ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.5

Secure medical Image classification based on Azure Cloud

Chaithra .R
M. Tech Student
Department of Computer Science and Engineering
MVJ College of Engineering, Bangalore, India

Abstract:
In cloud computing applications, users data and applications are hosted by cloud providers. With internet and cloud availability increasing numbers of people are storing their content on cloud. With cloud availability many of the medical services are also getting hosted on cloud. In this paper we design a secure medical image classification service on top of Microsoft Azure cloud. The Medical classification is implemented as a generic service, so that by changing the training images, features and classifier used it can be made to work for any kind of medical image classification. The Medical image service is available as a network service, so that users anywhere can invoke this service and classify his medical image in a secure manner.

I. INTRODUCTION

Cloud computing for medical image-based research has attained significant attention in the recent years. While the number of medical image-based studies have grown at a steady rate of 3%-5% per year, data-storage requirements have significantly grown at 10%-25% per year. The advent of major commercial cloud-service providers between 2006 and 2008 such as Amazon web service (AWS), Google App Engine (GAE) and Microsoft Azure has led to the development of platforms, software and infrastructure that promote collaborative research among multiple investigators at different institutions, with the advantage of minimal overhead for maintaining the storage and computation systems. This is particularly useful for medical image-based studies using computed tomography (CT) or fundus images; that have been impacted by long wait times for storage and transfer across workstations. Thus, there is an impending need for raw medical image data management, image processing and image based evaluation systems that have cloud-based high-volume data storage, computation and sharable capabilities. Medical image-based research requires heavy computational workload associated with image analysis and collaborative device independent platforms to incorporate expert opinions from multiple institutions. The service hosted on cloud must be generic so that it can be invoked for any kinds of images. To make it generic, the service should be customizable for type of features to be extracted and the type of classifier. Also medical images are private information of patient and it should be handled in a secure manner , so that image does not reaches the hand of attackers , to achieve this security secure transmission of images is needed. In this paper we design such a secure generic medical image classification service on Microsoft Azure cloud.

II. RELATED WORK

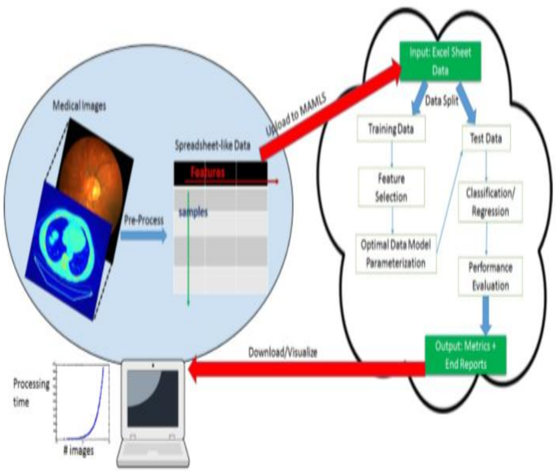
In this section we survey the current solutions for generic medical image classification systems.

In paper [1] ,.author focuses on creating a novel Generalized Flow within the cloud-based computing platform: Microsoft Azure Machine Learning Studio (MAMLS) that accepts multi-class and binary classification data sets alike and processes them to maximize the overall classification accuracy. First, each data set is split into training and testing data sets, respectively. Then, linear and nonlinear classification model parameters are estimated using the training data set. Data dimensionality reduction is then performed to maximize classification accuracy. For multi-class data sets, data-centric information is used to further improve overall classification accuracy by reducing the multi-class classification to a series of hierarchical binary classification tasks. Finally, the performance of optimized classification model thus achieved is evaluated and scored on the testing data set. The classification characteristics of the proposed flow are comparatively evaluated on 3 public data sets and a local data set with respect to existing state-of-the-art methods. On the 3 public data sets, the proposed flow achieves 78-97.5% classification accuracy. Also, the local data set, created using the information regarding presence of Diabetic Retinopathy lesions in fundus images, results in 85.3-95.7% average classification accuracy, which is higher than the existing methods. In paper [2] authors apply a Bag-of-Features approach to malignant melanoma detection based on epiluminescence microscopy imaging. Each skin lesion is represented by a histogram of codeword's or clusters identified from a training data set. Classification results using Naive Bayes classification and Support Vector Machines are reported. The best performance obtained is 82.21% on a dataset of 100 skin lesion images. Furthermore, since in melanoma screening false negative errors have a much higher impact and associated cost than false positive ones, authors used the Neyman-Pearson score in their model selection scheme. In the paper [3] , author reviews various methods applied to handwritten character recognition and compares them on a standard handwritten digit recognition task. Convolutional neural networks, which are specifically designed to deal with the variability of 2D shapes, are shown to outperform all other techniques. Real-life document recognition systems are composed of multiple modules including field extraction, segmentation recognition, and language modeling. A new learning paradigm, called graph transformer networks (GTN), allows such multimodal systems to be trained globally using gradient-based methods so as to minimize an overall performance measure. Two systems for online handwriting recognition are described. In paper [4] authors developed a cloud data analytics service based on

