A Systematic Approach to Healthcare Knowledge Management Systems

Subjects: Computer Science, Artificial Intelligence | Computer Science, Information Systems | Health Care Sciences & Services
Contributor: Thanh Ngoan Trieu, Anh-Cang PHAN, Thuong-Cang PHAN

Big data in healthcare contain a huge amount of tacit knowledge that brings great value to healthcare activities such as diagnosis, decision support, and treatment. However, effectively exploring and exploiting knowledge on such big data sources exposes many challenges for both managers and technologists. A healthcare knowledge management system that ensures the systematic knowledge development process on various data in hospitals was proposed. It leverages big data technologies to capture, organize, transfer, and manage large volumes of medical knowledge, which cannot be handled with traditional data-processing technologies. In addition, machine-learning algorithms are used to derive knowledge at a higher level in supporting diagnosis and treatment.

Keywords: KMS ; big data ; machine learning ; high blood pressure ; brain hemorrhage ; Spark

1. Introduction

Knowledge represents an important resource that needs effective management to capture, organize, transfer, and apply this kind of intellectual property. A knowledge management system (KMS) is a class of information systems for managing organizational knowledge. Unlike traditional information systems that only focus on capturing, organizing, and managing explicit knowledge, KMS explores and exploits explicit and tacit knowledge. The advancement of knowledge management systems has changed the way organizations operate, especially medical organizations, in which healthcare is a knowledge-intensive industry. Healthcare data come from many sources such as hospital databases, national databases, or private analytic databases. An example of private analytic databases is the Premier Hospital Database, which comprises data from more than 1 billion patient encounters from over 700 private and academic hospitals in the United States, corresponding to approximately 20% of all hospitalizations in the country [1]. Many studies [2][3][4] leverage the available databases to reveal valuable knowledge, which is meaningful in public healthcare. The large-volume databases including patient information, disease diagnosis, and medical treatment allow for the investigation of rare diseases and uncommon complications that are not always possible with prospective clinical studies. However, the rapid increase in healthcare records in these databases poses many challenges for KMS to improve the decision-making support process. Specially, with the advent of technology in the field of the Internet of Things, many wearable sensor devices are launched to remotely monitor patients' health. This will rapidly enlarge the size of the health records in healthcare systems. The large amount of data needs to be managed and analyzed appropriately. Big data in healthcare contain explicit and tacit knowledge that supports a wide range of medical functions such as disease monitoring, clinical decision support, and healthcare management. Thus, it is necessary to build an effective KMS managing the precious knowledge to support medical diagnosis decision-making in the context of big data and artificial intelligence.

Alavi and Leidner presented discussions about knowledge, knowledge management, and knowledge management systems [5]. They described issues, challenges, and benefits of knowledge management systems [6]. Brent Gallupe considered three levels of knowledge management technologies: tools, generators, and specific KMSs [7]. Some studies discussed knowledge management in the age of big data related to some aspects such as knowledge bases, knowledge discovery, and knowledge fusion. Suchanek and Weikum gave an overview of the methods for building large knowledge bases [8]. Begoli and Horey presented three system design principles that can be integrated into knowledge discovery infrastructure and provided development experiences with big data problems [9]. Dong et al. introduced a web-scale probabilistic knowledge base that employed supervised machine-learning methods in knowledge fusion from existing repositories [10]. These studies considered the presentation of big data in their systems, but they did not provide a comprehensive process of knowledge development. Tretiakov et al. [11] adapted and extended a generic model of knowledge management systems including relevant factors to healthcare. Experiments were conducted on data collected from 263 doctors within two district health boards in New Zealand. Maramba et al. [12] presented a comprehensive synopsis of the challenges in the implementation of computer-based KMS in healthcare institutions. Manogaran et al.

[13] proposed a big-data-based KMS supporting clinical decisions. They provided an overview of big data tools and technologies that can be used in KMS. These observed studies remain at the level of knowledge exploration that do not apply new knowledge in concrete practice. Recently, Le Dinh et al. proposed an architecture for implementing big-data-driven knowledge management systems [14]. A knowledge management system in a big data context must fully ensure the development process of knowledge including four stages: capture, organize, transfer, and apply. The study stays on the abstract level of KMS without any implementation.

In order to overcome the above challenges, researchers propose to build a big-data-driven healthcare knowledge management system supporting the diagnostic decision in a parallel and distributed environment. The large-scale healthcare system ensures a complete and comprehensive knowledge development process, including knowledge exploration and knowledge exploitation. Additionally, the involvement of artificial intelligent and big data processing is to provide real-time diagnosis decision supports with the massive volumes of medical records for a reasonable response time. The proposed healthcare knowledge management system for supporting medical diagnosis includes four layers: a data layer, an information layer, a knowledge layer, and an application layer. An illustration of the proposed system is presented using machine-learning techniques in the knowledge layer to generate knowledge for hypertension and brain hemorrhage diagnosis. Data used in this system are collected from several hospitals and health-monitoring devices. Hypertension is one of the most leading causes of disability and death worldwide. According to the World Health Organization (WHO), an estimated 9.4 million deaths are caused by high blood pressure. This dangerous disease needs to be promptly detected and treated to limit the risks of death as well as disease complications. Researchers use decision trees to generate knowledge for hypertension diagnosis and classification. Decision trees learn and generate simple rules from a complex decision-making process that is similar to the way of human thinking. In addition, researchers use deep-learning techniques to generate knowledge for brain hemorrhage detection and classification. A brain hemorrhage is a type of stroke that is caused by an artery bursting in the brain. Stroke is the second leading cause of death according to the World Health Organization. The diagnosis of the disease is based on cerebral CT/MRI images; thus, researchers proposed to use deep-learning techniques for hemorrhage detection and classification. The trained model with Faster R-CNN Inception ResNet v2 achieves the mean average precision of 79% in classifying four types of brain hemorrhage.

