

Research Article

Research on Detection and Diagnosis Technology of Subarachnoid Hemorrhage Based on New Association Classification Algorithm

Long Chen and Changlong Zhou 

Department of Neurosurgery, Yongchuan Hospital, Chongqing Medical University, Chongqing 402160, China

Correspondence should be addressed to Changlong Zhou; 13598747336@hait.edu.cn

Received 1 November 2021; Revised 15 December 2021; Accepted 17 December 2021; Published 24 March 2022

Academic Editor: Rahman Ali

Copyright © 2022 Long Chen and Changlong Zhou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Subarachnoid hemorrhage (SAH) is one of the most common cerebrovascular emergencies, which can lead to serious consequences. Spontaneous subarachnoid hemorrhage accounts for about 15% of acute cerebrovascular accidents. Among them, spontaneous subarachnoid hemorrhage caused by rupture of an intracranial aneurysm or vascular malformation is more common, accounting for about 85%. Therefore, it is very important to detect the early symptoms of subarachnoid hemorrhage through reasonable means and to carry out appropriate clinical intervention and treatment. With the development of imaging technology, computed tomography angiography (CTA) is widely used in clinical practice. However, the accuracy of manual recognition of CT images is not very high, and the efficiency is low. The emergence of data mining technology is gradually solving this problem. In this paper, we introduce and summarize the development of data mining, domestic and foreign research progress, the application status of data mining in the medical field, and the main technologies and methods of data mining. We study the application of association rule extraction technology in data mining in the medical field. The Apriori algorithm for finding frequent item sets in association rule extraction and its series of improved algorithms are studied and finally combined the characteristics of medical CT images, an image mining method of association rules based on the gray-level cooccurrence matrix is proposed. Based on the FP-growth algorithm, the NCFP-growth algorithm based on association rules is proposed and compared with the mining effect of several other algorithms. The proposed algorithm achieves a classification accuracy of above 90%, which is higher than the Apriori algorithm and its improved variations.

1. Introduction

Subarachnoid hemorrhage (SAH) is a type of acute cerebrovascular disease caused by multiple causes. It is caused by the rupture of blood vessels on the surface of the brain or spinal cord, and blood flows into the subarachnoid space. Its incidence is second only to cerebral thrombosis. Formation and hypertensive intracerebral hemorrhage [1], according to different causes, clinically, subarachnoid hemorrhage is divided into two major categories: spontaneous and traumatic. Spontaneous subarachnoid hemorrhage accounts for about 15% of acute cerebrovascular diseases. Intracranial aneurysms, brain/spinal arteriovenous malformations, and hypertensive atherosclerosis are the most common causes of this disease. In addition, moyamoya disease, brain vasculitis,

malignant tumors, hematological diseases, meningitis, encephalitis, and complications of anticoagulation therapy can cause the occurrence of this disease [2]. Among them, spontaneous subarachnoid hemorrhage caused by rupture of an intracranial aneurysm or vascular malformation is the most common, and the others are relatively rare. In spontaneous subarachnoid hemorrhage caused by rare etiologies, perimesencephalic nonaneurysmal subarachnoid hemorrhage (PNSH) and atraumatic convexal subarachnoid hemorrhage (cSAH) are their main subtypes. In 1991, Rinkei proposed the definition of PNSH: The central part of the subarachnoid hemorrhage is limited to the front of the midbrain. The hemorrhage may be accompanied by expansion to the base of the annular cistern. The anterior part of the longitudinal fissure cistern is not completely filled. It

does not extend to the lateral fissure cistern, and there is no clear intracerebral hematoma formation [3]. PNSH accounts for about 15% of all spontaneous subarachnoid hemorrhage [4]. sSAH refers to the subarachnoid hemorrhage confined to the sulcus gyri of the superficial cerebral cortex, it usually does not involve the adjacent large and cerebellar parenchyma, the anterior and posterior longitudinal fissure cistern, the subcutaneous cistern or the ventricle, etc., and the bleeding range is small [5], which accounts for about spontaneous subarachnoid hemorrhage. 7.45% [6]. Therefore, early diagnosis and early treatment are an important part of the prevention and treatment of the sequelae of subarachnoid hemorrhage. Intracranial aneurysmal subarachnoid hemorrhage is a typical representative of spontaneous subarachnoid hemorrhage, and it is clinically one of the most common cerebrovascular diseases. The annual incidence is about 8–16 cases per 100,000, accounting for about 85% of patients with spontaneous subarachnoid hemorrhage [7]. Intracranial aneurysm rupture is insidious and develops rapidly. Rapidly, the fatality rate of the first hemorrhage is as high as 40%, and the disability rate is as high as 33% [8, 9]. If it is not diagnosed and treated in time, the mortality rate of rerupture is as high as 60–70%, and the damage is huge [10]. There are about 60 million patients with aneurysms, and more than 200,000 cases of intracranial aneurysms rupture and hemorrhage each year, which seriously threatens people's lives and health and brings heavy psychological and economic burdens to society and families. Therefore, early detection should be conducted through reasonable means.

It is very important to extract intracranial aneurysms and conduct appropriate clinical interventions. Data mining technology has received more and more attention since its birth. It is good at dealing with massive incomplete, noisy but noisy practical applications. For data with hidden value, data mining conducts a highly intelligent analysis of these data. Through induction, generalization, and reasoning, the potential information in the data is found. At the same time, the data itself is continuously improved through the mining process so that the data can be fully obtained, explained, and used to the maximum [11]. Medicine and research in the relevant fields are becoming dependent on new technologies and they impact the practice. We need to store and process large amounts of data to be able to use the data and extract insights from it [12]. These data are extremely valuable for disease diagnosis, disease analysis, and pathological research, but most of the current database systems cannot intelligently process these data, and they cannot be found in them. The knowledge of data cannot predict the future development trend of data. Efficient medical data mining can improve the level of hospital information management, provide accurate information and patterns in the diagnosis and treatment of diseases, and help doctors make scientific and correct decisions [13, 14]. This article introduced the development and current status of data mining and knowledge discovery and the current status of data mining in medical applications. The basic ideas, overall framework, and main technologies and methods of data mining are introduced in detail, and then the particularity of medical

data is analyzed. And combine the two to propose a process model suitable for medical data mining. Then according to the extraction of association rules and their application in medical image mining, the related theories of association rules are introduced, the main algorithms of association rule extraction are studied, and then images are combined. Gray-level symbiosis matrix extracting subarachnoid hemorrhage texture features of CT images and the association rules are extracted to achieve the purpose of auxiliary diagnosis and classification.

