Research Article

Key Technology of the Medical Image Wise Mining Method Based on the Meanshift Algorithm

Jinli Cui,1 Yadong Wang,2 and Ke Wang3

1Department of Radiology, Affiliated Hospital of Changchun University of Traditional Chinese Medicine, Changchun 130000, Jilin, China
2Department of Computer Science and Technology, Changchun University of Technology, Changchun 130000, Jilin, China
3Department of Radiology, Xi’an Kidney Diseases Hospital of Traditional Chinese Medicine, Xi’an 710300, Shaanxi, China

Correspondence should be addressed to Ke Wang; wangke1128520@njau.edu.cn

Received 2 April 2022; Revised 4 May 2022; Accepted 19 May 2022; Published 15 June 2022

Academic Editor: Hang Chen

Copyright © 2022 Jinli Cui et al. 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.

Mean-shift originally refers to the mean vector of the offset. The algorithm idea is to assume that the data sets of different clusters conform to different probability density distributions, and the area with high sample density corresponds to the center of the cluster. With the wide application of hospital information system, especially the popularity of the meanshift algorithm in the outpatient system, it has greatly improved the efficiency of medical staff. Medical imaging refers to the technology and process of obtaining internal tissue images of the human body or a certain part of the human body in a noninvasive manner for medical treatment or medical research. It contains the following two relatively independent research directions: medical imaging system and medical image processing. In this paper, we expect to improve the mining ability of medical image information with the help of the meanshift algorithm based on the key technology of the medical image intelligent mining algorithm. This paper proposes a method to enhance image feature extraction and data mining and how to apply relevant analysis rules for mining. Applying this integrated algorithm to extract simplified rules is more beneficial to people’s understanding than the raw data and helps doctors quickly understand the patient’s condition.

1. Introduction

Today’s society is an era of information and networking. The amount of information is growing rapidly. It also brings a lot of problems: the first is excessive information and difficult to digest; the second is true and false information. The fast-growing massive amounts of data are stored in large and large databases, without powerful tools, and understanding them has gone far beyond human capabilities, resulting in data being collected in large databases becoming “data graves” rare to visit again. Data archives and a lot of important information are hidden behind the surge of data. Decision makers’ decisions are often not based on useful information in the database, but on intuition, because decision makers lack the tools to extract valuable knowledge from massive data [1–3]. There are many types of data available in the medical field [4], including complete code information, clinical information about the patient’s medical history, testing, and treatment [5], drug management information [6], hospital management information, and [7], data mining theory is applied in medicine [8–10]. According to the amount of data, it is divided into memory level: the amount of data does not exceed the maximum memory of the cluster; BI level: the amount of data that is too large for memory; the amount of data that has completely failed or is too expensive for databases and BI products. The meanshift algorithm, also known as the mean shift method, has at its core the mean shift vector, which estimates the gradient of the probability density function. The meanshift algorithm needs to first calculate the offset mean of the current point, move the point to this offset mean, and then use this as a new starting point and continue to move until the final
conditions are met. The function does not require the aid of a priori knowledge and is classified as a nonparametric estimation method. The meanshift algorithm provides an effective means of analyzing medical data and extracting valuable and meaningful information. The study of genetic patterns of human disease and health is extremely important in promoting human health and maintaining a healthy quality of life [11, 12].

Electronic devices have been widely used in the medical field [13, 14]. However, most hospitals currently deal with databases as low-end operations in medical databases, lacking data integration and analysis, not to mention medical decision-making and automatic acquisition of knowledge. A large amount of data are described as "rich data, but information is poor." How to use data mining technology to find valuable knowledge and rules from these massive data is a problem that needs to be solved at present [15, 16], and mining the hidden rules in the data to use them to provide scientific decisions for the diagnosis and treatment of diseases, summarizing various treatments. The efficacy of the program, better for hospital decision-making management, medical, scientific research, and teaching services, has become a very important research topic. Rough sets are widely used, including a variety of logical, mathematical, and philosophical aspects of rough sets [17, 18]. Rough sets are also associated with many other methods and are also associated with a very broad hybrid system. Rough sets are now associated with data mining and knowledge discovery and decision systems. Rough sets have been used in reasoning and computational fields [19, 20].

