Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 127 (2018) 388–397

Procedia Computer Science

www.elsevier.com/locate/procedia

The First International Conference On Intelligent Computing in Data Sciences

Security Enhancement in Healthcare Cloud using Machine Learning

Mbarek Marwan*, Ali Kartit, Hassan Ouahmane

Chouaïb Doukkali University, LTI Laboratory, ENSA
Avenue Jabran Khalil Jabran, BP 299, El Jadida, Morocco

Abstract

Image processing requires sophisticated platform because it is usually very expensive in terms of memory space and computational time. Consequentely, it is important to adopt economical solutions to replace traditional systems. These considerations led us to use cloud computing to meet large-scale data processing requirements. Meanwhile, this approach provides rapid access to on-demand services with high availability and scalability. Therefore, using cloud services instead of in-house applications would undoubtedly help healthcare organizations outsource computations to an external party, thereby minimizing operating expenses. Nevertheless, strong data protection against both untrusted clouds and unauthorized users is required to prevent malicious data disclosure. Today, various frameworks are developed to enable users to store and process their data using cloud computing. In genaral, they are built up using cryptosystems, distributed systems and sometimes a combination of both. In particular, homomorphic cryptosystems, Service-Oriented Architecture (SOA), Secure Multi-party Computation (SMC) and Secret Share Schemes (SSS) are the major security mechanisms for almost all existing implementations. The main problem in the process of massive data analysis over cloud using these techniques is the computational costs associated with image processing tasks. The first and foremost challenge is to prevent unauthorized access to medical records and personal health information. In this regard, we propose a novel approach based on machnine learing techniques to secure data processing in cloud environement. Typically, we use Support Vector Machines (SVM) and Fuzzy C-means Clustering (FCM) to classify image pixels more efficiently. Additionally, we incorporate a further level, the CloudSec module, into the conventional two layered architecture to reduce the risk of the potential disclosure of medical information. We perform two sets of experiments to evaluate the proposed technique. The simulation results demonstrate that the use of Support Vector Machines (SVM) is an efficient concept for simultaneous image segmentation and data protection. In fact, we obtain some encouraging findings which reveal new insights so as to promote cloud services in the healthcare domain.

Keywords: Cloud computing; image processing; security; machine learning; SVM; FCM

* Corresponding author. Tel.: +212678418459.
E-mail address: marwan.mbarek@gmail.com

10.1016/j.procs.2018.01.136

1. Introduction

Cloud computing can offer the possibility of outsourcing computation operations, thereby allowing healthcare organizations to leverage the cost advantages. The use of this concept enables consumers to process patients' electronic health records remotely using cloud-based imaging solutions. The principal benefit of this concept is the ability to provide ubiquitous access to imaging applications without having to buy and maintain these tools. It is commonly agreed that cloud computing has completely revolutionized the way health records are being managed, stored, processed and used. In this model, cloud providers deliver a wide variety of Health Information Technology (HIT) solutions to support E-health systems and meet the growing need for healthcare services. Interestingly, the easy utilization of cloud applications, flexible resources management, a pay-per-use basis model for charging clients are some basic characteristics that define a cloud service [1]. This obviously implies that there will be a substantial growth in demand for cloud adoption since it offers enormous opportunities for healthcare organizations. Unlike a traditional model, the usage of cloud brings security issues because clients' data are commonly stored and processed on distant data centers. There is actually a large number of factors that can specifically affect cloud solution considerations, including virtualization security risks [2], data storage location challenges [3], potentially insecure storage web technology [4], systems interoperability issues [5], [6] and legal constraints [7]. For instance, homomorphic cryptosystems, Service-Oriented Architecture (SOA), Secure Multi-party Computation (SMC) and Secret Share Schemes (SSS) are widely used as a security mechanism to ensure a minimum level of confidentiality and privacy of patient's health records. However, in many cases, these techniques cannot guarantee the Quality of Service (QoS) requirements, which are expressed and well-defined in the Service Level Agreement (SLA). First, digital records are so large that they are time-consuming in the processing stage. Second, these mechanisms are still insufficient with regard to medical data protection compliance.

Our main objective is to provide healthcare organizations with a simple and efficient cloud framework to analyze digital records using only cloud resources. We are going to show how machine learning algorithms can be used to overcome security problems in data processing. Especially, we rely, in this study, on Support Vector Machine (SVM) classifiers because they have received a growing interest in multi-region segementation techniques. In addition to providing more secure operations, the proposal is a good distributed processing solution to noticeably improve the running time. The objective of the three-level architecture is to handle additional cloud security challenges and risks, especially anonymity and unlinkability.

