PARASCRIPT®

White Paper

Using “Voting” in Computer Aided Detection for Mammography to Lower False-Positive Rates

Takeaway

- One of the biggest problems facing developers of CAD systems, as well as radiologists has been the high number of false-positive CAD marks on the reviewed mammograms.
- Only when CAD systems achieve sensitivity and false positive rates similar to a radiologist, they will be fully embraced as a second opinion tool.
- The integration of several complementary algorithms and sophisticated voting methods within one CAD system can enable the next generation CAD systems to achieve high sensitivity rates and low false-positive marks for greater levels of certainty.

Summary

The use of voting methodology in medical imaging is expected to eventually improve the false positive rates of current CAD systems by 80%, bringing false positive rates close to those of a general radiologist in the 10 – 20% range. This reduction of false-positive readings will ultimately result in decreased medical costs, emotional stress, follow-up examinations, and recalls.

What’s in this Paper?

6273 Monarch Park Place, Longmont, CO 80503 USA
T: 303.381.3100 | toll free: 888.225.0169
F: 303.381.3101 | info@parascript.com

Introduction

Breast cancer is taking women away in the prime of their lives. Excluding cancers of the skin, breast cancer is the most common type of cancer in women in the United States, accounting for 1 of every 3 cancers diagnosed. The chance of developing invasive breast cancer at some time in a woman's life is about 1 in 7 (13.4%). The National Cancer Institute estimates that there will be 226,870 new cases detected and 39,510 women will die of breast cancer in 2012[1]. While in the past approximately only 75% of women diagnosed with breast cancer survived 5 years after detection, today nearly 90% of women diagnosed with breast cancer will survive their disease 5 years after detected. This increase in survival rates is largely attributed to routine mammography screening and other advanced treatment methods.

Today nearly 90% of women diagnosed with breast cancer will survive their disease 5 years after detected. This increase in survival rates is largely attributed to routine mammography screening and other advanced treatment methods.

Advances in Screening Mammography

Quality mammography screening can help detect and evaluate breast abnormalities, both in women who have no complaints or symptoms and in women who have warning signs. Often it can detect breast cancer in its earliest, most treatable stages – when it is too small to be felt or identified by any other method. Finding small breast cancers early by a screening mammogram greatly improves a woman's chance for successful treatment.

While mammograms detect most breast cancers, a small percentage is still missed. Also, sometimes signs on a mammogram that appear abnormal, and may require a biopsy, will turn out not to be breast cancer (a positive reading). A positive reading refers to a suspicious finding that indicates possible cancer: a major area of concern for radiologists and physicians. Patients with a positive reading undergo a biopsy to determine if the suspicious area is cancerous or benign. If the biopsy results show it is non cancerous, it is considered a false-positive reading.

For every 1,000 women that have a screening mammogram, approxi-mately 100 of them will get called back, about 10 to 20 of them will need a biopsy, and only four to six will have breast cancer.

The average statistics in this area are very well known. For every 1,000 women that have a screening mammogram, approximately 100 of them will get called back, about 10 to 20 of them will need a biopsy, and only four to six will have breast cancer.

However, the variability of false positive results among radiologists interpreting a test set of mammograms varies widely. The study on "Screening Mammograms by Community Radiologist" published in the Journal of the National Cancer Institute[2] reported false-positive rates ranging from 2.6% to 15.9%. This variability in readings by individual radiologists can be both alarming and expensive for the patient. The study made a conclusion that the ultimate goal is to enhance mammography performance by reducing the rate of false-positive interpretations while maintaining high levels of sensitivity and accuracy.

False-positive readings result in increased medical costs, undue patient anxiety and distress, increased time demands on the medical system, pain and anguish. Additional medical costs are incurred due to follow up examinations like repeated biopsies; recalls for screening examinations are required before the patient can be declared cancer free. The extra costs, monetary and emotional, of false-positive mammograms, especially in women under 50, are a neglected but substantial problem. The emotional stress due to false-positives on women, until now considered a soft issue, is now gaining prominence.

Computer Aided Detection and False Positive Rates

Computer-aided detection (CAD) involves the use of software and computers to bring suspicious areas on a mammogram to the radiologist's attention. It is employed after the radiologist has done the initial review of the mammogram.

The image is displayed on a video monitor, with suspicious areas highlighted for the radiologist to review. The radiologist can compare the digital image with the digitized conventional mammogram or previous digital images to see if any of the highlighted areas were missed on the initial review and require further evaluation. It is the radiologist who decides whether true areas of concern are present at the highlighted locations before making the final diagnosis.

