instructions on how to obtain data from that IP address. Before responding to a request for data by a registered data service 130 system 100 validates the request against market rules provided by data sources 130 who have contributed requested data into system 100. This way access to data in system 100 is always controlled by data access rules established by the data sources 201 contributing data into system 100.

**[0026]** Data audit service to log data entering and leaving the system 100.

**[0027]** Data quality engine (DQE) to monitor the quality of data entering the system 100 which in turn drives better analytics, which can include the following dimensions: Attribute-level data quality (for example age should be in a certain range or last name should not be blank), Context or aggregate data quality (for example patient data on average should be approximately 50% male 50% female plus or minus certain margin), Operational data quality (for example, reject if patient demographics data is present but medication data is not present thereby rendering the entire patient record not useful)

**[0028]** Data termination service to remove from the system 100 any data from a data source upon request by the data source owner makes it easy for the source owner to remove their data with control from the platform.

**[0029]** In some embodiments, where the rules cannot sufficiently resolve inconsistencies in the data, those elements/fields are flagged for manual intervention from a data steward user who is authorized to see identifiable data from all stakeholders and reconcile the differences. This can include examples in which patient matching/indexing algorithms have identified two records, from two stakeholders, which records appear to belong to the same patient but are missing.

**[0030]** Patient match service to identify and match data on the same patient ingested from multiple data sources 201. For example, using patient record elements that are less likely to change throughout a patient's life (e.g., name, date of birth, gender, SSN—as opposed to a phone number or address for example), the health data server 120 can match a patient across multiple records of different types from different sources. Accordingly, in some embodiments, the health data server 120 can be configured to assign a weight to each field to create a total match score, with the weight implying the importance of that field in the matching process (e.g., social security number (SSN) and date of birth (DOB) can get a 35% weight while gender, last name, and first name can get 10% weight each), which weighting may be operable to improve accuracy of the match. Upon matching data from two or more data sources 201, the switch module 311 tags the data with a unique id for that patient in the health data system 100.

**[0031]** The resulting cleansed, structured, standardized data can be stored in one or more databases defined by the common information model module 312, (see, e.g., block 413 and 415 of FIG. 3). For example, the common information model module 312 can standardize information in accordance with terminologically robust standards such as Systematized Nomenclature of Medicine—Clinical Terms (SNOMED) standards (e.g., for procedure codes, medication method), International Organization for Standardization (ISO) standards (e.g., for Country Codes), Fast Healthcare Interoperability Resources (FHIR) standards (for certain fields), or the like. For example, medication route and dosage method can be standardized to the SNOMED values, country code can be standardized to 3 digit ISO-3166 country codes, gender and race can be standardized using FHIR standard values, or the like.

**[0032]** In some embodiments, when standardizing and integrating data, the system 100 can also facilitate patient matching. For example, using patient record elements that are less likely to change throughout a patient's life (e.g., name, date of birth, gender, SSN—as opposed to a phone number or address for example), the system 100 can match a patient across multiple records of different types from different sources. Accordingly, in some embodiments, the system 100 can be configured to assign a weight to each field to create a total match score, with the weight implying the importance of that field in the matching process (e.g., social security number (SSN) and date of birth (DOB) can get a 35% weight while gender, last name, and first name can get 10% weight each), which weighting may be operable to improve accuracy of the match. Additionally, the business/market rules can also identify unstructured data and prepare it for storage in clusters separate from the rest of the structured data (e.g., in big data storage 313) without losing lineage information (see, e.g., block 412 of FIG. 3).

**[0033]** Although various embodiments discussed herein relate to processing data for storage on the health data server 120, in further embodiments, data processing described herein can apply to data in-motion. In other words, in some embodiments, data may not be stored on the server 120 and can be passed between data sources 201 and data consumers 207 using the health data server 120 as an intermediary, but without the data being stored in the health data server 120. In such embodiments, data can be processed as described herein.

**[0034]** Referring to FIGS. 1, 2, and 4, the system 100 can comprise a plurality of data consumers 207 that are configured to consume data from the server 120 for the purpose of research, clinical care, commercial purposes, or the like, using the one or more user devices 130. For example, a given data consumer 207 can comprise one or more user devices 130 that are configured to request and/or receive various types of data from the server 120 as described in more detail herein and as shown in FIG. 4, through the user devices 130.