The rest of this paper is organized as follows. Section 2 examines the related works in this field and discusses the limitations of existing approaches. In Section 3, we describe the fundamentals of the proposed solution to address security problems in cloud-based services. Section 4 provides details about techniques involved in data protection. Section 5 reports experimental results of our proposal. Conclusion and perspectives are given in Section 6.

2. Related work

In general, significant privacy concerns might emerge when outsourcing data storage and computation to cloud providers. The confidentiality of health records in cloud storage usually involves encrypting clients'data before transmitting them into off-site servers. In this sense, several techniques are available to help deal with this issue, including AES, RSA DES, 3DES, ECC, ECDH. These methods, unfortunately, do not provide the ability to achieve secure data processing because they cannot be used in encrypted domain. In response to this challenge, we provide and discuss, in this section, techniques for implementing a secure cloud-based image processing.

In [8], the authors developed a framework that uses XML standard and Service-Oriented Architecture (SOA) to build distributed software systems. In this approach, a block of many web services is created to process digital records. Accordingly, each node is responsible for performing a specific task. Technically, the process of image analysis is first decomposed into smaller tasks each of which is mapped to a distinct cloud provider. This method is meant to ensure collaboration between business components during data processing. However, privacy protection is the main disadvantage of this technique because it processes only raw images. In the same line, image processing as a service is developed in [9]. Chiang et al. use an APIs (Application Programming Interface) that integrates frontend services to analyze clients'data easily. Mainly, SOA and ImageJ tools are used to simplify the development and deployment processes on the cloud platform. The proposal is composed of various modules to process data

remotely. They include Presentation layer, Service layer, Business Logic and ImageJ tools. One of the major concerns that arises while using SOA technology in cloud computing is security. Regarding privacy protection, SOA is not designed to prevent data disclosure both internally and externally.

Unlike the previous methods, privacy and security considerations are the key reasons behind using homomorphic encryption in data processing over cloud computing. More importantly, these types of algorithms support homomorphism so that one can do some arithmetic operations on encrypted data. In [10], a secure cloud application is developed for achieving successful results by adopting homomorphic encryption. Authors have taken advantage of Learning With Error (LWE) approach in order to process digital records safely. In this scenario, clients encrypt medical data, while cloud providers process them on behalf of their customers without decryption. In particular, both addition and multiplication are implemented in a real-world application to evaluate this concept. In spite of all its advantages, homomorphic approach is not well suited for larger volumes, especially medical images. Concretely, data encryption using this technique is very expensive in calculations, even for sophisticated computers, and hence cannot be run in practical imaging tools. Similarly, Gomathisankaran et al. [11] present a framesork which is capable of outsourcing image processing. In this system, cloud providers can analyze images without revealing any information about plaintext. This can be easily achieved with homomorphic encryption algorithms to secure image processing. In this respect, Residue Number System (RNS) scheme is used to encrypt consumers' data before they are transferred to the cloud. Basically, edge detection through the Sobel filter is carried out in order to practically test this proposed framework. The main drawbacks and limitations of this method is the time complexity, which obstructs its utilization in cloud environment.

There is a rapidly emerging need to develop a rapid method to deal with large data. In the light of this, distributed data processing approach was introduced. Accordingly, an image is split into many shares using Secret Share Scheme (SSS), allowing parallel processing. In [12], the authors have applied the SSS algorithm for conducting effective image analysis. Precisely, the proposal provides the ability to carry out data visualization using remote cloud services. In this case, a combination of SSS technique and volume rendering techniques is utilized to maintain privacy and confidentiality. In fact, this solution enables one to process created small portions in their encrypted format. Lathey et al. [13] propose a new technological method that enables online image processing via cloud computing. The main idea of the SSS technique is to prevent cloud providers to have access to confidential data. Indeed, this solution is an ecncryption tool via which a query is sent to different nodes and each server will try to process a part of it. In doing this, consumers can outsource image processing to off-site servers without disclosing the user's data. Other potential benefits of this solution include better reliability and performance. Clearly, the proposal provides efficient mechanisms to efficiently carry out some image enhancing operations such as noise removal, contrast and edge enhancement and antialiasing. Consequentely, the proposal will enable cloud computing services to be deployed in healthcare domain. Generally, the main challenge of using this approach is the application of image processing algorithm on small encrypted portions.