2. Knowledge Management Systems

Knowledge management systems have a dramatic impact on the decision-making support of organizations. However, an effective KMS needs to ensure the whole process of knowledge management, including knowledge exploration and knowledge exploitation. Le Dinh et al. proposed an architecture for big-data-driven knowledge management systems including a set of constructs, a model, and a method [14]. This architecture has complied with the requirements of the knowledge development process and the knowledge management process. Based on the research of Le Dinh et al., researchers have proposed an architecture for a knowledge management system supporting medical diagnosis including four layers: data layer, information layer, knowledge layer, and application layer (**Figure 1**). This knowl- edge management system ensures all four stages of the knowledge development process, including data, information, knowledge, and understanding, corresponding to four main activities, which are capture, organize, transfer, and apply. The objective of this entry is to present the architecture for medical diagnosis decision-supporting systems by collect- ing and analyzing big data. This proposal addresses two major challenges: knowledge management and knowledge organization from disparate data sources.

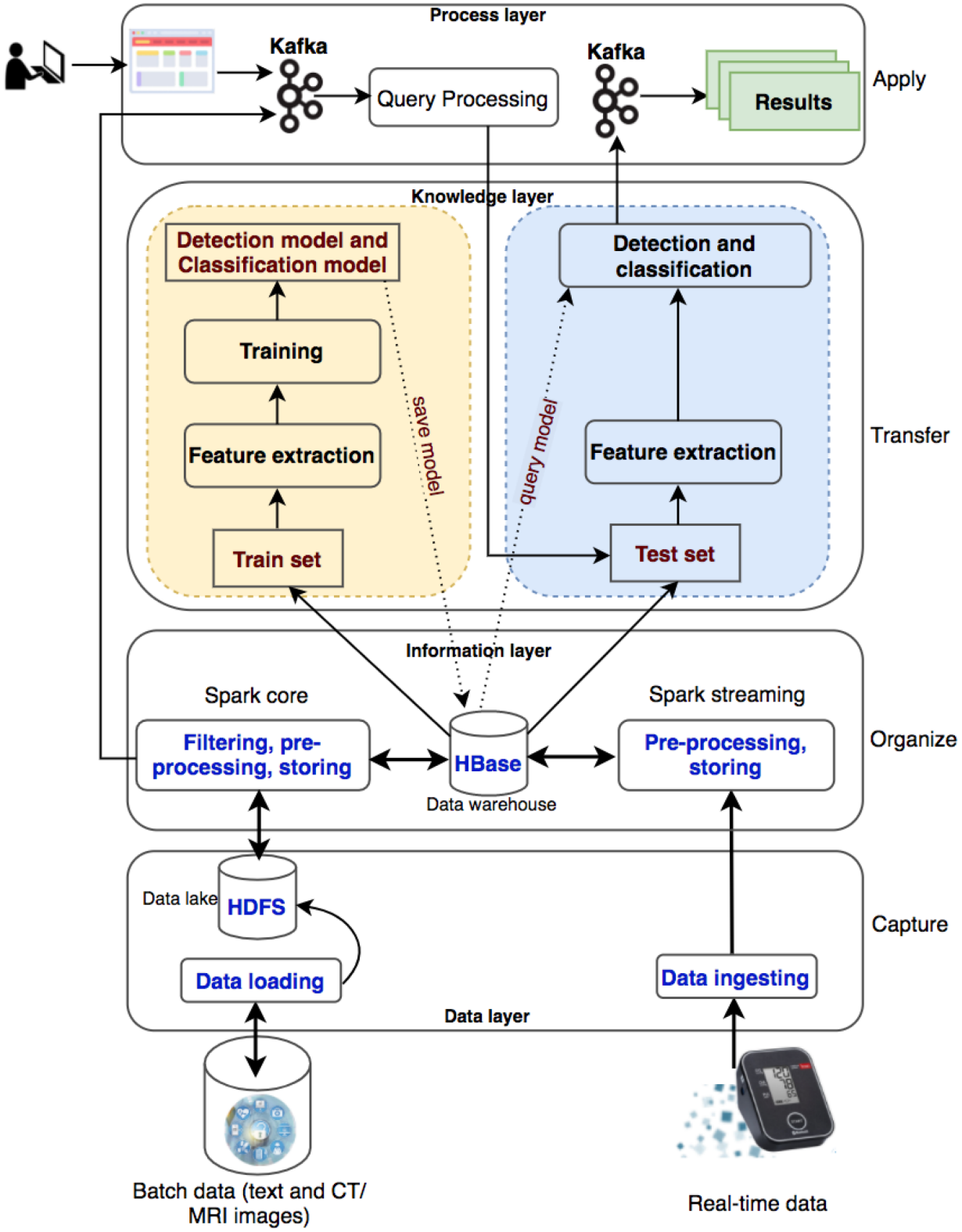

Figure 1. Proposed architecture for healthcare knowledge management systems.

The system processes two types of data: batch data (patient records collected over a long time period) and real-time data (collected from wearable devices). The batch data are loaded into the data lake (HDFS) and the real-time data are ingested into the processing system with Kafka and Spark streaming. With a large amount of medical data, the system will filter out useful information for disease diagnosis and classification, preprocess information, and store information into HBase. The information will be used for knowledge transformation to create machine-learning models. New knowledge is created and made available to users through queries from websites or wearable devices.