This paper combines the meanshift algorithm with medical image intelligent mining technology, by mining association rules and useful information from a large number of images, and investigates the feature extraction and loading method of medical image data [21, 22], the classification method of medical image data, and the coarse set mining data aspect of brain tumor MRI images, which helps doctors identify the relationship between the high incidence of disease population, the severity of disease, and various hidden information, aids the decision-making diagnosis process, and improves the accuracy rate, with important theoretical significance and broad application prospects.

2. Proposed Method

2.1. Data Mining. Data mining is a cross-cutting discipline. Data mining is the process of getting valuable information of scrappy data. Data mining refers to the use of data mining technology on a very large amount of irregular data to obtain hidden information that is beneficial to value. The complexity of the data can thus be seen. Data mining can derive a new value from the complexity of information. Through data mining, new data rules can be found from the data and the data can be analyzed from different perspectives to obtain new values. Figure 1 shows a schematic diagram of the data mining model:

Data mining is a hot problem in the field of artificial intelligence and database research, which can save time and improve efficiency by quickly capturing the needed information from a large amount of complicated information according to the condition constraints. Data mining is a decision support process that can also make inductive reasoning and uncover potential information to help decision makers make the right decisions. Data mining technology has experienced four steps of data collection, data access, data warehouse, and data mining, which can meet the needs of query and storage and provide a basis for auxiliary decision-making.

Data mining not only can query and traverse past data but also can predict future trends and behaviors and automatically detect patterns that have not been discovered before so that people can make good decisions. The object of data mining can be any type of data source: it can be a relational database, which contains structured data; it can also be data warehouse, text, and multimedia data, which contains semistructured data or even heterogeneous data source. The discovered knowledge can be used for knowledge management, decision support, process control, and many other applications. The functions of data mining are mainly divided into the following categories: Association analysis: the correlation analysis refers to finding the similarity of values in the database, that is, the hidden association rules between the discovered data, which is the description of the data correlation in the database. Clustering: the process of grouping data in a database into multiple classes of similar data is called clustering, and its purpose is to establish a macroscopic concept of objective reality. Each class generated by clustering is a collection of data, the data in the same class are similar to each other, and the data in different classes are different. Figure 2 shows a diagram of the data mining process.

In recent years, data mining technology has been widely used in the medical field. Promising results have been obtained in disease diagnosis, treatment, and scientific research. In the pathological study, a large amount of data of pathological sections were analyzed by data mining and the key indicators were summarized to establish a normal and pathological virtual cell model. Virtual cell is an emerging discipline that analyzes, integrates, and applies the structure and function of cells through mathematical calculation and analysis to simulate and reproduce the phenomena of cells and life. This can be used to understand the physiological mechanisms of the occurrence, activity, and regulation of virtual cells. It can also understand and reveal the pathogenesis of diseases, find effective pathogenic molecules and marker molecules, conduct early warning diagnosis of diseases, and propose prevention and intervention measures. Data mining can make better use of these data, help physicians improve the efficiency and accuracy of diagnosis, reduce the work intensity of physicians, discover new medical laws, explore human mysteries, minimize medical risks, and improve the success rate of cure.

2.2. Meanshift Algorithm. The core of the meanshift algorithm is the mean shift vector, which is classified as a nonparametric probability density estimation, which
includes histogram estimation and kernel density estimation methods, among others.

\[ \bar{g}(a) = \frac{1}{m} \sum_{z=1}^{m} h(Q_z - a). \]  

The estimation algorithm is a histogram estimation algorithm, and \( g(a) \) represents the probability density function and \( h(a) \) represents the index.

\[ \tilde{g}(v) = \frac{1}{ti} \sum_{z=1}^{i} W(v - V_z). \]  

\( W \) represents the kernel function and \( V \) represents the sample set.