3. Proposed framework

Obviously, the level of data security provided by off-site systems can vary depending on the chosen cloud providers and Service Level Agreement (SLA). In general, the privacy issue becomes more complicated under the cloud technology, especially when it comes to a domain that is as sensitive as healthcare. Based on these considerations, strong confidentiality protection for medical records is urgently needed to prevent internal or external parties from reaching the secret data. More importantly, cloud providers have the possibility to process encrypted data easily while preserving privacy. In this respect, we propose a novel methodology that tries to alleviate privacy concerns that occur while performing image analysis remotely using cloud computing. Basically, our proposed cloud framework consists of two major elements for efficiently dealing with security problems, i.e., system architecture and privacy protection technique.

3.1. Proposed architecture

In the most cases, the classical architecture of cloud computing usually exposes customers' data to various and serious security threats. In response, we suggest a three-level architecture to enhance data protection in cloud environment. We consider such a system to involve three different components, i.e., Client, CloudSec and Cloud

provider. In this model, the CloudSec component first encrypts all health data via the HTTPS/SSL protocol to secure data transfer. Second, this module uses segmentation approach to keep medical images safe. Once encrypted, the CloudSec sends clients' data to an external cloud provider to process them securely. In this context, this module is responsible for ensuring adequate privacy and security for clients' data during the utilization of cloud resources. In this context, Fig. 1 presents a general overview of the proposed solution to reduce security issues in cloud-based image processing.

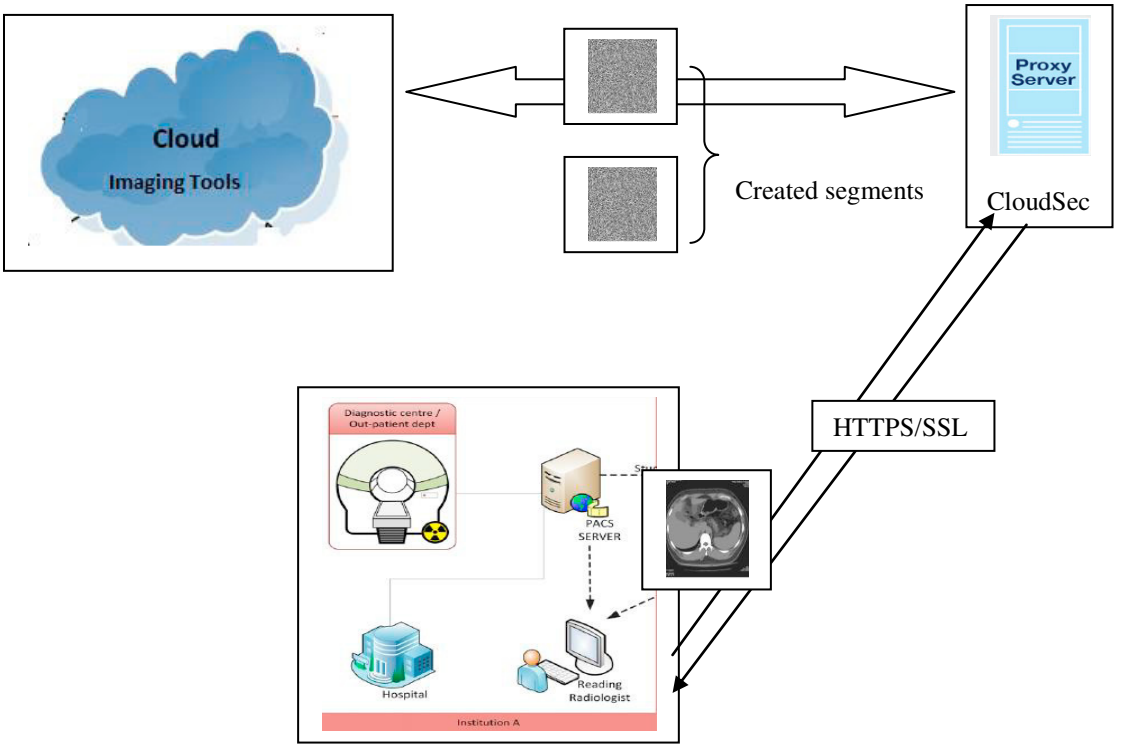


Fig. 1. An overview of cloud-based image processing

Additionally, we will restrict specific access requests to cloud services using a set of rules that defines the lists of authorized consumers and the permissions assigned to each clients. Here, the CloudSec module provides a variety of measures to allow cloud providers to process health records without disclosing confidential patients' data. In particular, this module can explicitly restrict access and detect misuse of cloud resources according to a predefined security policy. In order to support several control scenarios and strategies, there is a wide range of access control modules that prevents unauthorized access and use of patients' data nemely Discretionary Access Control (DAC), Mandatory access control (MAC), Role-based access control (RBAC), and Attribute-Based Access Control (ABAC) [14]. The latter is an extremely powerful, dynamic and flexible module, frequently used in the cloud, to meet security requirements.