2. Related Work

Perimetral nonaneurysmal subarachnoid hemorrhage (PNSH) accounts for about 21–68% of the first spontaneous subarachnoid hemorrhage with negative DSA, and it is also an important subtype of subarachnoid hemorrhage. At the onset of the disease, the symptoms and signs are mild, the bleeding site is limited, the clinical treatment process is good, the cerebral angiography is negative, complications such as vasospasm or hydrocephalus are rare, the prognosis is good, and there are few recurrences. It is related to the aneurysmal subarachnoid serious consequences of cavity hemorrhage are completely different, and it can be considered as a special type of benign subarachnoid hemorrhage. For such a disease with low morbidity and low mortality and disability rate, the literature believes that it is meaningful for the complications caused by examination methods for the disease to be less than 0.5%. Therefore, the inspection method we choose should be noninvasive and effective. It not only reduces the damage caused by the inspection method itself to the patient, but also has a higher sensitivity, which can exclude aneurysms with a high mortality and disability rate. At the same time, it also needs to have a higher negative predictive value to correctly diagnose PNSH. In addition, another important subtype of subarachnoid hemorrhage has been discovered this year, namely spontaneous localized subarachnoid hemorrhage. Spitzer et al. first reported 12 cases of cSAH in 2005, and only a few cases have been reported before [5]. The study of Kumar et al. indicated that its incidence is about 7.45% [6]. In addition, there is no large-scale epidemiological survey to report its incidence and gender differences. Its bleeding site is different from PNSH, but the prognosis is relatively good, and it can also be regarded as another benign subarachnoid hemorrhage. The principles of diagnosis and treatment of such subarachnoid hemorrhage should also be distinguished from aneurysms. Then, for the inference of the cause of SAH, in addition to the clinical data of the patient, we often use the following imaging examination methods. Digital subtraction angiography (DSA) is currently recognized as the gold standard for intracranial aneurysm diagnosis, preoperative evaluation, and evaluation of other vascular imaging indicators [12, 15]. However, as an invasive examination, DSA has a complication rate of about 1–2%, and about 0.5% will have permanent neurological dysfunction. In severe cases, it can lead to death. In addition, there are invasive and use of angiography. The shortcomings of drug, radiation hazard, long inspection time, high cost, etc., also limit its wide

application in the screening and follow-up observation of intracranial aneurysms [16, 17], especially not suitable for the exclusion screening of benign SAH such as PNSH and follow-up. Therefore, noninvasive inspection methods are getting more and more attention. With the development of imaging technology, computed tomography angiography (CTA) is widely used in clinics, making it quiet to rely solely on DSA technology as the gold standard for the detection of intracranial aneurysms, endovascular treatment, and surgical procedures. These noninvasive vascular imaging methods ensure the effective detection of aneurysms while also effectively avoiding complications of cerebral angiography [18].

However, the accuracy of manual recognition of CT images is not very high, and the efficiency is low, and the emergence of data mining technology is gradually solving this problem. Image mining is a frontier field developed in recent years. It is the intersection of multiple disciplines, including computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence. Although these disciplines are relatively mature in their respective fields, image mining is still in the exploratory research stage. Many scholars at home and abroad have made active explorations in this field and made meaningful attempts in the following aspects:

- (1) Celestial image mining: the system uses the sky images carefully classified by astronomers as the training set to construct model is used to identify galaxies, and people have successfully used this method to identify volcanoes on Venus [19].
- (2) Satellite remote sensing image mining [20]: satellite remote sensing images are now widely used in various fields, and the goal is to use remote sensing images to solve surface problems. Detecting moving targets in remote sensing images and storing the moving target information in the database together with the original image can dig out rich knowledge, such as the relationship between targets.
- (3) Spatial data mining [21]: used for understanding spatial data, discovering spatial relationships, and relationships between spatial and nonspatial data. People hope to construct a spatial data cube and mine spatial data based on the data cube. Association analysis of spatial data is a hot research topic in this field, and some algorithms have been proposed. The NASA Jet Propulsion Laboratory of the National Aeronautics and Space Administration of the United States has researched and developed a prototype image data mining software system, namely the "Diamond Eye System," which can automatically extract knowledge containing semantic information from the image and detect the terrain of the crater. And analysis and satellite detection have been specifically applied [22].
- (4) Medical image mining [23]: a large number of medical images have become an important factor in promoting the development of image mining technology. Medical images are always accompanied by doctor's diagnosis records. There may be a large number of correlations between the diagnosis records and the visual characteristics of medical images. People have begun to devote themselves to research in this field. For example, some research groups study the space of diseased brain tissues. The correlation between the characteristics and the pathological characteristics in the diagnostic record can help doctors find the location of the lesion; some research groups use similar methods for the judgment of early breast cancer [24]. In 1998, Simon Frase University in Canada researched a prototype of image data mining software called Multimedia Miner. This system was researched and developed on the basis of the original relational data mining system "DB-Miner" and C-BIRD [25]. The system includes the use of multidimensional analysis technology to create multimedia data cubes and can be used to discover a variety of knowledge, including summary knowledge, classification knowledge, association rule knowledge, etc. One of the modules is MM-Associator, which mainly mines the association rules of images. The information connected by these rules includes image size, color, and image description [26]. The prototype system consists of three functional modules: (1) The MM-Characterizer module describes the characteristics of multimedia data from multiple abstract layers, allowing users to observe data from a multilevel perspective, providing scroll-up and drill-down functions. (2) The MM-Associator module discovers association rules from image or video data sets. (3) The MM-Classifier module classifies multimedia data according to class tags and gives a description of each class. Chen et al. [27] believe in reducing the death rate and for accurate treatment of hemorrhage, machine learning based algorithms play an important role. They propose an IoT based system using a support vector machine and feedforward network for classification. The machine learning based application is capable of giving information about the type of brain hemorrhage, eventually helping the expert's diagnosis and the treatment procedure. Wang et al. [28] present an overview of the application of deep learning algorithms for automatic detection and classification of hemorrhage in CT images. They believe that the application of AI-based systems may help to automate the process of diagnosis and eventually lead us to better and timely cure of the disease. They use a CNN-based deep learning model to achieve accurate hemorrhage detection.