\[ \sup_{a \in I} |W(a)| < \infty, \]  

\[ \int_{-\infty}^{\infty} |W(a)| da < \infty, \]  

\[ \lim_{a \to \infty} a^n |W(a)| = 0, \]  

\[ \int_{-\infty}^{\infty} |W(a)| da = 1. \]  

As the nonreference functions are generally continuous single-peaked, they need to satisfy the aforementioned function expression conditions in the practical


2.3. Medical Images. Medical imaging refers to the technology and processing of the human body or a part of the human body to obtain images of internal tissues in a noninvasive manner for medical treatment or medical research. The former is the process of image generation; the latter is to recover images that are not clear enough, or to highlight certain features in the image, or to classify patterns in the image, etc. In addition to X-ray, there are also techniques such as MRI. Medical imaging has been developed so far. While techniques such as MRI focus on measurement and recording, but have no data display capabilities, the derived data can be considered an alternative form of sonography due to their locational characteristics. MRI applies a radio frequency pulse of a certain frequency to the human body in the static magnetic field, so that the hydrogen protons in the human body are excited and the magnetic resonance phenomenon occurs.

Medical images have many unique features compared to general images such as natural landscapes and art paintings. Among them, multimodality is one of them, which causes the multimodality of medical images because the imaging principles of modern medical imaging equipment are different. However, according to their application range, medical images are mainly divided into two types: anatomical images and functional images. Anatomical images have higher resolution and can more clearly reflect the structural information of a tissue or organ. Functional images can reflect information about the body’s metabolism but at a lower resolution. Specific to clinical applications, according to different imaging parameters and imaging conditions, each modality can also produce different performance images, such as T1 and T2 imaging modes in MRI. If you can combine this information, you can get more medical information for the doctor to diagnose, and image registration technology is the main tool to combine different modal image information. Compared with ordinary images, medical images have intrinsic nonuniformity and blur characteristics. Medical images have gradation in the gray scale. There is a large change in CT values of the same tissue. For example, the density of the femur, sinus bones, and teeth in the bones is very different. The CT values are not uniform for the same object, such as the density of the outer surface of the femur and the internal bone marrow. In addition, the noise signal caused by technical reasons tends to blur the high-frequency signal at the edge of the object and the ambiguous effect of the image due to the conscious or unconscious activity of the human body. In a voxel on a boundary, the relationship between the edge, the corner, and the region of the object in the image of both the boundary and the object is often difficult to accurately describe. Some of the diseased tissue cannot be clearly defined due to its infiltration into the surrounding tissue.

2.4. Image Enhancement. Image enhancement is to purposefully emphasize the overall or local characteristics of the image, make the original unclear image clear, expand the difference between the characteristics of different objects in the image, and improve the visual effect of the image. Traditional contrast enhancement algorithms are divided into direct and indirect methods. The direct method mainly achieves the purpose of enhancement by correcting the histogram, and the overhead calculation first borders the contrast and then enhances the contrast based on this. Image enhancements include a wide range of content; for example, it can enhance comparison, improve noise, reduce shadow, refine, and filter margins. Traditional contrast enhancement techniques are mostly based on global and neighborhood, and the results of these methods tend to fall under under-enhancement or overenhancement. In the global histogram equalization, since a small number of adjacent gradations are combined into one gradation during the equalization process, the contrast is lowered. Through histogram equalization, the image brightness can be better distributed on the histogram, which can be used to enhance the local contrast without affecting the overall contrast. In order to overcome this defect, the equalization function can be obtained in a small area. In the equalization process, the combination of adjacent pixels is reduced and the image contrast is less reduced than the global equalization. It is called adaptive neighborhood histogram equalization. There are many factors that cause low image quality: uneven lighting will make the grayscale image too concentrated, transmission lines are generated when the noise pollution and image quality inevitably reduced; the lighter the noise, the more difficult it is to see the details of the image; and then, the image will be blurred. Therefore, before the image analysis and processing, the image must be improved, that is, the image must be enhanced. Table 2 shows the evaluation of the image enhancement effect.

