Ilkogretim Online - Elementary Education Online, 2020; Vol 19 (Issue 1): pp. 1315-1318
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2020.661953

ANALYSIS OF DATA MINING TECHNIQUE FOR HEALTHCARE DATA CLASSIFICATION

Abhay Shukla, Shail Dubey, Rituraj Kushwaha, Ashish Shukla, Pooja Diwivedi

Department of Computer Applications, Axis Institute of Higher Education, Kanpur, Uttar Pradesh, India

ABSTRACT- The Scientific Records Class is pivotal in a wide range of medical imaging applications, either by streamlining or automating the process of evaluating image statistics. It addresses the difficulties in diagnosing, analyzing, and training within the medical sector. As a result, a variety of statistical analysis methods and software, all derived from statistical techniques, have been created and refined. This research introduces a comparative study of statistical analysis software. Hence, the existing literature suggests that the role of statistical analysis software in classifying medical data, which remains largely untapped and holds potential, should not be underestimated. Therefore, there's a significant potential to utilize statistical analysis techniques in the classification of medical data, an area that has been scarcely explored and stands as a promising future area for research.

Keywords: medical records class, statistics mining, neural networks, texture class.

I. INTRODUCTION

In today's age of creating facts, accessing medical records has become easily available through different methods like X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound, among others. As a result, there has been a noticeable rise in the amount of medical data stored, and the variety of databases needed to handle these digital records has grown dramatically (Mitra et al., 2002). The first steps into the computerized evaluation of medical images were taken in the 1960s, followed by in-depth and thorough studies on using computers to assist in diagnosing illnesses in the 1980s, which marked a major change in how computers are used in healthcare, moving from computer-assisted analysis to computer-assisted diagnosis (Doi, 2007).

Medical scans are essential for diagnosing and treating illnesses and are equally important for learning in healthcare. With the huge quantity of medical information out there, there's an urgent need to create new tools that can manage and access important data for doctors and research groups efficiently and effectively. This has resulted in the growth and progress of records mining, a technique that involves extracting meaningful, previously unknown, and possibly useful information from big data collections. Thus, records mining is the intricate method of extracting important information from medical records.

The rest of this document is organized as phase 2, which focuses on exploring strategies for data mining. Phase 3 will showcase a comparative examination of these data mining strategies. The conversation in phase 4 will be enhanced with case studies and illustrations, leading to phase 5, which will offer a detailed conclusion.

II. DATA MINING TECHNIQUES

Data mining, a field of study that began in 1980, aims to extract knowledge hidden in data. This field focuses on extracting insights from various records for organizational use. It involves the use of sophisticated statistical tools to find patterns in extensive data collections. Dealing with medical data, for example, poses significant challenges because of its complexity. The methods used in data mining are mainly rooted in statistical and computational understanding, utilizing algorithms. These methods can generally be grouped into two types: parametric and non-parametric models. Parametric models depend on established statistical algorithms to establish a mathematical connection between input and output

variables, assuming the data's structure is understood. On the other hand, non-parametric models do not need the data's structure to be known; instead, they use statistical techniques to establish this connection.

In the field of data mining, algorithms vary in their approach, ranging from those that lean more towards statistical models to those that focus on more flexible, non-parametric approaches. For instance, when dealing with health data, the focus is increasingly on non-parametric models because they are more effective in extracting valuable insights from information like medical images. This encompasses various activities such as segmenting images (Zaidi & El Naqa, 2010), creating records for images (Kharrat et al., 2010; Tu, Shin & Shin, 2009), assisting in the diagnosis using computer systems (Hadjiiski et al., 2004; Doi, 2007), annotating images (Russell et al., 2008), and more. Traditionally, there hasn't been much research into applying data mining methods to medical images. Nonetheless, there has been a growing trend, with methods like nearest neighbor-based approaches initially being the predominant and efficient techniques. More recently, support vector machines (Kharrat et al., 2010; Aarthi et al., 2011) have become increasingly favored and superior to these earlier techniques.

In the same way, as seen in the area of selecting features, merging different classifiers is a well-known method for improving results. This strategy has proven especially useful when dealing with data from medical images. The advancement of data mining methods, moving from simple nearest neighbor approaches to the implementation of support vector machines, indicates a wide and powerful use of these techniques in identifying and distinguishing different types of irregularities in medical images collected from a variety of tests using different imaging techniques.