At present, the research of image mining is relatively mature, which can perform various processing on medical images, so that the lesions that are difficult to observe become clearer, and at the same time provide a certain degree of auxiliary diagnosis. It has achieved a leap from film to the digitization of medical imaging, greatly simplifying the

workload of medical staff on equipment, allowing them to focus more on the diagnosis and treatment of diseases, and promoting communication between hospitals through this technology, in order to improve the overall medical level make a contribution.

3. Method

3.1. Overview of Association Rules. Association rules refer to the correlation between different items that appear in the same event. The extraction of association rules was first proposed by Agrawal et al. in 1993. After more than 20 years of development, it has become one of the most important mining techniques in data mining, and its mode is a descriptive mode. On the one hand, in the process of mining association rules, we can obtain association relationships at different conceptual levels. In the context of domain-related concept hierarchical tree support, the association rule mining method can be used to obtain associations rules that reflect the laws of different levels. On the other hand, the association rules contained in different data sets are different. There are two indicators to evaluate the rules extracted through association analysis, which are the degree of support reflecting the degree of interest of the rule and the degree of confidence reflecting the reliability of the rule. Normally, the purpose of association rule mining is to find out which is higher than the minimum support and minimum confidence threshold rules. However, in some specific cases, it may also be necessary to pay attention to low-support rules, such as disease surveillance. The typical association rule mining process is divided into two steps: the first is to find the itemsets that meet the minimum support degree, that is, frequent itemsets. The second is to generate the lowest confidence rules from frequent itemsets, that is, strong association rules. The main technical difficulty of this mining method lies in the first step, i.e., to efficiently find frequent itemsets, which has a greater impact on the performance of the algorithm. Many classic algorithms for mining association rules use this two-step mining method, such as Apriori and DHP.

3.2. Classification of Association Rules

- (1) Based on the categories of variables handled in the rules, it is divided into Boolean association rules and numerical association rules. If the association considered by the rule is the presence and absence of items, it is a Boolean association rule. If the rule describes the association between quantified items or attributes, it is a quantitative association rule.
- (2) Based on the dimensionality of the data involved in the rules, it is divided into single-dimensional association rules and multidimensional association rules. If each item or attribute in an association rule involves only one dimension and deals with some relationships in a single attribute, then it is a single-dimensional association rule. If the rule designs two or more dimensions and deals with the relationship

between various attributes, then it is a multidimensional association rule.

- (3) Based on the data abstraction level involved in the rules, it is divided into single-level association rules and multilevel association rules. If in a given rule set, the rule does not involve items or attributes of different abstraction layers, then the set contains a single-level association rule. The multilayer association rules involve items or attributes of different abstract layers.

3.3. Association Rule Process. Association rule mining refers to finding the rules that meet the minimum support threshold and the minimum confidence threshold from a given transaction database. The more primitive mining method is to calculate all possible rules a and b . But this method is obviously too inefficient. Just a small data set can extract hundreds of rules. If the minimum support threshold and the minimum confidence threshold are set to 25% and 50%, respectively, more than 80% of the rules will be eliminated. Therefore, in order to improve the efficiency of mining, the rules must be pruned first. It can be seen from the calculation formula of the support of the rule that the support of the rule $X \rightarrow Y$ only depends on the support count of the itemset $\{X \cup Y\}$. From this, it can be seen that most of the association rule extraction algorithms decompose the mining task, dividing it into two steps: finding frequent itemsets and mining association rules. The goal of the former is to find out the itemset that people are interested in (that is, higher than the preset support threshold), which is called frequent itemsets. The latter is to extract association rules higher than the preset confidence threshold from the frequent itemsets, that is, strong association rules.

3.4. Association Rule Mining Algorithm

3.4.1. Apriori Algorithm. In the analysis of association rules, the most basic and commonly used algorithm is the Apriori algorithm [29], which was proposed by the scholar Agrawal when analyzing shopping basket data. It uses multiple scans of the database to mine the single-layer Boolean association rules required frequent itemsets, the name of the algorithm is taken from the Latin Apriori, which refers to the result from the cause. Its core idea is recursion based on the frequency set theory. The Apriori algorithm is named based on the prior knowledge about frequent itemsets. It uses a circular hierarchical search to mine frequent itemsets. This loop uses k -itemsets to generate $(k + 1)$ -itemsets. First the frequent itemset of 1-item set is extremely T_1 , and then T_1 is used to mine T_2 , and T_2 is used to mine T_3 . At the same time, the data must be scanned once for each layer of mining, and it will continue to loop until no more frequent items can be mined. The algorithm first generates T_1 , then generates a candidate set of T_2 through T_1 , scans the database D , deletes part of the itemset in the candidate set, and obtains T_2 . It generates a candidate set of T_3 through T_2 and then scans the database D again. The process is repeated this way until there are no itemsets that contain more items. The Apriori

algorithm has an important property. According to the definition, if a certain itemset I does not meet the minimum support threshold, then I is not a frequent itemset. If an item a is added to the itemset I , the new itemset $a \cup I$ is in the database. The number of occurrences must be less than I , so $a \cup I$ cannot be a frequent itemset. Its inverse proposition is: if an itemset cannot meet the minimum support threshold, all possible supersets also cannot be satisfied. The following section describes the specific process of T_k generating T_{k+1} in detail and explains the application of Apriori properties to the two steps of joining and deleting in frequent itemsets mining:

- (1) Connection step: in order to mine T_{k+1} , two itemsets in T_k can be connected to obtain a candidate set of T_{k+1} , which is set as the set L_{k+1} . Let t_1 and t_2 be the two itemsets in T_k . If the itemsets t_1 and t_2 , except for the last item and the penultimate item, other items are the same; at the same time, the penultimate item is not the same. It is assumed that the records in the database have been arranged in lexicographical order, then t_1 and t_2 in T_k can be connected, and this connection principle can ensure that all and non-duplicate candidate itemsets are generated.
- (2) Deletion step: the candidate set L_{k+1} generated by the above connection principle is a superset of T_{k+1} . The deletion process here is also divided into two steps. First, delete all previously found supersets of infrequent itemsets according to the Apriori property. Then scan the database to delete itemsets whose support is lower than the minimum support threshold, and at the same time classify this part of the itemsets as infrequent itemsets.
- (3) According to the deletion step, we can know that each itemset in L_{k+1} needs to be scanned in the database before deciding whether to add it to T_{k+1} . This verification process is the bottleneck of the Apriori algorithm. For example, the coefficient $k=10$, then the database needs to be scanned ten times, which requires a large I/O load, and most of the improvements to the Apriori algorithm are also aimed at this point. After finding all frequent itemsets, it is necessary to generate association rules through frequent itemsets. The rule generation process of most algorithms including Apriori is basically the same. For each frequent itemset T , all nonempty subsets S are generated. For each nonempty subset, if $\sigma(T)/\sigma(S) \geq \text{min_conf}$, a strong association rule $S \rightarrow (T - S)$ is generated.