Daytona, which is an iterative MapReduce runtime optimized for data analytics. In our model, Excel and other existing client applications provide the data entry and user interaction surfaces, Daytona provides a scalable runtime on the cloud for data analytics, and our service seamlessly bridges the gap between the client and cloud. Any analyst can use our data analytics service to discover and import data from the cloud, invoke cloud scale data analytics algorithms to extract information from large datasets, invoke data visualization, and then store the data back to the cloud all through a spreadsheet or other client application they are already familiar with. In paper [5] author proposed MLI. MLI is an Application Programming Interface designed to address the challenges of building Machine Learning algorithms in a distributed setting based on data-centric computing. Its primary goal is to simplify the development of high-performance, scalable, distributed algorithms. Our initial results show that, relative to existing systems, this interface can be used to build distributed implementations of a wide variety of common Machine Learning algorithms with minimal complexity and highly competitive performance and scalability. In the paper [6] author presents a computer-aided screening system (DREAM) that analyzes fundus images with varying illumination and fields of view, and generates a severity grade for diabetic retinopathy (DR) using machine learning. Classifiers such as the Gaussian Mixture model (GMM), k-nearest neighbor (kNN), support vector machine (SVM), and AdaBoost are analyzed for classifying retinopathy lesions from nonlesions. GMM and kNN classifiers are found to be the best classifiers for bright and red lesion classification, respectively. A main contribution of this paper is the reduction in the number of features used for lesion classification by feature ranking using Adaboost where 30 top features are selected out of 78. A novel two-step hierarchical classification approach is proposed where the nonlesions or false positives are rejected in the first step. In the second step, the bright lesions are classified as hard exudates and cotton wool spots, and the red lesions are classified as hemorrhages and micro-aneurysms. This lesion classification problem deals with unbalanced datasets and SVM or combination classifiers derived from SVM using the Dempster– Shafer theory are found to incur more classification error than the GMM and kNN classifiers due to the data imbalance. The DR severity grading system is tested on 1200 images from the publicly available MESSIDOR dataset. The DREAM system achieves 100% sensitivity, 53.16% specificity, and 0.904 AUC, compared to the best reported 96% sensitivity, 51% specificity, and 0.875 AUC, for classifying images as with or without DR. The feature reduction further reduces the average computation time for DR severity per image from 59.54 to 3.46 s. In [7] authors presents a novel three-stage blood vessel segmentation algorithm using fundus photographs. In the first stage, the green plane of a fundus image is preprocessed to extract a binary image after high-pass filtering, and another binary image from the morphologically reconstructed enhanced image for the vessel regions. Next, the regions common to both the binary images are extracted as the major vessels. In the second stage, all remaining pixels in the two binary images are classified using a Gaussian mixture model (GMM) classifier using a set of eight features that are extracted based on pixel neighborhood and first and second-order gradient images. In the third post processing stage, the major portions of the blood vessels are combined with the classified vessel pixels. The proposed algorithm is less dependent on training data, requires less segmentation time and achieves consistent vessel segmentation accuracy on normal images as well as images with pathology when compared to existing supervised segmentation methods. The proposed algorithm achieves a vessel segmentation accuracy of 95.2%, 95.15%, and 95.3% in an average of 3.1, 6.7, and 11.7 s on three public datasets DRIVE, STARE, and CHASE_DB1, respectively.

III. PROPOSED SOLUTION

The architecture of the proposed solution



The System works in client server model and has two parts.
Medical Classification Server: This system is hosted on cloud. The administrator of this server will load training images, select the features to extract from the image, select the classifier to train and then train the service. Once the service is trained it is available for remote invocation from clients. It waits for user invocation. When user sends encrypted images, the image is decrypted using AES and features are extracted from image. The classification model is invoked with extracted features to get the result of classification. The result is then packed and sent to Medical Client user. Medical Client User: Client can be distributed anywhere on internet. When the user uploads the image to classify, the image is encrypted using AES algorithm and sent to Medical Classification Server via a TCP port. It then waits for result from server and once result arrives, the result is displayed to the user.