III. COMPARATIVE ANALYSIS OF DATA MINING TECHNIQUES

The specified section offers a range of research papers on the classification of medical images, which are outlined in table 1. It is clear that each classification method produces impressive outcomes, but it appears that no single method is adaptable enough to achieve broad recognition among researchers in the field of medical image classification. Suguna & Thanushkodi (2010) explored the development of a k-NN system enhanced with a genetic algorithm to simplify computations and lessen reliance on the training dataset, leading to improved results. Additionally, a method for classifying MRI of the brain using a genetic algorithm with a Support Vector Machine (SVM) was introduced, accurately categorizing brain tissue into either benign, malignant, or benign categories. Yet, SVM performs more effectively with features that have multiple dimensions and are continuous, necessitating a substantial dataset to ensure the highest level of prediction precision (Kharrat et al., 2010). The use of decision tree-based classification techniques is common in data mining for providing support in decision-making for medical practitioners. Moreover, the analysis of image texture is feasible across all digital image formats, offering insight into an image's spectral properties through various analytical approaches (Kassner & Thornhill, 2010).

IV. DISCUSSION

Over the past several decades, there's been a significant advancement in scientific data classification across three main areas (1) the development and application of complex classification algorithms, (2) the utilization of specific capabilities, and (3) the integration of related statistical data with classification algorithms. Nonetheless, several challenging problems persist, including but not limited to, the methodology of analyzing medical records, interactions with patients, overall efficiency, and the ability to scale these processes; other problems involve understanding the practical applications of data mining in medical settings and its societal impacts. As per Table1 and Table2, each imaging technique presents its unique challenges for classification. Despite these advancements, there remains no universally adopted approach to categorizing medical data. This is because the medical field demands a high level of accuracy, particularly in reducing false negatives. While general approaches do exist and can be applied to various datasets, specialized approaches that cater to specific needs are often more efficient by leveraging previous experiences. Choosing the appropriate approach to a classification problem can, therefore, be quite challenging. Nevertheless, there is a considerable opportunity for further improvement in contemporary scientific data classification tasks. Thus, there's a significant potential for the application of data mining techniques to medical data, which has not yet been fully explored and could represent an exciting avenue for future research.

V. CONCLUSION

The document analyzes contemporary methods, challenges, and potential advancements within the area of categorizing scientific records. It concentrates on detailing the primary sophisticated classification methods and approaches aimed at improving the precision of categories. Researchers have demonstrated an increasing level of curiosity and have dedicated resources to investigate promising applications for data mining in this context. A significant amount of research literature exists on the subject of classifying scientific data.

After examining this extensive collection of studies, it becomes clear that the methods of data mining have been successful for categorization purposes. This document presents an analysis that compares the latest progress in the field of categorizing medical data. It provides new insights into the methods described in previous research, focusing on the classification approaches used. Medical data typically includes a combination of numbers and text, which can be scattered and repetitive. As a result, it's essential to have algorithms that are not only efficient but also strong and adaptable.

The research on categorizing medical data indicates that there is still a lot of potential for improvement in the current categorization of tasks. Nonetheless, additional studies are required to further investigate and minimize uncertainties related to the classification of medical data, aiming to enhance the accuracy of the process.

REFERENCES

[1]. A Antonie, M.L., Zaiane, O.R., & Coman, A. (2001). Application of datamining techniques for medical image classification/KDD, 94-101. Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medicalimagingand graphics: the official journal of the Computerized Medical Imaging Society, 31(4-5), 198.

[2]. Fesharaki, N. J., & Pourghassem, H. (2012). Medical X-ray Images Classification Based on Shape FeaturesandBayesianRule.In computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on (pp. 369-373). IEEE.

[3]. Hadjiiski, L., Chan, H.P., Sahiner, B., Helvie, M. A., Roubidoux, M. A., Blane, C. & Shen, J. (2004). Improvement in Radiologists„ Characterization of Malignant and Benign Breast Masses on Serial Mammograms with Computer- aided Diagnosis: A ROC Study1. Radiology, 233(1), 255-265.