In this work, we mainly focus on data protection using a segmentation approach because it is a simple and efficient technique. In order to achieve this, an alternative solution based on machine learning techniques is used to classify data efficientely into different groups based on the image features such as color, texture and shape.

3.2. Proposed data protection approach

Image segementation algorithms have been successfully used in a wide range of applications, but none of them is directly applicable in data protection. Our contribution consists mainly in applying this approch to address security issues in cloud services. The idea is to form clusters of pixels with similar features; on this basis it will be possible to

process each region separetely. In this framework, we first extract the pixel level color of the input image to identify each pixel. Next, we use these features as inputs to the Support Vector Machines (SVM) for classification purpose. The latter actually is a supervised learning algorithm for training data set from practical experiences. To this aim, we rely on Fuzzy C-means (FCM) to train SVM Classifiers in order to increase the efficiency of linear classification. The general idea of the classification approach based on SVM technique is illustrated in Fig. 2.

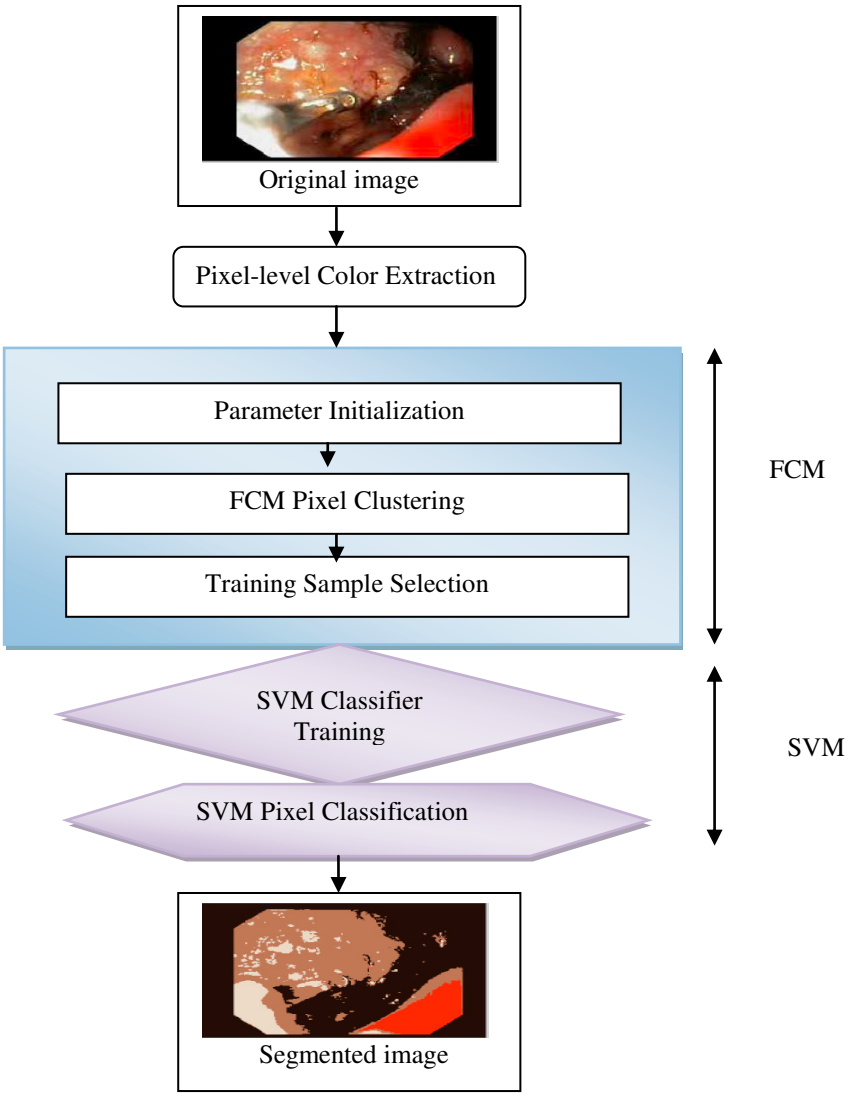


Fig. 2. The proposed method for data protection

4. Used methods

In this paper, we introduce a new privacy protection method using Support Vector Machines (SVM) combined with Fuzzy C-means Clustering (FCM). The main goal is to divide an image into distinct regions containing pixels with level image features. In most of the cases, color, texture and shape features are the fundamental units used to represent an image. Consequently, the use of FCM in conjunction with SVM is an effective approach to prevent the potential disclosure of confidential data. Meanwhile, this technique is designed to achieve fast and accurate results. This section aims at providing deep insight into the proposed method.