IV. RESULTS

The proposed solution was implemented in real cloud in Microsoft Azure. Following are features implemented in medical image classification server

1. HOG
2. HOG Mean
3. HOG Variance
4. GLCM Energy
5. GLCM Contrast
6. GLCM Homogeneity
7. GLCM Correlation

Medical image classification service implements two classifiers SVM and KNN. Service has to created on the Medical image classifications server portal and hosted.

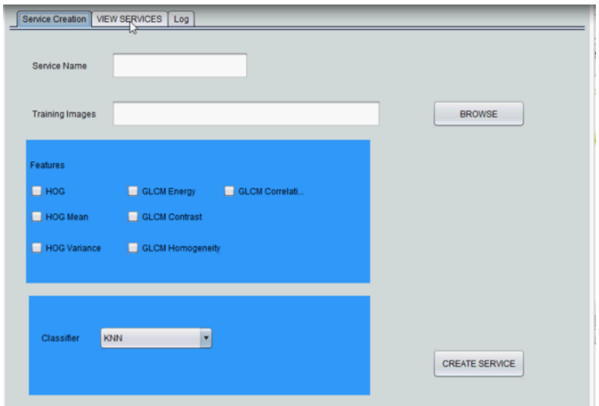


Figure.1.All the hosted services are displayed.

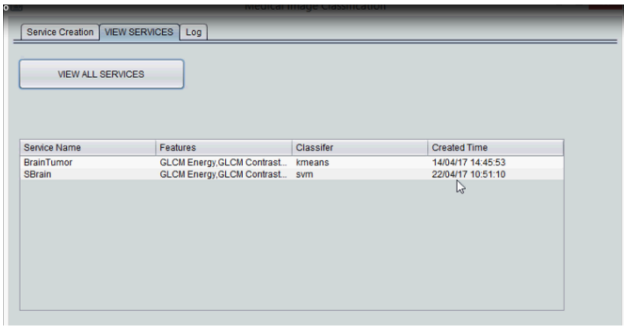


Figure.2.User can upload their images to server and view the results

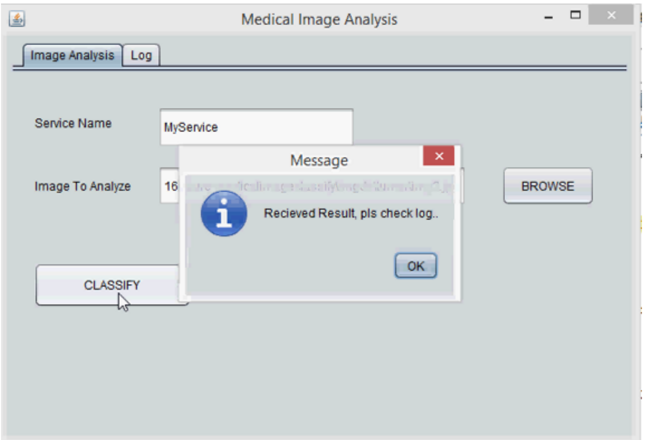


Figure.3.Once the result arrives pop up comes informing of result arrival.

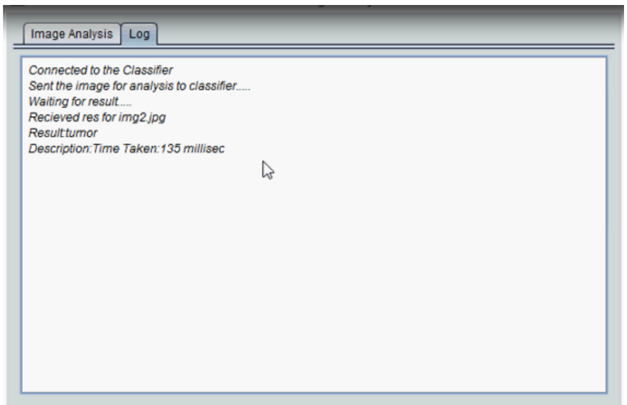


Figure.4.The result of classification and the time taken is displayed on the log.

V. CONCLUSION

The proposed medical image classification service is tested for various medical images and it found to be accurate in classification. In future we will add additional features and classifier to the system, so that the capability of system increases. Right now medical images are sent without compression, so there is a wastage in bandwidth consumption, in future we plan to reduce it by using compression schemes.

VI. REFERENCES

[1]. G. C. Kagadis, C. Kloukinas, K. Moore, J. Philbin, P. Papadimitroulas, C. Alexakos, P. G. Nagy, D. Visvikis, and W. R. Hendee, “Cloud computing in medical imaging,” Medical physics, vol. 40, no. 7, p. 070901, 2013.