**[0035]** The flowchart and block diagrams in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods and computer program products according to various embodiments. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of code, which comprises one or more executable instructions for implementing the specified logical function(s). It should also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

**[0036]** The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the embodiments. As used herein, the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

**[0037]** The corresponding structures, materials, acts, and equivalents of all means or step plus function elements in the claims below are intended to include any structure, material, or act for performing the function in combination with other claimed elements as specifically claimed. The description of the embodiments has been presented for purposes of illustration and description, but is not intended to be exhaustive or limited to the embodiments in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the embodiments. The examples disclosed were chosen and described in order to best explain the principles of the embodiments and the practical application, and to enable others of ordinary skill in the art to understand the various embodiments with various modifications as are suited to the particular use contemplated.

**[0038]** Though embodiments have been described with reference to certain versions thereof; however, other versions are possible. Therefore, the spirit and scope of the embodiments should not be limited to the description of the preferred versions contained herein.

**I/WE CLAIM**:

1. A method for implementing deep learning techniques for efficient big data analytics in healthcare environment, the method comprising:
   one or more data stores to store and manage data within a network;
   one or more servers to facilitate operations using information from the one or more data stores; and
   a processing system using machine learning and big data analytics, the processing system comprising:
   a data access interface to receive data associated with a subject, wherein the data is received from a data source comprising an internal data source and an external data source, wherein the internal data source comprises a financial database of a financial institution associated with the subject, and the external data source comprises a public database and a web feed associated with the subject;
   generate a recommendation based on the at least one of the financial forecast, the ratio, and the index, wherein the recommendation comprises at least the financial action for the subject to take based on the predicted life event; and
   an output interface to transmit the recommendation to at least one of a user at a financial institution, the subject, and the processing system for continuous machine learning, statistical analysis, simulation, or modeling.

2. The method for implementing deep learning techniques for efficient big data analytics in healthcare environment, as claimed in claim 1, wherein the one or more attributes define at least one of clinical health data, biometrics, wearables, social media data, mobile application data, and device instrumentation.

3. The method for implementing deep learning techniques for efficient big data analytics in healthcare environment, as claimed in claim 1, transmitting from the message handler module, the one or more tenant-specific business objects to the platform persistence module for storage.

4. The method for implementing deep learning techniques for efficient big data analytics in healthcare environment, as claimed in claim 1, receiving at the platform persistence module, external data from an external services orchestrator module.

5. The method for implementing deep learning techniques for efficient big data analytics in healthcare environment, as claimed in claim 1, receiving at the platform analytics module, data from the platform persistence module, wherein the data from the platform persistence module comprises at least one of: messages from the message collector module, tenant-specific business objects from the message handler module, and external data from the external services orchestrator module.

**METHOD AND SYSTEM FOR DEPLOYING DEEP LEARNING TECHNIQUES FOR EFFICIENT BIG DATA ANALYTICS IN HEALTHCARE ENVIRONMENT**


## ABSTRACT

The present invention provides an approach to deploying deep learning techniques for efficient big data analytics in healthcare environment. The method comprises one or more data stores to store and manage data within a network and one or more servers to facilitate operations using information from the one or more data stores; and a processing system using machine learning and big data analytics. The invention includes a data access interface to receive data associated with a subject, wherein the data is received from a data source comprising an internal data source and an external data source, wherein the internal data source comprises a financial database of a financial institution associated with the subject, and the external data source comprises a public database and a web feed associated with the subject.