Using image gray histogram processing technology, there are many ways to enhance the image in the spatial domain, such as enhanced contrast and dynamic range compression. The grayscale histogram is to arrange all the pixels in the digital image according to the size of the grayscale value. It represents the number of pixels with a certain gray scale in the image and reflects the frequency of a certain grayscale in the image. This kind of processing method is more flexible and convenient, and the processing effect is also good, but for some images with dense gray distribution or weak contrast. Although it has a certain enhancement effect, it is not obvious to identify. The gradation transformation processing of an image is realized by changing the probability distribution of each pixel of the original image at each gray level. Table 3 shows the image enhancement metrics for the different regions.
The histogram function expressions can be obtained by counting the images.

\[ p(sk) = \frac{nk}{n}, \quad k = 0, 1, \ldots, L - 1. \] (7)

This expression is expressed in the first \( k \), the number of pixels at the gray level is \( nk \), total number of pixels is \( n \), ratio is \( p(sk) \), and the right \( sk \) should be given, which is an estimate of the probability of occurrence. Through this function, you can clearly understand the dynamic range of the image and you can understand the main concentration range of the grayscale image.

The purpose of enhancing an image is to sharpen the original image to emphasize certain features of the image, making it easier to improve the visual effect on the picture or perform other processing. The histogram equalization enhancement method is a grayscale enhancement algorithm performed in the spatial domain.

2.5. Feature Extraction. A single measurement is defined in eigen extraction treatment. Feature extraction is a major action in picture processing, which examines pixels. The input image, which is a prerequisite for feature extraction action, has usually been processed before the final result is produced. Figure 3 shows the block diagram of an image enhancement algorithm.

The first step in SIFT feature extraction is to design the space, the second step is to target the image focus, and the third step is to perform orientation assignment. The fourth step is to describe the key points. In order to ensure the stability of the image when describing the key points, the angle is adjusted in the area near the key points beforehand and then the histogram of the sampled area is calculated. Finally, a vector technique is applied to the image.

2.6. Rough Set. Rough set theory is a powerful data analysis tool that does not require prior knowledge in the application. Rough set theory has some unique viewpoints, which makes it irreplaceable in many fields and makes it particularly suitable for data analysis. It does not need any prior knowledge beyond the set of data the problem needs to deal with and is highly complementary to theories dealing with other uncertainty problems. Different attribute knowledge descriptions can be used to generate different categories. It divides the knowledge of the research domain by the indistinguishable relationship and forms the knowledge expression system to use above and below approximation sets to approximate the description object through knowledge reduction so as to obtain the simplest knowledge. A main goal of rough set theory is to obtain the conceptualization of the target through background knowledge. Its more prominent feature is its objectivity. It does not need to prescribe certain features or quantitative descriptions of attributes, such as the membership of fuzzy sets, degree, and the probability distribution of statistics, but directly from the description set of a given problem, it finds the inherent law in the problem through the indiscernible relationship and the approximate domain. Table 4 shows examples of rough set decision-making.

Rough theory differs from theory as it does not require a priori information, so rough theory is a more objective description of the uncertainty of the problem. Rough theory extends the old set theory by embedding the knowledge used for classification into the set as a collective component. Rough theories are thus divided into three types: (1) object a must belong to set \( X \); (2) object a must not belong to set \( X \); and (3) object a may or may not belong to set \( X \).