The Apriori Algorithm 1 [29] in general form is given as follows.

3.4.2. DHP Algorithm. In the process of mining frequent itemsets, the Apriori algorithm obtains the candidate set of the next layer through the connection of the frequent itemsets of the previous layer. After further filtering the candidate set to obtain the frequent itemsets of this layer, the cycle continues. The biggest bottleneck is that this screening

process needs to calculate the support of each itemset in the candidate set through a one-to-one comparison with the database so that when the number of itemsets in the candidate set increases, the efficiency of the algorithm will be great. Therefore, reducing the number of candidate sets to the greatest extent can reduce the number of comparisons and improve the efficiency of the algorithm. DHP (Direct Hasmgarld Prulling) algorithm introduces a hash table structure to prune unnecessary candidate sets so as to improve the efficiency of mining association rules. The specific method is to establish a hash bucket in the layer-by-layer loop, that is, in the process of deriving $(k+1)$ -itemsets through k -itemsets, and use the hash buckets to further filter the candidate sets. At the same time, to update the database with frequent itemsets. Although the DHP algorithm spends resources to maintain the hash bucket and update the database, it can greatly reduce the number of itemsets in the candidate set, thereby greatly reducing the number of comparisons and significantly improving efficiency.

3.4.3. Partition Algorithm. In terms of mining frequent itemsets and obtaining rules, the principle of the Partition algorithm is basically the same as that of the Apriori algorithm. The difference is that the Partition algorithm reduces the comparison cost of each scan of the database by segmenting the database. The core idea is as follows: A certain itemset is a frequent itemset for the entire database, then the itemset must be frequent on a segment of the database. The algorithm is mainly divided into two steps: the first step is to segment the database and perform a frequent itemset mining algorithm for each segment to obtain the frequent itemset of the segment. The second step is to combine all the segmented frequent itemsets to obtain a large candidate set, and then through the comparison of the large candidate set and the entire database to filter and verify the true frequent itemset of the entire database. The idea of divide and conquer adopted by the Partition algorithm, the entire process actually only searches the database twice, so I/O consumption is greatly reduced, but in order to avoid repeated frequent itemsets between different segments, it must be implemented before mining to sort the database, the time and resources consumed also limit the application of the Partition algorithm to a certain extent.

3.4.4. FP-Growth Algorithm. The FP-growth algorithm can skip the candidate itemset step and directly generate frequent itemsets. The FP-growth algorithm also uses the idea of divide and conquer, but its strategy is divided into two steps: The first step is to compress the entire database into a FP-tree for the first scan, and retain the itemset information, and then divide the compressed database into a group of condition databases that are associated with frequent items one by one; the second step is to mine these condition databases one by one. Among the several improved Apriori algorithms listed, the FP-growth algorithm is the most different from it. Their classification rules for frequent itemsets are different. The Apriori algorithm divides them into 1-itemsets, 2-item set,, k -item set according to the

```

L1 = {large 1-itemsets}
For (k = 2; Lk-1 ≠ ∅; k++) do
  Ck = Apriori-gen (Lk-1)
  for (all transactions t ∈ Ct, do
    increment c.count
  end
  Lk = {c ∈ Ck | c.count ≥ minsup}
End
Solution = Uk Lk

```

ALGORITHM 1: The Apriori Algorithm 1 in general form.

length of frequent itemsets, mining is carried out in the order of increasing length, while the FP-growth algorithm classifies frequent items according to the decreasing order of support. The frequent items in any transaction are inserted into the FP-tree after sorting and then recursively mining on the FP-tree.

3.4.5. NCFP-Growth Algorithm. Since the problem of mining association rules was put forward, people have constantly pointed out the limitations of association rules. In order to avoid generating illusory association rules, people have introduced various new thresholds to strengthen the evaluation of association rules. Therefore, based on the FP-growth algorithm, combined with the new threshold, this paper proposes an improved frequent pattern tree construction algorithm NCFP-growth (New Criteria Frequent Pattern-growth) algorithm. Through the introduction of interest degree weights, the algorithm effectively further filters frequent items, thereby reducing a large number of redundant and false rules generated when the system adopts the FP-growth algorithm. In addition, compared with the FP-growth algorithm, this algorithm effectively reduces the size of the tree and the system storage space when constructing the frequent pattern tree, and the search space of the algorithm is also effectively compressed. The construction of NCFP-tree: The input is the transaction database DB; the minimum support threshold is min_sup; the minimum interest weight is min_up; the output is the complete set of frequent patterns. The method is as follows: (1) scan the transaction database DB; (2) find the set F of frequent items and the corresponding support from the minimum support min sup; (3) arrange the items of W in descending order according to the degree of support, and record the result as table L; (4) create the NCFP-tree root node, record it as root, and its value is NULL; (5) Perform each

transaction in the transaction database DB (6) and (7) two steps; (6) sort the frequent items that meet min_up in each transaction in the order in L; the sorted table is marked as [p | P, T], where p is the first element, and P is the list of remaining elements; (7) if T has a child N such that N_item name = p_item name, the count of N is increased by 1; otherwise, a new node N is created, and its count is set to 1, and link to its parent node T; link the node chain of N to the node with the same item name through the node chain structure; if P is not empty, repeat this step. After constructing NCFP-tree, the frequent pattern mining process of the NCFP-tree is the same as the mining method of the FP-tree. The flowchart of the NCFP-growth algorithm is shown in Figure 1.

Observing the NCFP-growth algorithm, it can be seen that due to the addition of a new threshold, the original frequent items in the database are further filtered to avoid the system from generating a large number of redundant and false rules, making it easier for users to mine more practical association rules that they are interested in.