2.1. Data Layer

There are two data sources used in this entry, including historical datasets collected from hospitals and real-time data collected from patients via health-monitoring wearable devices. The batch data are loaded into Hadoop Distributed File System (HDFS), a well- known fault-tolerant distributed file system. HDFS is designed to store very large datasets reliably and to stream those datasets at high bandwidth to user applications. The real- time data are ingested

into the system with Apache Kafka, a distributed, reliable, high-throughput and low-latency publish-subscribe messaging system. Kafka has become popular when it and Apache Spark are coordinated to process stream data as well as to use both of their advantages. Researchers use Kafka to ingest real-time event data, streaming it to Spark Streaming. The data can be in text format or images, especially CT/MRI images that are commonly used in medical diagnosis. These raw data are collected and fed into the system for storage at the data layer.

2.2. Information Layer

Data will be sorted, organized, and filtered accordingly to transform into meaningful information in an organized and retrievable form. This information will be stored as datasets on a distributed file system HBase to serve for distributed and parallel processing in a big data environment. Apache HBase is a distributed column-oriented NoSQL database built on top of HDFS. The system requires the ability to handle batch and real-time data. Consequently, researchers use Apache Spark for both the batch and real-time data processing. Spark has emerged as the next-generation big-data-processing engine because it works with data in memory that are faster and better able to support a variety of compute-intensive tasks. Spark Core processes the batch data from HDFS to organize content according to their semantics and to create and maintain the knowledge base (HBase) as an organizational memory. Spark Streaming involves mapping continual input of the data from Kafka into real-time knowledge views. Every single event is sent as a message from Kafka to the Spark Streaming. Spark Streaming produces a stream and executes window-based operations on them.

The data collected from the hospital management system consist of many tables and many data fields. Depending on the goals of the medical diagnostic support systems, the appropriate data should be extracted. The historical datasets collected from hospitals will be used for the knowledge generation process, which is the input to the knowledge layer. These data are authentic, and the diagnostic results are given by the doctors with high professional confidence to help the labeling process in building knowledge models more effectively.

2.3. Knowledge Layer

Machine-learning algorithms can be used in the Spark distributed environment to build models for knowledge generation consisting of two phases: the training phase and the testing phase. Spark MLib is a core component to execute the learning service that allows for quickly experimenting and building data models. The appropriate models supporting diagnosis decisions will be made based on accuracy. In this layer, it is necessary to perform preprocessing of the data, which is to select the necessary information for the construction of a diagnosis support system. The diagnosis results previously given by doctors are used for labeling purposes. After data preprocessing, 70% of the random dataset will be used for the training phase and 30% for the testing phase.

The machine-learning algorithms used in the knowledge layer are decision trees and deep neural networks. Decision trees have been successfully used in a wide range of fields such as speech recognition, remote sensing, and medical diagnosis. The reason for choosing a decision tree at the knowledge layer here is that the patient records for hypertension are all in text format. The decision tree uses input data to learn and generate knowledge with the same rules as to how humans think. It breaks down a complex decision-making process into simple rules that are simple to understand and suitable to use for datasets of diverse attributes and data types. Deep learning with Faster R-CNN Inception ResNet v2 is another machine-learning algorithm to be used in the knowledge layer for brain hemorrhage diagnosis. Deep-learning techniques have been successfully applied in a wide range of fields, especially in medical images analysis.

Training phase: In this phase, researchers perform feature extraction on the input dataset and then train machine-learning models. Model training is performed in a distributed environment and stores the trained model on distributed file systems (**Figure 2**). Researchers build machine-learning models with the extracted feature dataset.

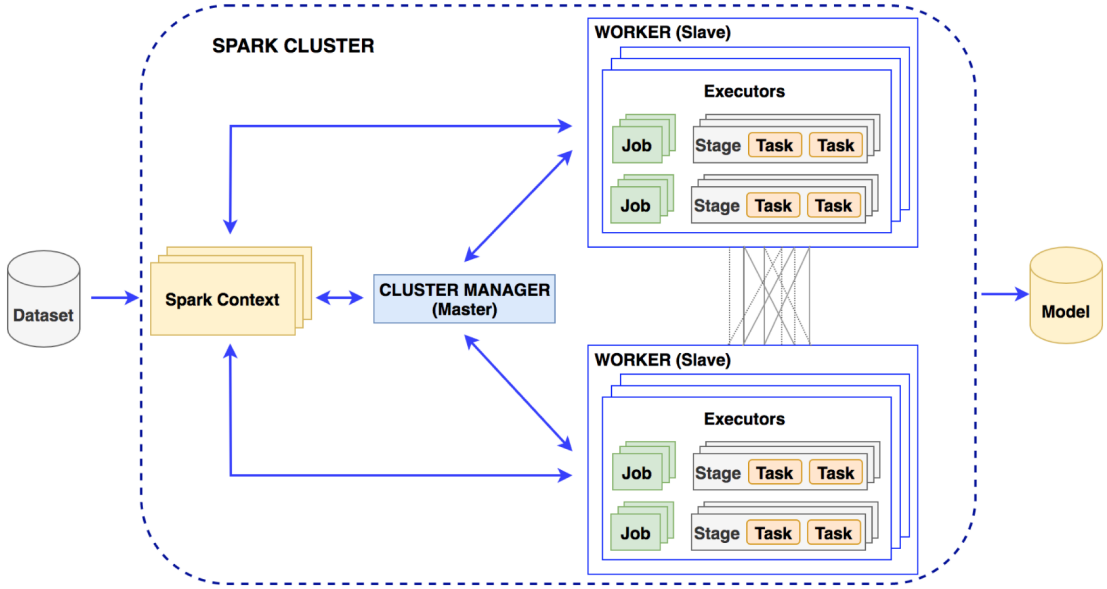

Figure 2. Training phase in a Spark cluster.