4.1. Fuzzy C-means Clustering (FCM)

The FCM algorithm [15], [16] is an efficient technique for unsupervised data classification and data analysis. In contrast to traditional partitioning methods, each pixel can belong to more than one cluster in fuzzy approach according to its membership grades (between 0 and 1). The fundamental objective is to select appropriate centroids that minimize the dissimilarity function. This can be easily achieved by initializing the membership matrix (U) randomly, as indicated in Eq. 1.

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,..., n \tag{1}$$

In addition, this algorithm uses the dissimilarity function, presented in Eq. 2, to classify data.

$$J(U, c_1, c_2,..., c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} \tag{2}$$

- u_{ij} is typically between 0 and 1;
- c_i refers to the centroid of cluster i;
- d_{ij} denotes the Euclidian distance between i_{th} centroid (c_i) and j_{th} data point;
- m ϵ [1,∞] represents a weighting exponent.

4.2. Support Vector Machines (SVM)

It is a simple and powerful method that automates both classification and regression by using machine learning theory. Typically, this technique uses generalized linear classifiers to evaluate data and identify patterns.To this end, it relies heavily on statistical learning theory developed by Vapnik [17] to maximize predictive accuracy. Actually, we almost always choose to apply linear functions called hypotheses, especially in case of a high dimensional feature space. The primary objective of SVM model is to calculate the hyperplane that best separates different groups and maximize the minimum distance between any data point (margin). In this sense, all points on the margin are called support vectors. In a simplified manner we may represent linear SVM model, which is the simplest form of this scheme, as illustrated in Fig. 3.

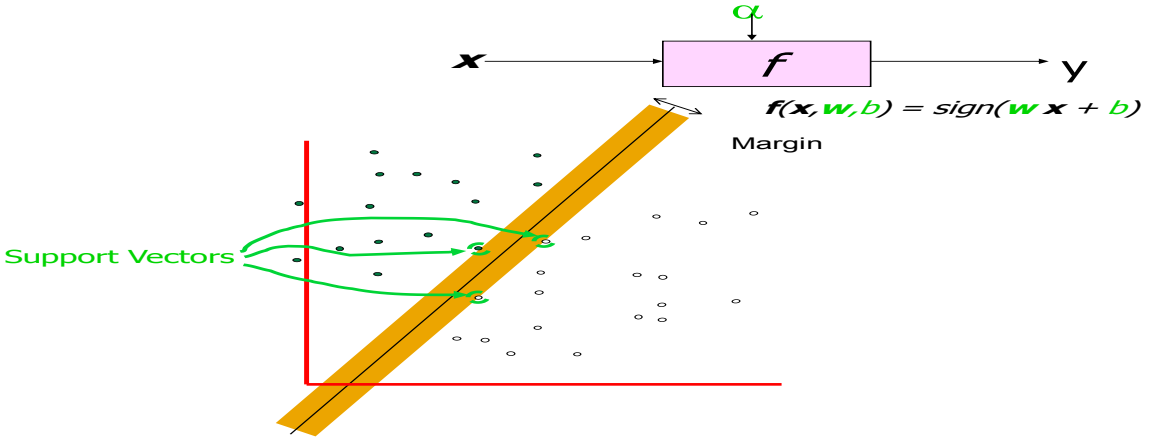


Fig. 3. Illustration of linear SVM model [18]

In this scheme, we use basically Eq. 3 to obtain the maximal possible margin of separation [19], [20].

$$\text{margin} \equiv \arg\min_{\mathbf{x}\in D} d(\mathbf{x}) = \arg\min_{\mathbf{x}\in D} \frac{|\mathbf{x}\cdot\mathbf{w}+b|}{\sqrt{\sum_{i=1}^{d} w_i^2}} \tag{3}$$

Where:

- **w**: refers to decision hyperplane normal vector
- $\mathbf{x}_i$: reprents data point *i*
- y_i: denotes class of data point *i* (+1 or -1) NB: Not 1/0
- Classifier is calculated based on the following formula: $f(\mathbf{x}_i) = \text{sign}(\mathbf{w}^T\mathbf{x}_i + b)$
- Functional margin of $\mathbf{x}_i$ is represented by the following expression: $y_i(\mathbf{w}^T\mathbf{x}_i + b)$