3.5. Gray-Level Cooccurrence Matrix of the Image. Any image is composed of many pixels, and the two pixels have the same or different gray levels. In the image, two pixels separated by a certain distance, the statistical form of the joint distribution of their gray levels can be used to analyze the texture information of the image well. The gray-level cooccurrence matrix is a texture analysis method based on the conditional probability density function of the second-order combination of estimated gray levels. Defined in the entire image, starting from a pixel with a gray level of i , the probability of another pixel with a distance of (Dx, Dy) having a gray level of j is called frequency, and the formula is as follows:

$$P(i, j, d, \theta) = \{(x, y) \mid f(x, y) = i, f(x + D_x, y + D_y) = j, x, y = 0, 1, 2, \dots, N - 1\}, \quad (1)$$

where x, y are the coordinates of the image pixels, i, j represent the gray level of the image, θ refers to the direction of the two pixels, and there are four values: 0° , 45° , 90° , and 135° . In this way, the texture information of the image is

described by the gray-level pair (i, j) . Obviously, the gray-level cooccurrence matrix formed is a symmetric matrix. Taking the element $M(1, 1)$ of the gray-level cooccurrence matrix as an example, $M(1, 1) = 1$ means that there is only a

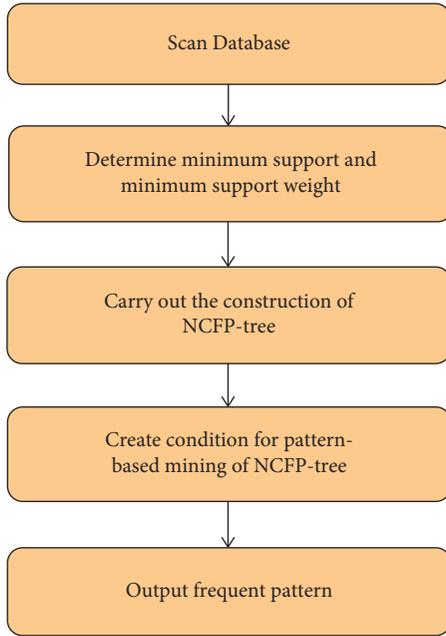


FIGURE 1: NCFP-growth algorithm process.

pair of pixels with a gray level of 1 that are horizontally adjacent in the original image. $M(1, 2) = 2$, because there are two pairs of pixels with gray scales of 1 and 2 horizontally adjacent in the original image.

3.6. Association Rule Mining Method. In summary, before mining the arachnoid CT images, the medical features of the arachnoid CT images must be converted into mathematical texture features through the gray-level cooccurrence matrix, as shown in Figure 2.

Before extracting features, it is necessary to perform image cropping, image noise reduction, and enhancement of the image. After completion, the gray-level cooccurrence matrix of the CT image can be constructed. The construction of the matrix is mainly to determine the gray-level L and step. The two parameters of long D , the amount of data, and the amount of calculation of the image gray-level cooccurrence matrix are determined by these two parameters. Suppose the image format is 288×288 . If the image does not undergo any compression and takes the gray-level $L = 256$ and the step size $D = 1$, the time consumption of calculating the gray-level cooccurrence matrix will be quite astonishing. On the basis of reducing L and increasing D to compress the original image, the smaller the gray-level L and the larger the step size D , the lower the cost of calculating the gray-level cooccurrence matrix, but the more information is lost, the less accurate the result will be. On the contrary, the larger the L , the smaller the D , and the more information retained, but the cost of calculating the gray-level cooccurrence matrix will increase. After determining L and D , the gray-level cooccurrence matrix of the image can be obtained and the 6 features of energy, contrast, entropy, median, local stability, and correlation of the matrix can be extracted. After calculating the above 6 features of the arachnoid CT image,

combined with the doctor's prediagnosis of the patient's subarachnoid hemorrhage (abbreviated as PD), a CT image mining database can be constructed. The characteristics of each case are organized and stored in the database in the form of a transaction, and each record is in the following format: PN, PD, H, I, J, K, L, M, Class, where PN represents the case number, PD represents the doctor's prediagnosis, H, I, J, K, L, and M are the features extracted from 6 gray-level cooccurrence matrices, respectively, and Class indicates whether the case is finally confirmed for subarachnoid hemorrhage. After the database is established, the association rule mining can be used to assist the diagnosis and classification. Here, the process of mining a database segment to determine the diagnostic criteria of subarachnoid hemorrhage is illustrated. Table 1 is a segment of the established database.

According to clinical medical experience and intelligent image diagnosis rules, the following diagnostic criteria are preliminarily assumed: If the doctor diagnoses in advance as 1, and at least one of the H, I, and J values of the arachnoid CT image is 1, and if at least two of the K, L, and M values are 1, the diagnosis is established, and the case can be diagnosed as subarachnoid hemorrhage, that is, the Class is 1. In the analysis, it was found that No. 81 in the database did not meet the initial diagnostic criteria, but the final case was still subarachnoid hemorrhage. This shows that the practicability and scientificity of the diagnostic criteria should be measured by mining association rules.

4. Experiment and Analysis

4.1. Data Collection. The third chapter has introduced the performance of subarachnoid hemorrhage in CT images, mainly including unsmooth edges, dark surface color, rough texture, uneven grayscale distribution, etc. For these performances, this article collected 50 groups of 1000 arachnoid CT images (two types of normal and abnormal) from the affiliated hospital of a certain top three medical hospital, and organized it in the image database for application research, the specific number of groups and the number of images in each group are shown in Table 2.

4.2. Comparison of NCFP-Growth with Other Algorithms

4.2.1. Algorithm Experiment on UCI Data Set. In order to test the performance of the NCFP-growth algorithm proposed in this paper, this paper selects the standard data in the data mining field to compare the performance of different algorithms. 8 data sets are in the UCI ML Repository: diabetes, glass, heart, hepatitis, Horse, iris, labor, led7. The five algorithms mentioned in Chapter 3 are used to obtain the comparative effects through comparative experiments, as shown in Figures 3 to 6.

It can be seen from the histogram that the NCFP-growth algorithm has a significant advantage in accuracy. This is because the NCFP-growth algorithm can prune unnecessary candidate sets when looking for frequent itemsets, thereby improving the accuracy of mining. The divide and conquer



FIGURE 2: CT image mining process.

TABLE 1: CT image feature data fragment.

PN	PD	H	I	J	K	L	M	Class
80	1	1	0	0	1	0	1	1
81	1	1	0	1	0	1	0	0
82	1	1	1	1	0	1	1	1
83	0	0	1	1	1	1	0	0
84	1	1	0	0	1	0	1	1
85	0	0	1	0	1	0	1	0
86	1	1	1	0	1	1	0	1
87	1	1	1	1	0	1	0	1
88	1	0	1	1	1	1	0	1
89	0	1	0	0	0	1	1	0

TABLE 2: Data image grouping and quantity.