Testing phase: Researchers extract features for the testing set, thereby evaluating the accuracy of the trained models with the test set. The trained model is used to predict whether or not a patient has a disease. The execution of queries in this phase is also implemented in a distributed parallel environment. Machine-learning models are used in the testing phase to evaluate the accuracy of the predictions. The models' performance can be evaluated with precision, recall, and F1 score. The appropriate models for the problem will be stored on a distributed storage system for future use.

2.4. Process Layer

In this layer, the applications are built to input patient information into the system and give outputs about diagnosis and diseases classification. The applications are designed to perform patient data entry and then execute knowledge queries to return new knowledge about the patient's health status. The execution of queries in this layer is implemented in a distributed environment.

3. Healthcare Knowledge Management Systems

3.1. High Blood Pressure Diagnosis Support

Blood pressure is the blood force exerted against vessel walls as it moves through the vessels [15]. Blood pressure is expressed as two numbers: systolic pressure and diastolic pressure. Systolic is the higher number, which corresponds to the period when the heart beats to push the blood in the arteries. Diastolic is the lower number, which corresponds to the rest period between two consecutive heartbeats. Typically, high blood pressure is when the blood pressure measured in medical facilities is greater than or equal to 140/90 mmHg. According to the seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7) [16], the classification of blood pressure for adults aged 18 and older is presented in **Table 1**.

Table 1. Classification of blood pressure for adults.

Class	Systolic	Diastolic
Normal	**<120**	**and <80**
Prehypertension	**120–139**	**or 80–89**
Stage 1 hypertension	**140–159**	**or 90–99**
Stage 2 hypertension	**≥160**	**or ≥100**

3.1.1. Decision Tree for High Blood Pressure Detection

Preprocessing: The text data have a lot of empty data, zero value data, and even non-viable values that will affect the operations of the knowledge layer. Therefore, data preprocessing will remove non-viable values from the dataset. The solution to empty data fields is filling values using mathematical interpolation. This dataset is saved as a csv extension file and put on HBase for later use in distributed environments.

Researchers label the data records based on the diagnosis results, which are concluded by professional doctors with high reliability. The data record is labeled 1 if the patient is diagnosed with high blood pressure and 0 otherwise. After labeling, researchers process the string information in the dataset to build a feature extraction model and receive the feature vectors.

Model training: Researchers fit a decision tree with a ratio of 70/30 for training and testing phases. A classification decision tree is built with the train set, and then researchers will use the test set to evaluate the model performance. **Table 2** contains the information of the dataset after labeling and feature extraction. This information is obtained during the steps researchers take before dividing train/test sets.

Table 2. Examples of data before training models.

Symptoms	Diagnosis	Label	Index	Symptoms Classification	Features
Headache, vomit	**Intracranial injury**	**0**	**194**	**(25,152, [194], [1.0])**	**(25,163, [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 205], [17.0, 100.0, 60.0, 80.0, 18.0, 1.57, 22, 53, 48.0, 37.0, 1.0])**
Fiver	**Chickenpox**	**0**	**7**	**(25,152, [7], [1.0])**	**(25,163, [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 18], [1.0, 36.0, 140.0, 60.0, 78.0, 20.0, 1.7, 39, 68, 50.0, 39.0, 1.0])**
Tired	**Hypertension**	**1**	**1**	**(25,152, [1], [1.0])**	**(25,163, [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12], [49.0, 210.0, 140.0, 104.0, 22.0, 1.73, 40, 55, 80.0, 37.0, 1.0])**
Abdominal pain	**Acute appendicitis**	**0**	**0**	**(25,152, [0], [1.0])**	**(25,163, [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], [23.0, 110.0, 70.0, 87.0, 20.0, 1.46, 40.0, 50.0, 40.0, 37.0, 1.0])**
Dizzy	**Vestibular dysfunction; Hypertension**	**1**	**4**	**(25,152, [4], [1.0])**	**(25,163, [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15], [1.0, 53.0, 170.0, 100.0, 84.0, 18.0, 1.5, 42, 55, 50.0, 37.0, 1.0])**

In addition, based on the trained model, researchers use the featureImportances function supported by PySpark library to select variables that have an important influence on the disease diagnosis in the dataset. The importance of a variable is weighted by Gini-importance defined by the total decrease in node impurity. It is calculated by the number of samples that reach the node, divided by the total number of samples. The higher the value, the more important the feature is. Researchers can rely on this result to remove unimportant data fields to reduce training time as well as increase the accuracy of the model. The results researchers obtained from the featureImportances are shown in **Figure 3**.