Mathematically speaking, hard-margin SVM can be represented by quadratic optimization problem, which is given by the following Eq. 4 [20], [21].

$$\min_{f,\xi_i} \|f\|_K^2 + C\sum_{i=1}^{l}\xi_i \tag{4}$$
$$y_i f(\mathbf{x}_i) \geq 1 - \xi_i, \text{ for all } i \quad \xi_i \geq 0$$

As in the case of dual problem, the SVM can be expressed most succinctly as Eq. 5 for learning linear classifiers.

$$\min_{\alpha_i} \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j) \qquad 0 \leq \alpha_i \leq C, \text{ for all i;} \tag{5}$$
$$\sum_{i=1}^{l}\alpha_i y_i = 0$$

Where ξ_i value denotes slack variables, ideally to be used as the measurement error at point ($\mathbf{x}_i$, y_i).

In reality, conventional techniques (hard-margin) do not usually take into account the noise and uncertainty in the data. To address this issue, soft-margin SVM uses typically the concept of slack to handle noisy data. In this case, SVM formula becomes $y_i(w'x + b) \geq 1 - S_k$, where S_k is the allowed distance that seperates data with the hyperplane without violating the predefined constraints.
Based on these considerations, we use Lagrangian variable to represent soft-margin SVM, as in Eq. 6 [20], [21].

$$min\ L = \tfrac{1}{2}\, w'w - \sum \lambda_k (y_k (w'x_k + b) + s_k - 1) + \alpha \sum s_k \tag{6}$$

Where, $0 \leq \alpha_i \leq C$ for all α_i.

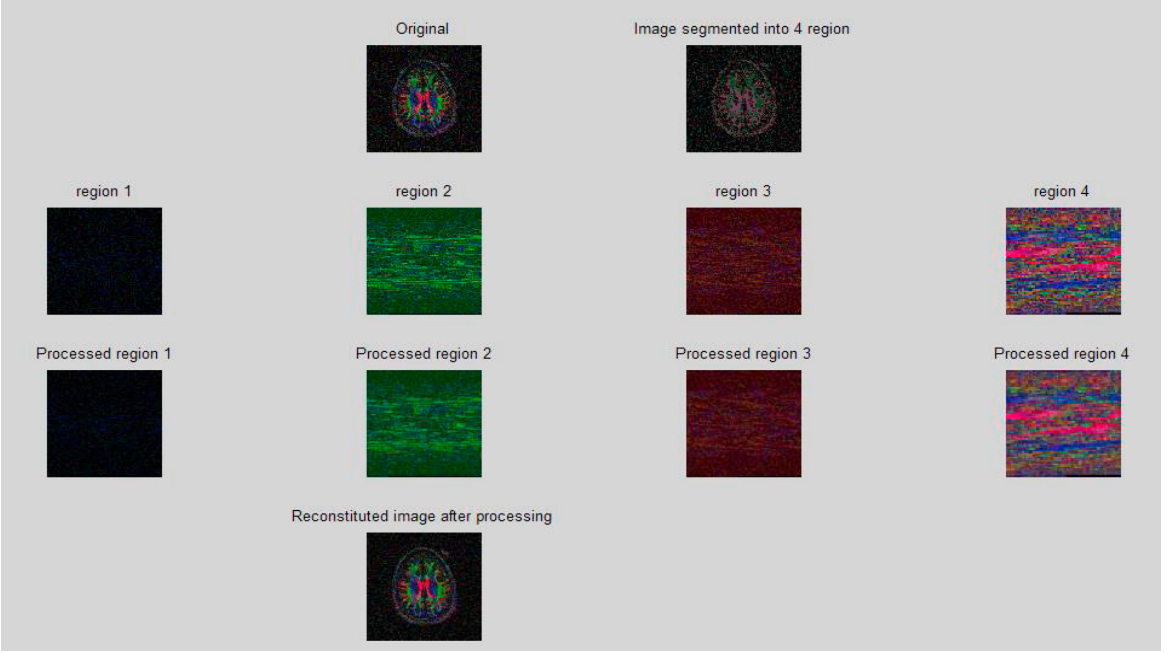
5. Simulation results

Obviously, segmentation is an efficient method to create homogeneous parts according to the selected features. In order to achieve this aim in a concise and simplified manner, we use color level to identify each image pixel to perform pixel-based segmentation. To implement our proposed method we rely on the LIBSVM [22] to easily build SVMs. In fact, LIBSVM is an efficient and powerfull open source library machine learning tool that performs data classification. In particular, we select the one-against-one approach in the multi-class SVMs implementation. Meanwhile, we implement FCM algorithm to train SVM Classifiers by using color feature. To demonstrate the accuracy of the proposed segmentation method, we apply this approach on RGB color image (JPG format). In this case, we divide the secret image into 4 regions according to pixel's color.