Classification	Groups	Image number
Normal	40	800
Abnormal	10	200

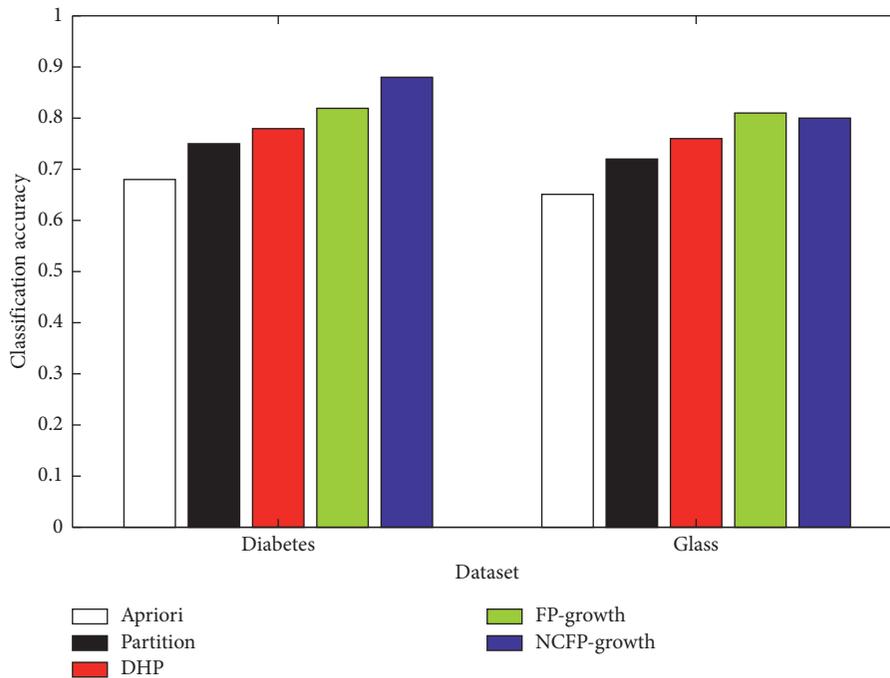


FIGURE 3: Classification accuracy of five algorithms in diabetes and glass.

idea of the Partition algorithm can significantly improve the performance of the algorithm.

4.2.2. Experiment of NCFP-Growth on the Subarachnoid Hemorrhage Dataset. 10-fold cross-validation is also used for the subarachnoid hemorrhage data set, and the

classification accuracy is compared with the several traditional classification algorithms introduced in Chapter 3, Apriori, DHP, Partition, and FP-growth. The experiment parameters are set as follows: min_sup is set to 1%, and min_conf is set to 50%.

On the subarachnoid hemorrhage data set, the experimental results in Table 3 show that compared with other

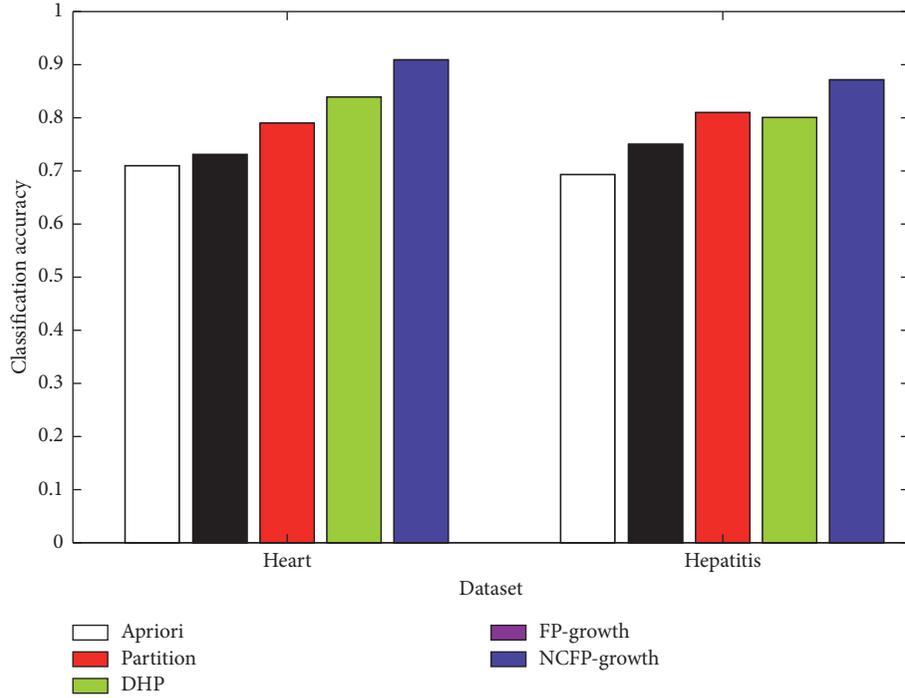


FIGURE 4: Classification accuracy of five algorithms in heart and hepatitis.

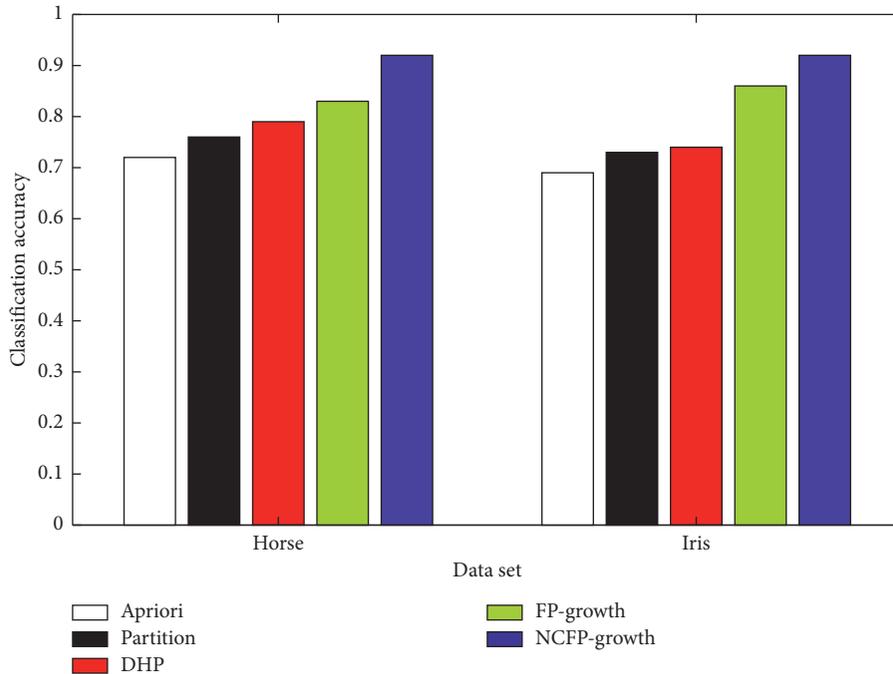


FIGURE 5: Classification accuracy of five algorithms in Horse and iris.