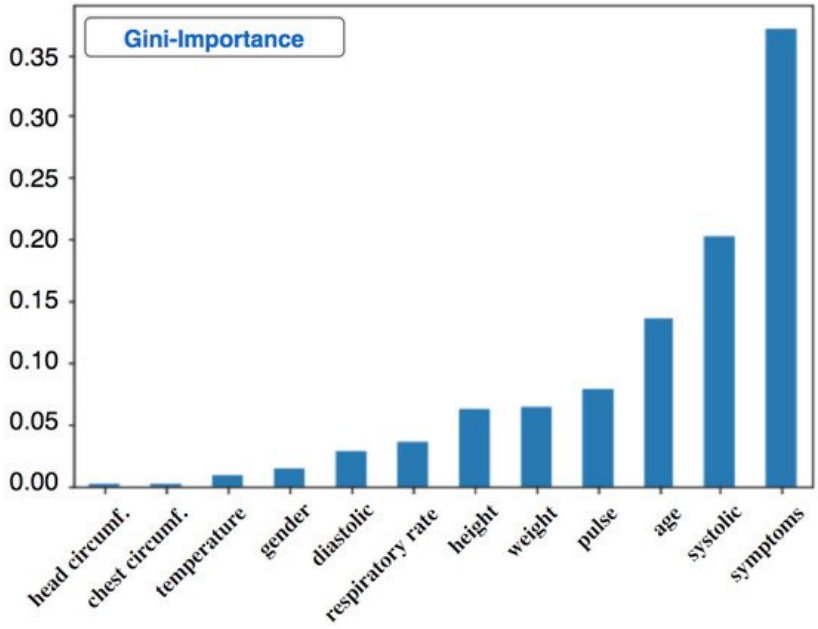

Figure 3. Feature importance in predicting high blood pressure.

Researchers decided to remove two unimportant data fields (head circumference and chest circumference) and retrain the models with the dataset consisting of only 11 data fields. Researchers train different decision tree models by varying the tree depth as well as performing the training phase in a distributed environment with three proposed scenarios.

Training results: Researchers construct decision trees with different depths. Each tree will have rules that give different prediction results. A tree of depth n will inherit inner branches from a tree of depth n−1 and has additional conditions for making predictions. An example to illustrate a decision tree with a depth of 4 is shown in **Figure 4**.

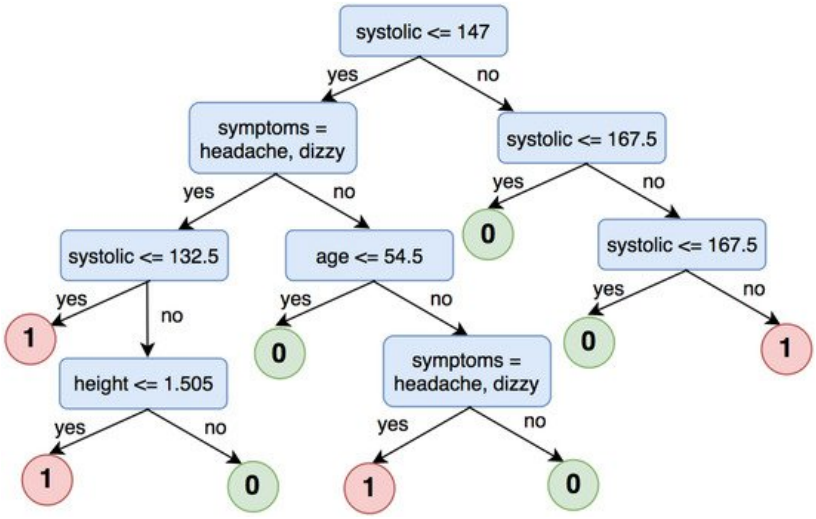

Figure 4. Decision tree of depth 4 for the problem high blood pressure detection.

In addition, based on the decision tree models and the rules generated, researchers found that several health factors of the patient are closely related to high blood pressure. For example, a patient with the systolic blood pressure of over 147 usually has some symptoms such as headache, dizziness, and fatigue. People over the age of 55 are likely to have a high risk of hypertension. Researchers train the models on a Spark cluster, and the training time is presented in **Figure 5**a. The deeper the tree, the more time it spends on the training process. After finishing the training process, researchers evaluate the detection models by applying the models for high blood pressure detection on the testing set. The accuracy of the models received is presented in **Figure 5b**. The precision of the models with different tree depth levels reaches 84% to 87%. After the process of training and evaluating the results of the models, researchers choose to stop training at a tree depth of 6 because the generated rules are consistent with reality. These things considered, if researchers increase the depth of the tree, researchers find that redundant branches start to appear, and the decision trees fall into over-fitting.

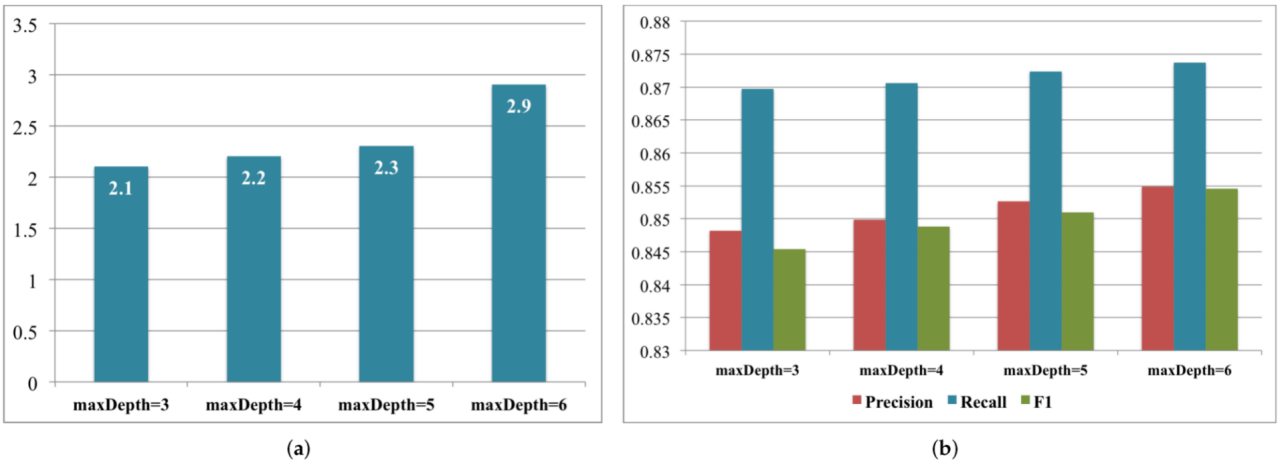

Figure 5. Training time and accuracy of the detection models. (**a**) Training time; (**b**) Accuracy.