[4]. Hadidi, M. R. A. A., Gawagzeh, M. Y.& Alsaaidah, B. (2012). Solving mammography problemsof breastcancerdetectionusingartificial neural networks and image processing techniques. Indian journal of science and technology. Volume5, No.4. pp.2520-2528.

[5]. Han, J., Kamber, M. & Pei, J. (2011). Data Mining: Concepts and Techniques, Third Edition. Morgan Kaufmann.

[6]. Kaur, H., & Wasan, S. K. (2006). Empirical studyon applications of data mining techniques in healthcare. Journal of Computer Science,2(2),194-200.

[7]. Kotsiantis, S. B. Supervised machine learning: A reviewof classificationtechniques, Informatica. 2007, Volume (31), pp. 249-268.

[8]. Kassner, A., Thornhill, R.E. (2010). Texture Analysis: A Review of Neurologic MR Imaging Applications, A Journal of Neuroradiology31:809

[9]. Kharya, S. (2012). Using data mining techniquesfor diagnosis and prognosis of cancer disease. International Journal of computer science and information technology (IJCSIT), Vol. 2, No.2, pp55-66.

[10]. Kharat, K.D., Kulkarni, P.& Nagori, M.B. (2012). Brain tumor classification using neural network-based methods. International journal of computer sciences and informatics, ISSN: 2231- 5292-, Vol-1, issue 4, pp 85-90.

[11]. Kharrat, A., Gasmi, K. Messaoud, M. B., Benamrane, N. &Abid, M. (2010). A hybrid approachfor automaticclassificationofbrainMRI using genetic algorithm and support vector machine", Leonardo Journal of science's-1582-0233, pp. 71-82.

[12]. Kumar, G. J., & Kumar, G. V. (2008). Biological Early Brain Cancer Detection Using Artificial Neural Networks. In Artificial Intelligence and Pattern Recognition (pp. 89-93).

[13]. Latifoglu, F., Polat, K., Kara, S., Gunes,S. (2008). Medical diagnosis of a atherosclerosis from carotidartery Doppler signals using principal component analysis (PCA), k-NN based weighting preprocessing and Artificial Immune Recognition System (AIRS). J. Biomed. Inform.,41,15–23
[14]. Mitra, S., Pal, S.K., & Mitra, P. (2002). Data mining in soft computing framework: Asurvey. IEEE transactions on neural networks,13(1),3-14. Miller, A.S., &Blott, B. H. (1992). Review of neural network applications in medical imagingand
[15]. signal processing. Medical and Biological Engineeringand Computing, 30(5),449-464
[16]. Russell, B.C., Torralba, A., Murphy, K.P., & Freeman, W. T. (2008). LabelMe: a database and web based tool for image annotation. International journal of computer vision, 77(1-3),157-173.
[17]. Rajini N.H. & Bhavani, R. (2011). Classification of MRI Brain Images using k- Nearest Neighbor and Artificial Neural Network, IEEE-International Conference on Recent Trends in Information Technology, ICRTIT.
[18]. Rajendran, P., Madheswaran, M., &Naganandhini, K. (2010). An improved pre- processing technique with image mining approach for the medical image classification. In Computing Communication and Networking Technologies (ICCCNT), 2010 International Conference on (pp. 1-7). IEEE.
[19]. Sarhan, A. M. (2009). Cancer classification based on micro array gene expression data using DCT and ANN. Journal of theoretical and applied informationtechnology, 6(2),208-216.
[20]. Smitha, P., Shaji, L., Mini, M.G. (2011). A review of medical image classification technique, International confernceon VLSI, Communication & Instrumentation
[21]. Suguna, N. &Thanushkodi, K. (2010). An improved k-nearest neighbor classification using genetic algorithm. International journal of computer science issues (IJCSI), Volume 7, issue4, No2.
[22]. Tu, M. C., Shin, D., & Shin, D. (2009). A comparative study of medical data classification Methods based on decision tree and bagging algorithms. In Dependable, Autonomic and Secure Computing, 2009.DASC'09. Eighth IEEE International Conference on (pp. 183-187). IEEE.
[23]. Perner, P. (2002). Image mining: issues, framework, a generic tool and its application to medical-image diagnosis. Engineering Applications ofArtificialIntelligence,15(2), 205-216