[2]. L. Wang, G. Von Laszewski, A. Younge, X. He, M. Kunze, J. Tao, and C. Fu, “Cloud computing: a perspective study,” New Generation Computing, vol. 28, no. 2, pp. 137–146, 2010.

[3]. Y. Shen, Y. Li, L. Wu, S. Liu, and Q. Wen, “Cloud computing overview,” Enabling the New Era of Cloud Computing: Data Security, Transfer, and Management: Data Security, Transfer, and Management, p. 1, 2013.

[4]. M. Kubat, R. C. Holte, and S. Matwin, “Machine learning for the detection of oil spills in satellite radar images,” Machine learning, vol. 30, no. 2-3, pp. 195–215, 1998.

[5]. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Advances in knowledge discovery and data mining. the MIT Press, 1996.

[6]. C. D. Manning, P. Raghavan, H. Sch¨utze et al., Introduction to information retrieval. Cambridge university press Cambridge, 2008, vol. 1, no. 1.

[7]. G. Press. (2014, Sept) 12 big data definitions: What’s yours? [Online]. Available: http://www. forbes.com/sites /gilpress/2014/09/03/12-big-datadefinitions-whats-yours/# 314 c45921a9

[8]. D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, “Interactions with big data analytics,” interactions, vol. 19, no. 3, pp. 50–59, 2012.

[9]. T. Shepherd, M. Teras, R. R. Beichel, R. Boellaard, M. Bruynooghe, V. Dicken, M. J. Gooding, P. J. Julyan, J. A. Lee, S. Lef`evre et al., “Comparative study with new accuracy metrics for target volume contouring in pet image guided radiation therapy,” IEEE Transactions on Medical Imaging, vol. 31, no. 11, pp. 2006–2024, 2012.

[10]. M. Bihis and S. Roychowdhury, “A generalized flow for multi-class and binary classification tasks: An azure ml approach,” IEEE International Conference on Big Data (Big Data), pp. 1728–1737, 2015.

[11]. Machine learning repository. University of California, Irvine. [Online]. Available: http://archive.ics.uci.edu/ml/

[12]. H. P. Shanahan, A. M. Owen, and A. P. Harrison, “Bioinformatics on the cloud computing platform azure,” PloS one, vol. 9, no. 7, p. e102642, 2014.

[13]. R. S. Barga, J. Ekanayake, and W. Lu, “Project daytona: Data analytics as a cloud service,” IEEE 28th International Conference on Data Engineering (ICDE), pp. 1317–1320, 2012.

[14]. E. R. Sparks, A. Talwalkar, V. Smith, J. Kottalam, X. Pan, J. Gonzalez, M. J. Franklin, M. I. Jordan, and T. Kraska, "Mli: An api for distributed machine learning," IEEE 13th International Conference on Data Mining (ICDM), pp. 1187–1192, 2013.

[15]. Y. Low, D. Bickson, J. Gonzalez, C. Guestrin, A. Kyrola, and J. M. Hellerstein, "Distributed graphlab: a framework for machine learning and data mining in the cloud," Proceedings of the VLDB Endowment, vol. 5, no. 8, pp. 716–727, 2012.

[16]. S. Mund, "Microsoft azure machine learning," Packt Publ., Birmingham, 2015.

[17]. J. Philbin, F. Prior, and P. Nagy, "Will the next generation of pacs be sitting on a cloud?" Journal of digital imaging, vol. 24, no. 2, pp. 179– 183, 2011.

[18] K.-W. Chang, T.-Y. Tsai, K.-C. Chen, S.-C. Yang, H.-J. Huang, T.-T. Chang, M.-F. Sun, H.-Y. Chen, F.-J. Tsai, and C. Y.-C. Chen, "ismart: an integrated cloud computing web server for traditional chinese medicine for online virtual screening, de novo evolution and drug design," Journal of Biomolecular Structure and Dynamics, vol. 29, no. 1, pp. 243–250, 2011.

[19] X. Chen and F. Luo, "Cloud computing in bioinformatics," Scientific Computing, vol. 29, no. 5, pp. 21–24, 2012.

[20] B. Da Mota, R. Tudoran, A. Costan, G. Varoquaux, G. Brasche, P. J. Conrod, H. Lemaitre, T. Paus, M. Rietschel, V. Frouin et al., "Machine learning patterns for neuroimaging-genetic studies in the cloud," Recent Advances and the Future Generation of Neuroinformatics Infrastructure, p. 188, 2015.