algorithms, NCFP-growth greatly reduces the number of candidate association rules, from 8756 in FP-growth to 3122 in NCFP-growth, and the classifier number of rules has been reduced from 63 in the FP-growth to 26 in the NCFP-growth. The experimental results in Figure 7 show that

compared with the traditional data mining classification algorithm introduced in Chapter 3, NCFP-growth has the highest classification accuracy rate of 95.2%, which meets the requirements of subarachnoid hemorrhage for the core classification algorithm.

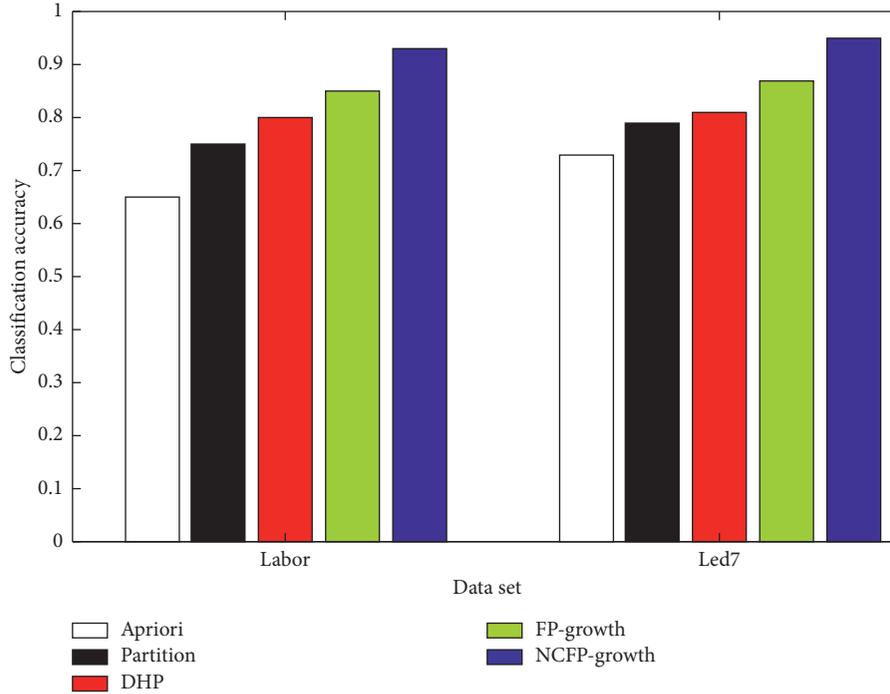


FIGURE 6: Classification accuracy of five algorithms in labor and led7.

TABLE 3: Number of FP-growth and NCFP-growth association rules.

Algorithm	Number of CARs	Number of classifiers
FP-growth	8756	63
NCFP-growth	3122	26

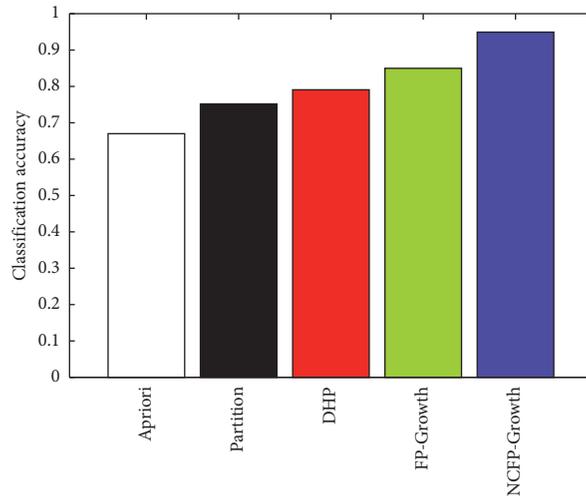


FIGURE 7: Classification accuracy of NCFP-growth and other algorithms.

5. Conclusion

This article has conducted an in-depth study on the application of association rules extraction in data mining in medical image diagnosis mining. Research shows that the technology in data mining can achieve good results in

medical image mining. The main work of this paper is as follows: (1) This paper firstly introduces the development of data mining, the research progress at home and abroad, and the application status of data mining in the medical field, etc., and introduces the main technologies and methods of data mining, combined with the characteristics of medical