3.1.2. Decision Tree for High Blood Pressure Classification

Model training: The classification of high blood pressure is based on **Table 1**. Researchers perform labeling by comparing the patient's systolic and diastolic blood pressure to make the classification as follows.

- Label 0: systolic < 120 and diastolic < 80
- Label 1: systolic ≥ 120 and diastolic ≥ 80
- Label 2: systolic ≥ 140 and diastolic ≥ 90
- Label 3: systolic ≥ 160 and diastolic ≥ 100

The classification of the disease is conducted after the disease detection; thus, researchers do not pay attention to label 0. Researchers train decision trees for classification problems on the same dataset with the ratio of 70/30 for train/test sets on the three proposed scenarios.

Results: Similar to the detection of hypertension, researchers build a classification model of high blood pressure with decision trees at different depths. Researchers choose to stop training at a tree depth of 4 because as the depth increases, redundant branches start to appear, and the tree falls into over-fitting. An example of a decision tree that classifies hypertension with a tree depth of 4 is shown in **Figure 6**.

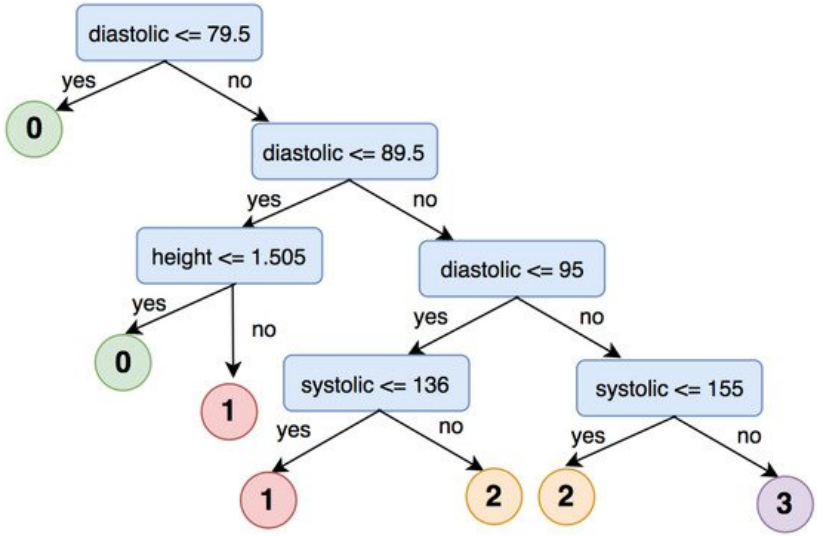

Figure 6. Decision tree of depth 4 for the problem of high blood pressure classification.

The classification models are trained on a Spark cluster. The training time is presented in **Figure 7**a. The deeper the tree, the more time it spends on the training process. Researchers evaluate the classification models based on precision, recall, and F1-score. The accuracy of the models received is presented in **Figure 7**b. Researchers receive a precision of over 92% all over the three models.

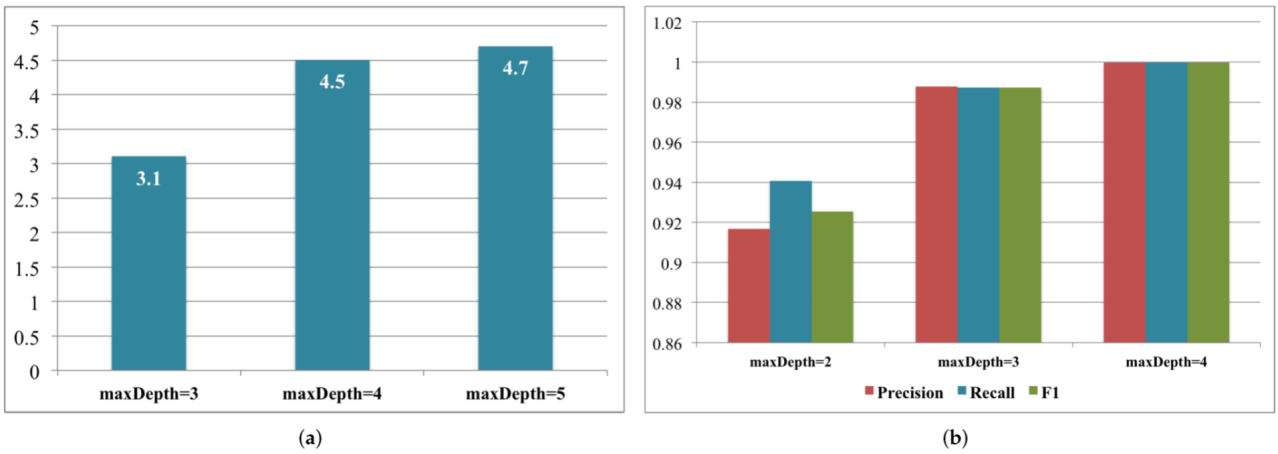

Figure 7. Training time and accuracy of the classification models. (**a**) Training time; (**b**) Accuracy.