COMPLETE SPECIFICATION

Sheet No. 1

Total No. of Sheets. 4

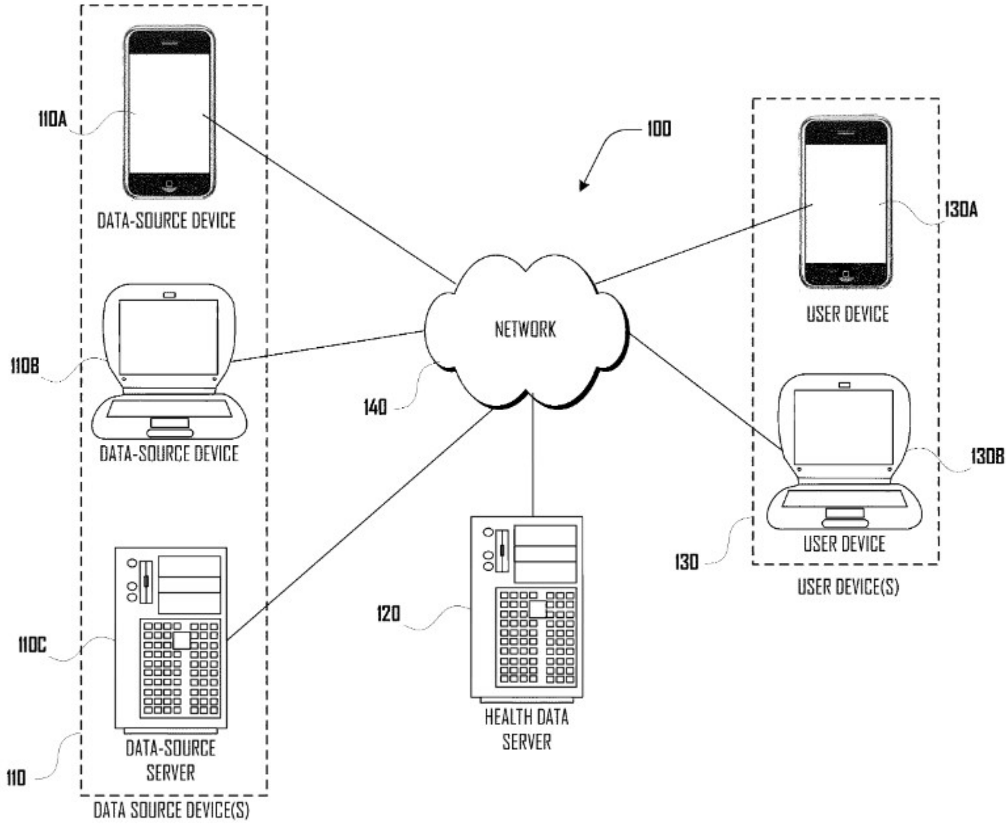


FIG. 1

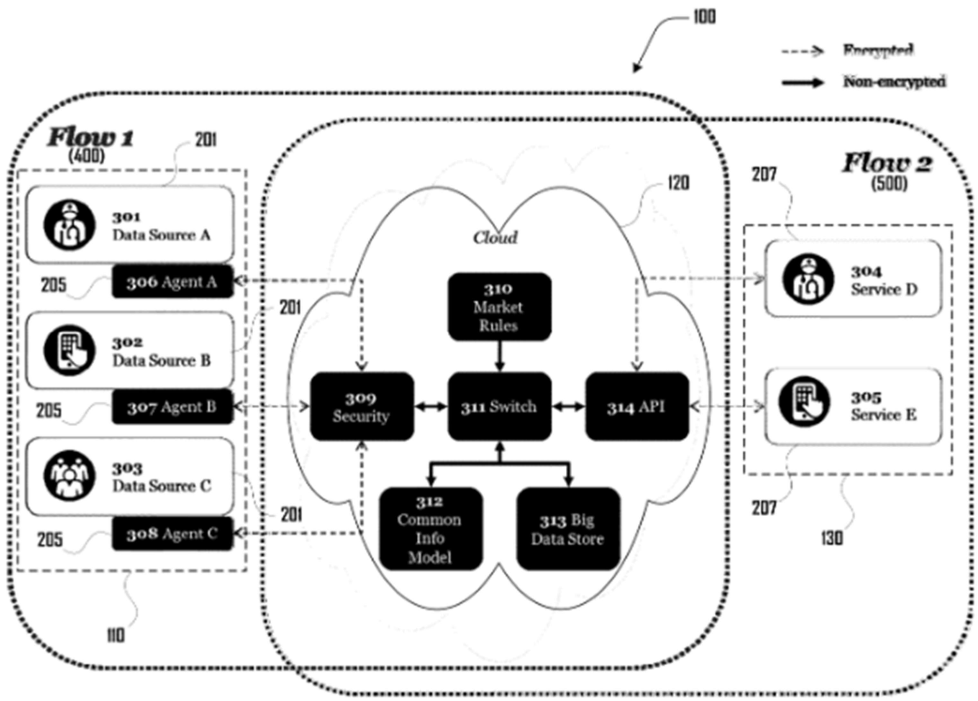
100
Encrypted
Non-encrypted
Flow 1
(400)
201
301
Data Source A
205
306 Agent A
201
302
Data Source B
205
307 Agent B
303
Data Source C
201
205
308 Agent C
110
120
Cloud
310
Market
Rules
309
Security
311 Switch
314 API
312
Common
Info
Model
313 Big
Data Store
207
Flow 2
(500)
304
Service D
305
Service E
207
130