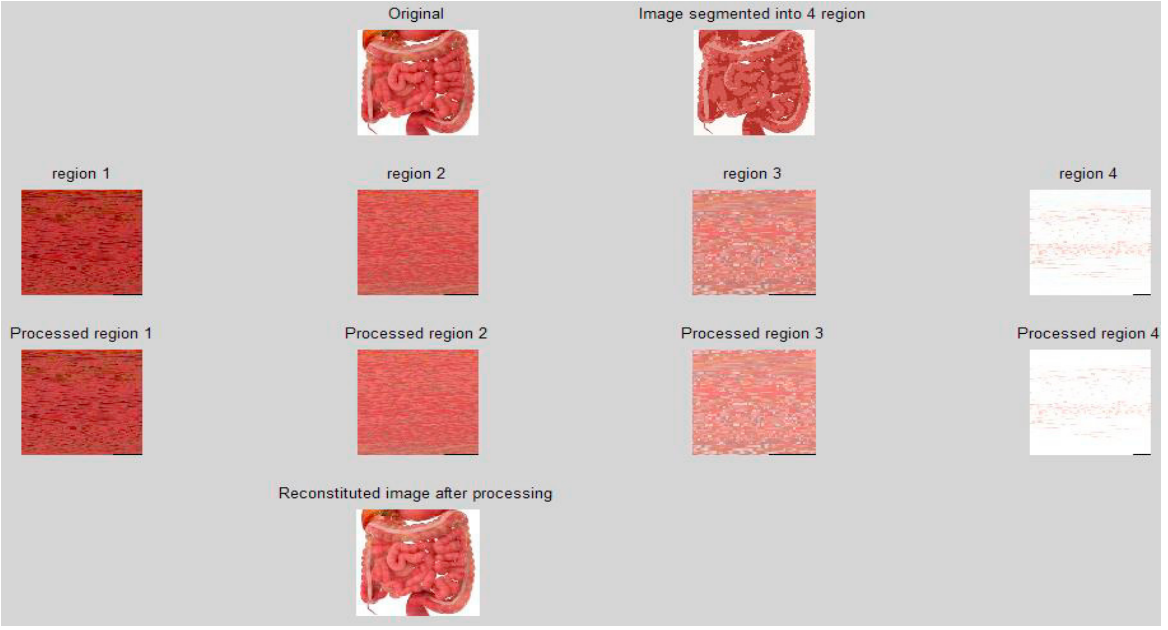
In this concept, we use segmentation approach to apply some basic image processing on each region. Next, we combine these processed parts to get the final image. In this work, we perform two basic functions, i.e., Gaussian filter and image enhancement.

First, we apply Gaussian filtering to blur images and remove noise and thereby enhancing the visual quality of the processed image. In this context, image pixels are grouped into four small regions. Next, we apply Gaussian filtering on each part. Lastely, we regroup these processed regions to get the final results. Second, we change pixel value to enhance image quality.

To evaluate the accuracy of our proposed approach, the experiments are conducted on two RGB images, as presented in Fig. 4.



(b) Pixel enhancement



(a) Gaussian filtering

Fig. 4. Image processing using SVM-FCM model (a) Gaussian filtering; (b) pixel enhancement

6. Conclusion

Thanks to cloud computing, Software-as-a-Service (SaaS) model is an efficient solution to facilitate the deployment of scalable Health Information Technology (HIT). However, the issue of security is one of the major concerns of medical image analysis over cloud computing. Results of the literature review show that existing works have limitations regarding security and privacy. In this respect, we use a hybrid approch based on SVM and FCM algorithms to split an image into many parts according to pixels' color. The initial idea behind this concept is that cloud providers could not reveal automatically hidden medical data. The simulation results demonstrate the utility of the combination of machine learning and segmentation in order to improve data security. Hence, our proposed method offers the possibility of processing RGB colorful images using cloud resources in a secure manner. More importantly, we apply two methods for image enhancement, i.e. Gaussian filtering and pixel intensity. Our future work will focus on working on more complex image processing operations. In addition, we intend to use pixel texture feature as an additional parameter to enhance image segementation. We will also use homogeneity model insteated of pixel color, which consists of two components, i.e., local image window and Gabor filter. In the same line, we will use robust medical image datasets with adequate data volume, annotation, truth and reusability. This approach may have significant ability to properly train data and extract the optimal feature. From this perspective, this technique would help cloud providers to easily protect data privacy, avoiding expensive encryption methods.