3. Experiments

Tumors that grow in the brain are collectively referred to as brain tumors, including primary tumors caused by changes in the brain’s own parenchyma and secondary tumors that
occur from other parts of the body and metastasize to the brain. It can occur at any age, and the cause is still unknown. The correct diagnosis of brain tumor grade has a great influence on the treatment and prognosis. Generally, low-grade brain tumor (LGG) has a survival time of 5 to 10 years, while high-grade brain tumor (HGG) has a survival time of about 1 year. The grading examination and diagnosis of brain tumors currently rely heavily on the analysis of medical images such as CT, angiography, and MRI. Among them, MRI examination is generally considered to be more sensitive than CT in the medical field and has three-dimensional imaging characteristics, which can reflect the brain more effectively and the essence of the tumor. At present, MRI diagnosis of brain tumors is mainly based on tumor location, tumor morphology, tumor occupying, tumor edge clarity, edema around the tumor, bleeding, calcification, and many other factors. In this experiment, MRI images of 120 pathologically confirmed brain tumors in a hospital were collected. The images were collected from the Maconi 1.5 T superconducting nuclear magnetic resonance system and organized in an image database, as shown in Figure 4 (image comes from the network) and collected according to the WHO classification. The classification of image data is shown in Table 5.

### 4. Discussion

The acquired MRI image has a large original size and needs to be compressed. The image format directly collected from the medical instrument is 512 × 512, which is compressed into 128 × 128 format, so as to improve the analysis efficiency. In Clementine, you can directly input the MRI image using the image database as the data source and perform cluster analysis based on the different settings of the edges, occupancy, bleeding, edema, and other factors of the MRI image. The data flowchart is shown in Figure 5; through the user input node, input source data after preprocessing, and K-mean cluster analysis.

The user input node is used to input the original data, and the features of the image are input into the data stream from the user input node in the form of a table. The middle type node, the derive node, and the SPSS transform node are feature inputs for processing the MRI image. Similarly, the characteristics of the MRI image are entered in a discrete type indicating discrete data and a flag type indicating Boolean data. Finally, the data are analyzed by the K-mean node.

Figure 6 shows the analysis results, according to the edge, hemorrhage, calcification, edema, and occupancy of a total of five MRI signs into a total of 9 categories, cluster1 to cluster9. In order to facilitate the subsequent data mining rough set, it is also required to carry out its treatment before discrete excavation, the attribute value range into a plurality of breakpoints through a number of discrete sections, and then different symbols representing each attribute value of the interval. After that, a rough set analysis data stream is established and the data are input into the input stream for rule mining. The rules explained in the excavation are shown in Table 6.

| WHO level | Percent |
|-----------|---------|
| Class I   | 14%     |
| Class II  | 42%     |
| Class III | 33%     |
| Class IV  | 11%     |
Table rules and inference rules presubstantially are the same as the remaining samples of the resulting test sample were evaluated rule, and the diagnostic rules using the WHO level above an 80% accuracy rate.

5. Conclusions

With the continuous development in health care, health care is gaining more and more attention. In this system, a wide variety of patient information can be stored. If this information could be reused, it would greatly facilitate the development of medical care. But this data simply record the surface information and do not analyze the deeper content of the information to its fullest extent.

In this paper, we focus on the meanshift algorithm, which provides an in-depth understanding of the vectors that are at the heart of the algorithm and provides an optimal optimisation algorithm for data analysis. Combining the meanshift algorithm with medical data will be of great use in disease diagnosis and treatment, medical research and teaching, and hospital management, greatly contributing to the advancement of healthcare.

Data Availability

This article does not cover data research. No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Z. He, B. Fan, T. C. E. Cheng, S.-Y. Wang, and C.-H. Tan, “A mean-shift algorithm for large-scale planar maximal covering location problems,” European Journal of Operational Research, vol. 250, no. 1, pp. 65–76, 2016.

[2] O. Hyrien and A. Baran, “Fast nonparametric density-based clustering of large datasets using a stochastic approximation mean-shift algorithm,” Journal of Computational & Graphical Statistics, vol. 25, no. 3, pp. 899–916, 2016.

[3] H. Abulkasim, A. Farouk, S. Hamad, A. Mashatan, and S. Ghose, “Secure dynamic multiparty quantum private comparison,” Scientific Reports, vol. 9, no. 1, pp. 17818–17916, 2019.

[4] W. Y. Zhu, “Video target tracking algorithm with combining robust estimation and Meanshift,” Journal of She-nyang University of Technology, vol. 39, no. 2, pp. 177–182, 2017.