3.2. Brain Hemorrhage Diagnosis Support

Brain hemorrhage is a dangerous disease, being a type of stroke that can lead to death or disability. There are four common types of cerebral hemorrhage [27]: epidural hematoma (EDH), subdural hematoma (SDH), subarachnoid hemorrhage (SAH), and intracerebral hemorrhage (ICH). Hypertension is the most common cause of primary intracerebral hemorrhage. To detect the brain hemorrhage, doctors usually rely on the Hounsfield Units (HU) of the hemorrhage region in a CT/MRI image. Thus, researchers propose a diagnosis supporting system for brain hemorrhage detection and classification using HU values. The machine-learning algorithm to be used in the knowledge layer for this type of disease is deep learning, which is mentioned in this entry as Faster R-CNN Inception ResNet v2.

Hounsfield unit represents different types of tissue on a scale of −1000 (air) to 1000 (bone). **Table 3** illustrates different tissues with their HU density. The hemorrhagic region will have HU values in the range of 40 to 90. The HU values are

calculated by Equation (1) with pvalue being the value of each pixel and rslope and rintercept being the values stored in CT/MRI images.

$$HU = p_{value} * r_{slope} * r_{intercept} \quad (1)$$

Table 3. HU density on CT/MRI images.

Matter	Density (HU)
Air	**−1000**
Water	**0**
White matter	**20**
Gray matter	**35–40**
Hematoma	**40–90**
Bone	**1000**

3.2.1. Training Phase

Preprocessing: The CT/MRI images will be converted into digital images (.jpg) according to the HU values. The location of brain hemorrhage is determined by HU values; thus, after preprocessing, researchers will have a digital images dataset with highlighted hemorrhagic regions. The hemorrhagic regions will be labeled with the supervision of specialists.

Feature extraction: Researchers perform feature extraction using a pretrained CNN of Inception ResNet v2 as the backbone of the Faster R-CNN to reduce the computation time. This step helps to quickly classify brain hemorrhage.

Model training: The extracted features are trained on Faster R-CNN. This training process is monitored with the Loss value. When the Loss value is not improved (or not decreased), researchers stop the training process. The Loss value of the model is very low (below 10%) after 60,000 training steps, as illustrated in **Figure 8**. This means that the error rate in the brain hemorrhage prediction of the proposed model is very low.

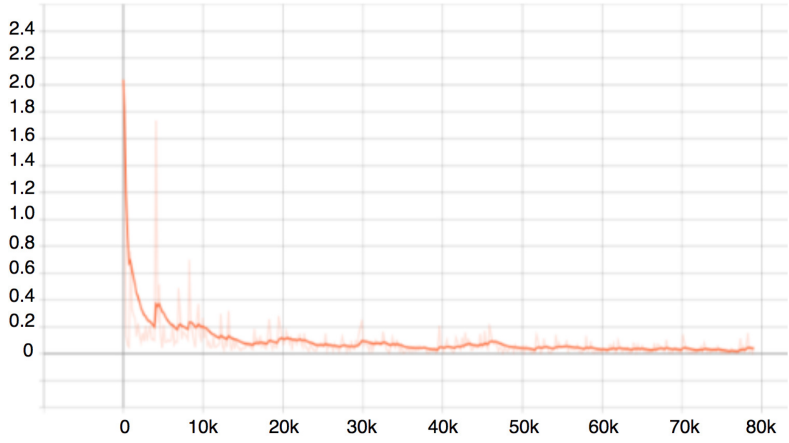

Figure 8. Loss values over training steps.

3.2.2. Testing Phase

After the training process, researchers evaluate the proposed model for brain hemorrhage detection and classification on the test dataset. The preprocessing and feature extraction are also performed on the testing set before evaluating the model. The trained Faster R-CNN Inception ResNet v2 is then used to detect and classify four common types of brain hemorrhage. It can correctly detect the contours of entire hemorrhage regions with an accuracy of 100%. An example of multiple hemorrhages detection on an image is presented in **Figure 9**. It can predict bleeding time from 2 to 3 days, recognize hemorrhage type as ICH and SAH, and accurately segment bleeding regions.

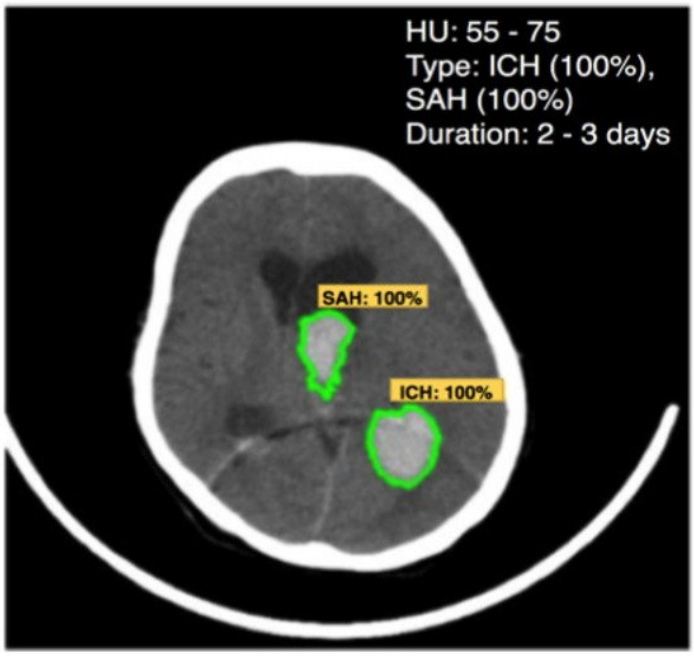

Figure 9. Multi- brain hemorrhages segmentation.