References

[1] Peter Mell, and Tim Grance. (2009) "The NIST definition of cloud computing." *Technical Report, National Institute of Standards and Technology,* vol. 15, pp. 1-3.

[2] Mazhar Ali, Samee U. Khan, and Athanasios V. Vasilakos. (2015) "Security in cloud computing: opportunities and challenges." *Information Sciences,* Elsevier, pp. 357–383 (2015).

[3] Fernandes D.A.B., Soares L.F.B., Gomes J.V., Freire M.M., and Inácio P.R.M. (2013) "Security issues in cloud environments: a survey." *International Journal of Information Security, Springer***13(2)**:113–170.

[4] Marwan Mbarek, Kartit Ali, and Ouahmane Hassan. (2016) "Cloud-based medical image issues." *International Journal of Applied Engineering Research*, vol. 11, pp. 3713-3719

[5] D. Petcu. (2011) "Portability and interoperability between clouds: challenges and case study", in Abramowicz, W., Llorente, I.M., Surridge, M., Zisman, A., Vayssière, J. (eds.) *ServiceWave 2011. LNCS*, Springer, Heidelberg, vol. 6994, pp. 62–74.

[6] S. P. Ahuja, S. Mani, and J. Zambrano. (2012) "A survey of the state of cloud computing in healthcare," *Network and Communication Technologies***1(2)**: 12–19.

[7] S. Pearson, and A. Benameur. (2010) "Security and trust issues arising from cloud computing", *in the Proceedings of the International Conference on Cloud Computing Technology and Science (CLOUDCOM),* IEEE, 693-702.

[8] V. Todica, and Vaida M.F. (2008) "SOA-based medical image processing platform", *in the Proceedings of the of IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR),* IEEE, pp. 398-403.

[9] W. Chiang, H. Lin, T. Wu, and C. Chen (2011) "Building a cloud service for medical image processing based on service-orient architecture", *in the Proceedings of the 4th International Conference on Biomedical Engineering and Informatics (BMEI),* pp. 1459–1465.

[10] R.K. Challa, Kakinada J., Vijaya Kumari G., and B. Sunny (2015) "Secure image processing using LWE based homomorphic encryption", *in the Proceedings of the International Conference on Electrical, Computer and Communication Technologies (ICECCT),* IEEE, pp. 1–6.

[11] M. Gomathisankaran, X. Yuan, and P. Kamongi (2013) "Ensure privacy and security in the process of medical image analysis", *in the Proceedings of the International Conference on Granular Computing (GrC),* IEEE, pp. 120–125

[12] M. Mohanty, P.K. Atrey, and W.T. Ooi (2012) "Secure cloud-based medical data visualization", *in the Proceedings of the International Conference on Multimedia (ACMMM'12),* ACM, Japan, pp. 1105-1108.

[13] Ankita Lathey, and Pradeep K. Atrey. (2015) "Image enhancement in encrypted domain over cloud," *ACM Transactions on Multimedia Computing, Communications, and Applications***11(3):** 1-24.

[14] Younis A. Younis, Kashif Kifayat, and Madjid Merabti. (2014) "An access control model for cloud computing." *Journal of Information Security and Applications,* Elsevier, vol. 19, issue 1, pp. 45–60.

[15] Jang J.S.R., Sun C.T., and E. Mizutani. (1997) "Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence." *In Prentice Hal, Englewood Cliffs (ed),* New Jersey, pp. 426-427.

[16] Bezdek J. C. (1973) "Fuzzy mathematics in pattern classification", *Ph.D. Thesis, Applied Math*., Center, Cornell University, Ithaca, USA.

[17] V. Vapnik. (1995) "The Nature of Statistical Learning Theory", Springer, N.Y., ISBN 0-387-94559-8.

[18] Andrew Moore. "Tutorial slides", [Online]. Available: http://www.cs.cmu.edu/~awm

[19] C. Burges. (1998) "A tutorial on support vector machines for pattern recognition", in Kluwer Academic Publishers (ed), *Data Mining and Knowledge Discovery,* Boston, volume 2.

[20] Nello Cristianini, and John Shawe-Taylor. (2000) "An Introduction to support vector machines and other kernel-based learning methods", in Cambridge University Press (ed).

[21] V. Vapnik. (1998) "Statistical learning theory", in Wiley-Interscience Publication (ed), New York.

[22] C. C. Chang, and C. J. Lin. (2001) "LIBSVM: a library for support vector machines", [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.