According to the American Cancer Society, early research results suggest that CAD systems help radiologists diagnose more early-stage cancers than mammograms alone. But some doctors disagree about how many cancers CAD software can identify correctly. Some of them feel that the CAD device is not as effective as simply having a second radiologist review the films or digital images, primarily due to false identification of benign changes as being suspicious for cancer. These fears also roused in April 2007 by a New England Journal of Medicine article by Fenton at al. alleging that CAD had, at best, a neutral impact on detection of lesions, and at worst, was missing lesions that otherwise might be spotted. One of the biggest problems facing developers of CAD systems, as well as radiologists, has been the high number of false-positive CAD marks on the reviewed screening mammograms.

According to the American Cancer Society, early research results suggest that CAD systems help radiologists diagnose more early-stage cancers than mammograms alone.

CAD systems currently available on the market have false-positive rates that are approximately 10 times higher than those of radiologists. This only aggravates the existing problem because the occurrence of a large number of false-positive marks by a CAD system can significantly hinder its usefulness by distracting the interpreting radiologist. A significant reduction in false positive rates is required in order for a CAD system to become a “true friend” of any radiologist. It is only when CAD systems achieve sensitivity and false positive rates similar to a radiologist that they will be fully embraced as a second opinion tool.

What Can be Done?

New advancements in image analysis and recognition technology can potentially enable CAD for mammography to be truly used as a second opinion with false-positive rates approaching those of a second radiologist.

Authors of this white paper suggest that the integration of several complementary algorithms within one CAD system or the utilization of two CAD systems and sophisticated voting methods can enable the next generation CAD systems to achieve high sensitivity and low false-positive rates for greater levels of certainty.

The Voting Mechanism Approach

Because orthogonal engines (based on different methodologies) make different errors, it is possible to reduce the error rate by combining the individual results of multiple engines using a voting mechanism. The voting mechanism produces a combined output, which is more reliable than any of the individual results and can improve the overall performance remarkably. However to be efficient, voting schemes might be very sophisticated. They have to be able to normalize results, focus on the individual engine’s strengths, avoid their weaknesses, and suppress unimportant results. In some cases voting eliminates results of an engine or algorithm; in others it combines them with the results of the others.

The entire process of applying voting methodology during image processing and analysis can be described as the work of a group of highly skilled experts. Each of them has unique skills, expertise and uses favorite approaches, which are especially efficient in some cases and produce good enough results in the others. When they work together as a team, their areas of expertise complement each other resulting in improved overall performance. Similarly, all algorithms apply various methods to analyze dozens of features of an image. Additionally, each of them has a special area of expertise and exploits a unique method, which may be viewed as a distinctive characteristic of a particular algorithm. Sophisticated logic analyzes and combines the interim results obtained from the individual engines to determine the final result and the corresponding confidence value.

First Applications of Voting Methodology in Postal and Payment Automation Industries

Implementation of a powerful combination of engines using a number of fundamentally different algorithms and techniques was initially applied in postal automation industry. A combination of a human-like holistic analysis, multiple neural networks and sophisticated statistical voting algorithms enabled a significant improvement in recognition rates and decrease in error rates in mail processing.

Parascript first started its work in mail automation in 1998, when the Remote Computer Reader (RCR) applied by the United States Postal Service was recognizing about 35% of machine printed and 2% of handwritten letter mail pieces[3]. Modern systems (based on Parascript technology) recognize 95% of machine printed and about 93% of handwritten letter mail[4]. Similarly, the application of multiple engines inside Parascript's software for payment automation in the banking and financial services industry allowed Parascript to raise the read rate from 40% in 1997 to the current 80% at 1% error rate.

The universality of these algorithms and methods makes them fully applicable to the medical imaging market. The image analysis algorithms automatically identify and clearly mark suspicious areas without obscuring the region of interest, enabling more accurate interpretation of mammograms and increased detection of calcifications and soft tissue densities.

History of Voting

Waterfall Voting

The predecessor of voting is a so called "waterfall" scheme where multiple recognition engines are configured sequentially, one after another. In this arrangement, the first recognition engine reviews all pieces of the image stream and finalizes what it can. The second recognition system gets a shot at the rejects. If this auxiliary system can recognize any of the images that the first system could not with high enough confidence, more images are finalized and the final recognition rate is improved. Generally, the second recognition engine makes moderate improvements in the overall read rate because it works only on a part of a stream. Moreover, there is a tendency to increase the final error rate if the second engine makes errors on the type of images the first engine successfully rejects.

Parallel Processing Voting

The simple voting algorithm, called "parallel processing," combines the power of two or more recognition engines or algorithms, leveraging the strengths and methods of each. Using this method, each engine reviews 100% of the image stream. The result is determined based on the majority ranking alone and not on confidence factors. Parallel processing is a simple and effective way for manufacturers of image processing software to reduce errors, and it is also possible to further improve performance by taking advantage of confidence levels reported by different engines. Even two engines are sufficient if confidence values are involved in the analysis because the system has a lot more information to work with.