data, proposed a process model suitable for medical data mining, and gave a detailed description of the process model. (2) This paper studies in detail the application of association rule extraction technology in data mining in the medical field. First, the theoretical basis and basic principles of the association rule extraction method are introduced. The Apriori algorithm for finding frequent itemsets in the association rule extraction and its series of improved algorithms are deeply studied. Finally, combining the characteristics of medical CT images, a grayscale-based algorithm is proposed. The association rule image mining method of the cooccurrence matrix is to calculate the gray cooccurrence matrix of the arachnoid CT image and obtain the texture features of the image. After the features are processed and organized, the association rules are mined. (3) Based on the FP-growth algorithm, the NCFP-growth algorithm based on association rules is proposed and compared with the mining effect of several other algorithms. Experiments on the subarachnoid hemorrhage data set show that the NCFP-growth algorithm has a higher accuracy rate in the CT diagnosis of subarachnoid hemorrhage, and can be used for actual case diagnosis.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] C. Z. Lü, *Neurology*, p. 154, Shanghai Science and Technology Press, Pudong, Shanghai, China, 2003.
- [2] Y. Q. Shi, *Practical Neurology*, pp. 649–653, Shanghai Science and Technology Press, Pudong, Shanghai, China, Second edition, 1994.
- [3] G. J. Rinkel, F. M. Wijdjcks, and M. Vermeulen, “Perimesencephalic nonaneurysmal subarachnoid hemorrhage: a follow-up study in 29 patients,” *Surgical Neurology*, vol. 57, no. 3, p. 160, 2002.
- [4] T. H. Schwartz and R. A. Solomon, “Perimesencephalic nonaneurysmal subarachnoid hemorrhage: review of the literature,” *Neurosurgery*, vol. 39, no. 3, pp. 433–434, 1996.
- [5] C. Spitzer, M. Mull, V. Rohde, and C. M. Kosinski, “Non-traumatic cortical subarachnoid haemorrhage: diagnostic work-up and aetiological background,” *Neuroradiology*, vol. 47, pp. 525–531, 2005.
- [6] S. Kumar, R. P. Goddeau, M. H. Selim et al., “Atraumatic convexal subarachnoid hemorrhage,” *Neurology*, vol. 74, no. 11, pp. 893–899, 2010.
- [7] R. Fogelholm, J. Hernesniemi, and M. Vapalahti, “Impact of early surgery on outcome after aneurysmal subarachnoid hemorrhage. A population-based study,” *Stroke*, vol. 24, no. 11, pp. 1649–1654, 1993.
- [8] Y. Kato, H. Sano, K. Katada et al., “Application of three-dimensional CT angiography (3D-CTA) to cerebral aneurysms,” *Surgical Neurology*, vol. 52, no. 2, pp. 113–122, 1999.
- [9] T. Inagawa and A. Hirano, “Ruptured Intracranial Aneurysms: An Autopsy Study of 133 patients,” *Surgical Neurology*, vol. 33, pp. 117–123, 1990.
- [10] J. B. Bederson, ES Jr. Connolly, H. H. Batjer et al., “Guidelines for the Management of Aneurysmal subarachnoid hemorrhage: a statement for healthcare professionals from a special writing group of the Stroke Council, American Heart Association,” *Stroke*, vol. 40, pp. 994–1025, 2009.
- [11] F. I. H. Witten, “Data mining,” *Practical Machine Learning Tools & Techniques with Java Implementations*, vol. 13, no. 1, p. 1, 2005.
- [12] K. J. Cios, B. Krawczyk, J. Cios, and K. J. Staley, “Uniqueness of Medical Data Mining: How the New Technologies and Data They Generate Are Transforming Medicine,” 2019, <https://arxiv.org/abs/1905.09203>.
- [13] J. Chen, K. Li, H. Rong, K. Bilal, N. Yang, and K. Li, “A disease diagnosis and treatment recommendation system based on big data mining and cloud computing,” *Information Sciences*, vol. 435, pp. 124–149, 2018.
- [14] P. Singh, S. Singh, and G. S. Pandi-Jain, “Effective heart disease prediction system using data mining techniques,” *International Journal of Nanomedicine*, T-NANO 2014 Abstracts, vol. 13, pp. 121–124, 2018.
- [15] R. Anxionnat, S. Bracard, X. Ducrocq et al., “Intracranial aneurysms: clinical value of 3D digital subtraction angiography in the therapeutic decision and endovascular treatment,” *Radiology*, vol. 218, no. 3, pp. 799–808, 2001.
- [16] H. J. Cloft, G. J. Joseph, and J. E. Dion, “Risk of cerebral angiography in patients with subarachnoid hemorrhage, cerebral aneurysm, and arteriovenous malformation,” *Stroke*, vol. 30, no. 2, pp. 317–320, 1999.
- [17] R. A. Willinsky, S. M. Taylor, K. TerBrugge, R. I. Farb, G. Tomlinson, and W. Montanera, “Neurologic complications of cerebral angiography: prospective analysis of 2,899 procedures and review of the literature,” *Radiology*, vol. 227, no. 2, pp. 522–528, 2003.
- [18] H. t. E. Westerlaan, A. M. van der Vliet, J. M. Hew, J. D. M. Metzemaekers, J. J. A. Mooij, and M. Oudkerk, “Magnetic resonance angiography in the selection of patients suitable for neurosurgical intervention of ruptured intracranial aneurysms,” *Neuroradiology*, vol. 46, no. 11, pp. 867–875, 2004.
- [19] C. Ordóñez and E. Omiecinski, “Discovering association rules based on image content,” in *Proceedings of the IEEE Advances in Digital Libraries*, IEEE Computer Society, Washington, DC, USA, March 1999.
- [20] M. Datcu and K. Seidel, “Image information mining: exploration of image content in large archives,” *IEEE Aerospace Conference Proceedings*, 2000.
- [21] J. Zhang, W. Hsu, and M. LiLee, “An information-driven framework for image mining,” in *Proceedings of the 12th International Conference on Database and Expert Systems applications (DEXA)*, pp. 232–242, Munich, Germany, September 2001.
- [22] M. C. Burl, C. Fowlkes, and J. Roden, “Mining for image content,” in *Systemics, Cybernetics, and Informatics/Information Systems: Analysis and Synthesis*, Orlando, FL, USA, 1999.
- [23] W. Hsu, M. L. Lee, and J. Zhang, “Image mining: Trends and developments,” *Journal of Intelligent Information Systems*, vol. 19, no. 1, pp. 7–23, 2002.
- [24] C. Glaser, R. Burgkart, A. Kutschera, KH. Englmeier, M. Reiser, and F. Eckstein, “Femoro-tibial cartilage metrics from coronal MR image data: technique, test-retest

- reproducibility, and findings in osteoarthritis,” *Magnetic Resonance in Medicine*, vol. 50, no. 6, pp. 1229–1236, 2010.
- [25] H. Laurent, “Wind extraction from meteosat water vapor channel image data,” *Journal of Applied Meteorology*, vol. 32, no. 6, pp. 1124–1133, 1993.
- [26] I. Dnvsls, “Count of equivalent pixel (cep): a novel algorithm to extract most relevant images from the medical image database,” *European Journal of Translational and Clinical Medicine*, vol. 7, no. 3, pp. 5047–5055, 2020.
- [27] H. Chen, S. Khan, B. Kou, S. Nazir, W. Liu, and A. Hussain, “A Smart Machine Learning Model for the Detection of Brain Hemorrhage Diagnosis Based Internet of Things in Smart Cities,” *Complexity*, vol. 2020, Article ID 3047869, 10 pages, 2020.
- [28] X. Wang, T. Shen, S. Yang et al., “A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans,” *NeuroImage: Clinica*, vol. 32, Article ID 102785, 2021.
- [29] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” vol. 1215, pp. 487–499, VLDB, September 1994.