FIG. 2

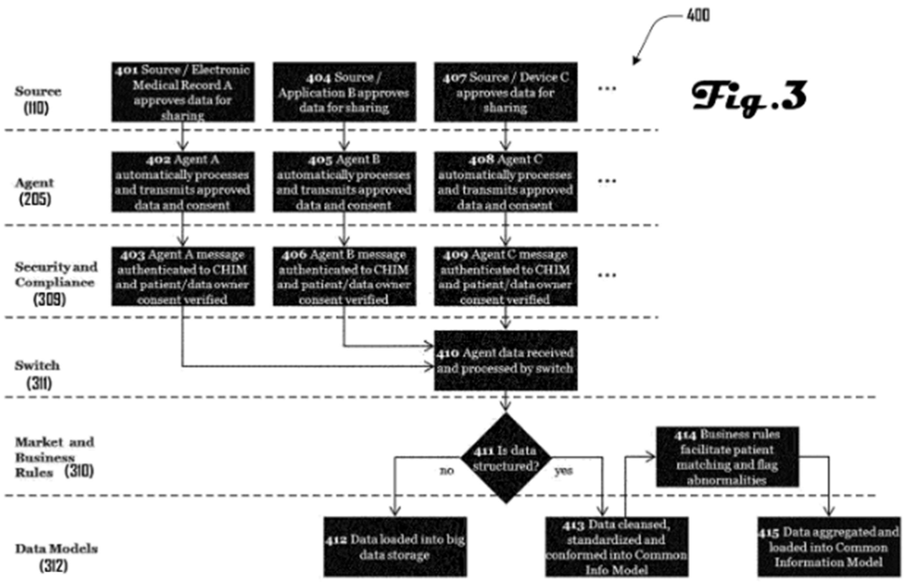
400
Fig.3
Source (110)
401 Source / Electronic Medical Record A approves data for sharing
404 Source / Application B approves data for sharing
407 Source / Device C approves data for sharing
Agent (205)
402 Agent A automatically processes and transmits approved data and consent
405 Agent B automatically processes and transmits approved data and consent
408 Agent C automatically processes and transmits approved data and consent
Security and Compliance (309)
403 Agent A message authenticated to CHIM and patient/data owner consent verified
406 Agent B message authenticated to CHIM and patient/data owner consent verified
409 Agent C message authenticated to CHIM and patient/data owner consent verified
Switch (311)
410 Agent data received and processed by switch
Market and Business Rules (310)
411 Is data structured?
no
yes
414 Business rules facilitate patient matching and flag abnormalities
Data Models (312)
412 Data loaded into big data storage
413 Data cleansed, standardized and conformed into Common Info Model
415 Data aggregated and loaded into Common Information Model

FIG. 3

Sheet No. 4

Total No. of Sheets. 4

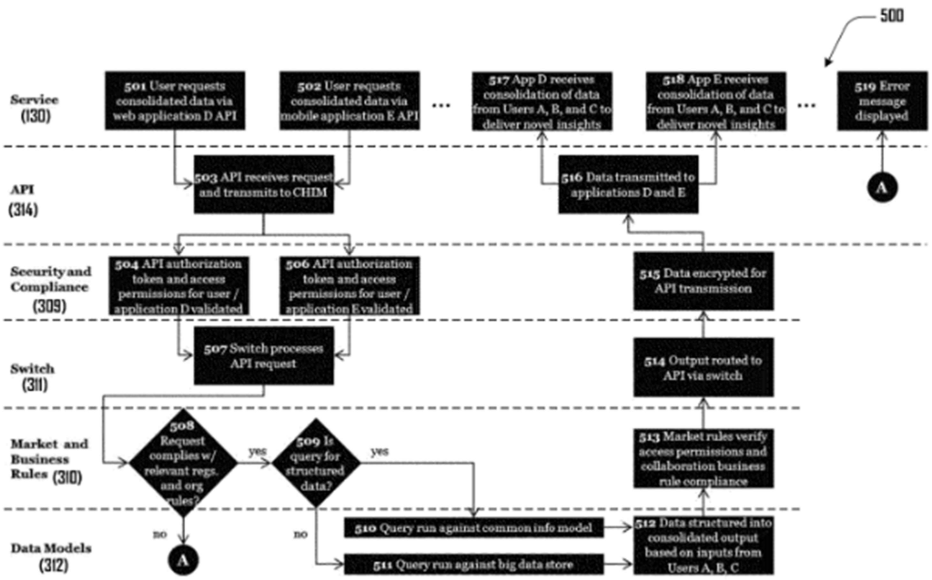


FIG. 4