The latest achievement in voting utilizes a few engines with known intermediate results. This allows a voting scheme to choose the best outcome leveraging from an understanding of the choices available at each stage. This method is efficient and delivers better recognition rates and accuracy.

Nowadays, voting algorithms are commonly used to improve the recognition results in postal and payment automation, such as reading and sorting mail and checks. However, sometimes generic methods produce disappointing results and force vendors to give up on the idea of efficiently combining multiple engines. Creating an efficient voting algorithm requires knowledge and expertise as well as a creative approach. It is important to consider the strengths and weaknesses of particular engines, their peculiarities and individual characteristics. For example, simple voting may not produce improved results if engines are not orthogonal. Let's assume a situation where there are three non orthogonal engines that come up with similar results. Two of them make mistakes in 1 of 10 cases. The third engine makes mistakes in 1 of 100 cases. How can it be decided which result to choose as a correct answer? Efficiency in each case depends on multiple factors including the ability to choose the best way to represent results, and a system of appraisals. Additionally, possessing know-how for converting a few satisfactory engines into a powerful recognition system is also very important. How does then voting mechanism apply to medical imaging?

Voting in Medical Imaging

Multiple parallel recognition processes can be applied to medical imaging as well. Each image recognition process may identify areas of interest on the mammogram image independently, without sharing information with other image recognition processes. However, image recognition processes might also work together to identify different areas of interest. After image recognition processes individually identify areas of interest or objects on the mammogram image, the different areas are compared to determine a confidence value related to the accuracy of the identifications. The comparison is done using a voting process. The voting process here is rather complicated, usually involves neural networks and differs significantly from previous methods used for address recognition and interpretation, or for capturing information from check images. Comparing the results of multiple image recognition processes allows for the mitigation of the inherent faults of the image recognition process, thus leading to reduced false-positive and false-negative rates. Additionally, methods utilizing multiple image recognition processes, rather than a single one, amicably lend themselves to multiple processor systems or networks. Thereby, they allow the image recognition processes to be processed in parallel, thus increasing computational efficiency and spreading the workload across multiple processors, which results in fast and accurate actionable information.

Expected Results

A significant reduction in false positive rates is required in order for a CAD system to become a "true friend" of any radiologist. It is only when CAD systems achieve sensitivity and false positive rates similar to a radiologist that they will be fully embraced as a second opinion tool.

Similar to improvements in payment and postal automation, the use of voting methodology in medical imaging is expected to eventually improve the false positive rates of current CAD systems by 80%, bringing false positive rates close to those of a general radiologist in the 10 – 20% range.

---

[1] 2012 National Cancer Institute. Available at: http://www.cancer.gov/cancertopics/types/breast

[2] Screening Mammograms by Community Radiologists: Variability in False-Positive Rates, by Joann G. Elmore, Diana L. Miglioretti and team, Journal of the National Cancer Institute, VOl. 94, No. 18, September 18, 2002

[3] Read and right. By Christine Velarde. PTI 10th anniversary issue, p.38

[4] USPS Postal Facts 2012. Available at: http://about.usps.com/who-we-are/postal-facts/welcome.htm

The use of voting methodology in medical imaging is expected to eventually improve the false positive rates of current CAD systems by 80%, bringing false positive rates close to those of a general radiologist in the 10 – 20% range.

Conclusion

The reduction of false-positive readings achieved through the utilization of voting methodology in medical imaging will ultimately result in decreased medical costs, emotional stress, follow-up examinations, and recalls. It is only when CAD systems achieve sensitivity and false-positive rates approaching those of a radiologist that they will be embraced as a second opinion tool with no hesitation.

Voting mechanism is believed to be able to alleviate the existing problem, by significantly reducing the occurrence of a large number of false-positive marks produced by a CAD system, improving usefulness of CAD and reducing unnecessary distractions for the interpreting radiologist.

While the voting mechanism discussed above may be used in detection of lesions in mammogram images, the methodology can be easily applied to other types of X-ray images, such as breast tomosynthesis, scans, magnetic resonance imaging (MRI), and other types of medical imaging.

About Parascript

The Parascript image analysis suite extracts meaningful information from images. Employing patented digital image analysis and pattern recognition technologies, the Parascript image analysis suite improves decision quality in medical imaging, postal and payment automation, fraud detection and forms processing operations. Parascript software processes billions of digital images per year. Fortune 500 companies, postal operators, major government and financial institutions rely on Parascript products. Organizations include the U.S. Postal Service, Bell + Howell, Fiserv, Elsag, Lockheed Martin, NCR, Siemens and Burroughs. Parascript is online at http://www.parascript.com.

6273 Monarch Park Place, Longmont, CO 80503 USA
T: 303.381.3100 | toll free: 888.225.0169
F: 303.381.3101 | info@parascript.com