[5] G. Beck, T. Duong, M. Lebbah, H. Azzag, and C. Cérin, “A distributed approximate nearest neighbors algorithm for efficient large scale mean shift clustering,” Journal of Parallel and Distributed Computing, vol. 134, no. Dec, pp. 128–139, 2019.

[6] O. I. Khalaf, G. M. Abdulsahib, and B. M. Sabbar, “Optimization of wireless sensor network coverage using the bee algorithm,” Journal of Information Science and Engineering, vol. 36, no. 2, pp. 377–386, 2020.

[7] T.-Y. Kim, S.-H. Kim, and H. Ko, “Design and implementation of BCI-based intelligent upper limb rehabilitation robot system,” ACM Transactions on Internet Technology, vol. 21, no. 3, 2021.

[8] W. Jianhong, Q. Welliang, and W. Xiaogang, “Convergence and stability analysis of mean-shift algorithm on large data sets,” Statistics and Its Interface, vol. 9, no. 2, pp. 159–170, 2016.
[9] D. Tan, W. Zhong, C. Jiang, X. Peng, and W. He, "High-order fuzzy clustering algorithm based on multikernel mean shift," Neurocomputing, vol. 385, no. 8, pp. 63–79, 2020.

[10] M. Elhoseny, G. Ramirez-Gonzalez, O. M. Abu-Elnasr, S. A. Shawkat, N. Arunkumar, and A. Farouk, "Secure medical data transmission model for IoT-based healthcare systems," IEEE Access, vol. 6, pp. 20596–20608, 2018.

[11] J. L. Zhang, P. Shi, and X. B. Wen, "A TLD tracking algorithm based on scale-adaptive mean-shift method," Kongzhi yu Juece/Control and Decision, vol. 34, no. 1, pp. 144–150, 2019.

[12] Z. Lv and H. Ko, "Introduction to the special issue on recent trends in medical data security for e-health applications," ACM Transactions on Multimedia Computing Communications and Applications, vol. 17, 2021.

[13] J. W. Yoon, "Bayesian interpretation to generalize adaptive mean shift algorithm," Journal of Intelligent and Fuzzy Systems, vol. 30, no. 6, pp. 3583–3592, 2016.

[14] A. Farouk, A. Alahmadi, S. Ghose, and A. Mashatan, "Blockchain platform for industrial healthcare: vision and future opportunities," Computer Communications, vol. 154, pp. 223–235, 2020.

[15] R. Niranjana, V. A. Kumar, and S. Sheen, "Darknet traffic analysis and classification using numerical AGM and mean shift clustering algorithm," SN Computer Science, vol. 1, no. 1, pp. 1–10, 2020.

[16] M. Adil, H. Song, J. Ali et al., "Enhanced AODV: a robust three phase priority-based traffic load balancing scheme for internet of things," IEEE Internet of Things Journal, 2021.

[17] C. Yi, W. Xiao-Hong, and X. Chun-Bo, "Positioning method of unmanned aerial vehicles based on Meanshift and camera calibration," Journal of Tianjin Polytechnic University, vol. 36, no. 1, pp. 55–60, 2017.

[18] S. Zhang, Y. Wang, P. Wan, J. Zhuang, Y. Zhang, and Y. Li, "Clustering algorithm based data fusion scheme for robust cooperative spectrum sensing," IEEE Access, vol. 8, 2020.

[19] D. Manea, "Hyperspectral imaging in the medical field: present and future," Applied Spectroscopy Reviews, vol. 49, no. 6, pp. 435–447, 2014.

[20] A. D. Kissling and K. D. Ballinger, "Implementation of a research information management system in a pediatric hospital," Medical Reference Services Quarterly, vol. 37, no. 2, pp. 184–197, 2018.

[21] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," Neuron, vol. 95, no. 2, pp. 245–258, 2017.

[22] B. Yang, "Research on customer value evaluation based on data mining and mathematical statistics," Advanced Materials Research, vol. 926-930, pp. 3890–3893, 2014.