The average precisions (AP) of the proposed model for four types of brain hemorrhage (EDH, SDH, SAH, and ICH) are 0.7, 0.59, 0.72, and 0.71, respectively (**Figure 10**). This model gives the mAP value of 0.68 for the detection and classification of four classes of brain hemorrhage. The results show that the system can support doctors in accurately diagnosing cerebral hemorrhage and providing appropriate treatment regimens.

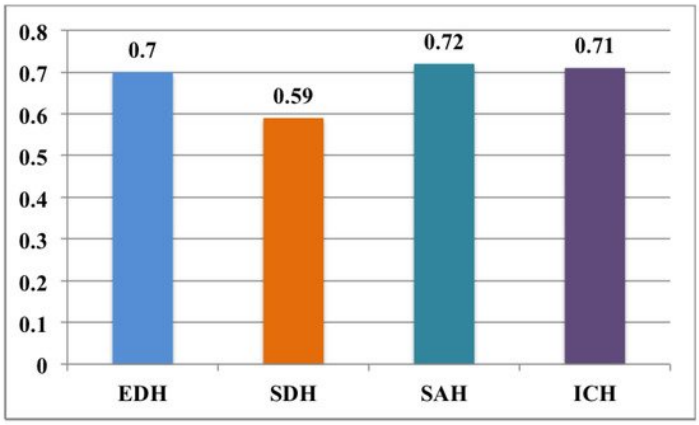

Figure 10. Average precision (AP) of four brain hemorrhage types.

References

1. Premier Healthcare Database being used by National Institutes of of Health to Evaluate Impact of COVID-19 on Patients Across the U.S . Premier. Retrieved 2022-5-12
2. Benjamin I. Chung; Jeffrey J. Leow; Francisco Gelpi-Hammerschmidt; Ye Wang; Francesco Del Giudice; Smita De; Eric P. Chou; Kang Hyon Song; Leanne Almario; Steven L. Chang; et al. Racial Disparities in Postoperative Complications After Radical Nephrectomy: A Population-based Analysis. *Urology* **2015**, *85*, 1411-1416, 10.1016/j.urology.2015.03.001.
3. Hoiwan Cheung; Ye Wang; Steven L. Chang; Yash S Khandwala; Francesco Del Giudice; Benjamin I. Chung; Adoption of Robot-Assisted Partial Nephrectomies: A Population-Based Analysis of U.S. Surgeons from 2004 to 2013. *Journal of Endourology* **2017**, *31*, 886-892, 10.1089/end.2017.0174.
4. Chung, Kyung Jin and Kim, Jae Heon and Min, Gyeong Eun and Park, Hyoung Keun and Li, Shufeng and Del Giudice, Francesco and Han, Deok Hyun and Chung, Benjamin; Changing trends in the treatment of nephrolithiasis in the real world. *Journal of Endourology* **2019**, *33*, 248--253, .
5. Maryam Alavi; Dorothy Leidner; Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly* **2001**, *25*, 107-136, 10.2307/3250961.

6. U. Of Maryland Maryam Alavi; Insead Dorothy Leidner; Knowledge Management Systems: Issues, Challenges, and Benefits. *Communications of the Association for Information Systems* **1999**, *1*, 7, 10.17705/1cais.00107.
7. Brent Gallupe; Knowledge management systems: surveying the landscape. *International Journal of Management Reviews* **2001**, *3*, 61-77, 10.1111/1468-2370.00054.
8. Fabian M. Suchanek; Gerhard Weikum; Knowledge bases in the age of big data analytics. *Proceedings of the VLDB Endowment* **2014**, *7*, 1713-1714, 10.14778/2733004.2733069.
9. Begoli, E.; Horey, J. Design principles for effective knowledge discovery from big data. In Proceedings of the 2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture, 24 Aug 2012; pp. 215–218.
10. Dong, X.; Gabrilovich, E.; Heitz, G.; Horn, W.; Lao, N.; Murphy, K.; Strohmann, T.; Sun, S.; Zhang, W. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014; pp. 601–610.
11. Nor'Ashikin Ali; Alexei Tretiakov; Dick Whiddett; Inga Hunter; Knowledge management systems success in healthcare: Leadership matters. *International Journal of Medical Informatics* **2016**, *97*, 331-340, 10.1016/j.ijmedinf.2016.11.004.
12. Maramba, George and Coleman, Alfred and Ntawanga, Felix F; Causes of Challenges in Implementing Computer-Based Knowledge Management Systems in Healthcare Institutions: A Case Study of Private Hospitals in Johannesburg, South Africa. *The African Journal of Information Systems* **2020**, *12*, 4, .
13. Manogaran, G.; Thota, C.; Lopez, D.; Vijayakumar, V.; Abbas, K.M.; Sundarsekar, R. Big data knowledge system in healthcare. In Internet of Things and Big Data Technologies for Next Generation Healthcare; 2017; pp. 133–157.
14. Le Dinh, T.; Phan, T.C.; Bui, T. Towards an architecture for big data-driven knowledge management systems. In Proceedings of the 22nd Americas Conference on Information Systems (AMCIS 2016), 2016; Association for Information Systems: USA, 2016.
15. American Heart Association. What Is High Blood Pressure? South Carolina State Documents Depository: Washington, DC, USA, 2017; pp. 1–2.
16. Chobanian, A.V.; Bakris, G.L.; Black, H.R.; Cushman, W.C.; Green, L.A.; Izzo, J.L., Jr.; Jones, D.W.; Materson, B.J.; Oparil, S.; Wright, J.T., Jr.; et al. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. Hypertension 2003, 42, 1206–1252.
