



ISSN : 2347 - 2243

*Indo - American Journal of  
Life Sciences and Biotechnology*



[www.iajlb.com](http://www.iajlb.com)

Email : [editor@iajlb.com](mailto:editor@iajlb.com) or [iajlb.editor@gmail.com](mailto:iajlb.editor@gmail.com)



# Implementing Generative Adversarial Networks and Cloud Services for Identifying Breast Cancer in Healthcare Systems

<sup>1</sup>Vijai Anand Ramar

IBLESOFT INC, Florida, USA

[vijaiandramar@gmail.com](mailto:vijaiandramar@gmail.com)

S. Rathna

Sri Ranganathar Institute of Engineering and Technology

Coimbatore, India.

[rathnajack@gmail.com](mailto:rathnajack@gmail.com)

## ABSTRACT

The study primarily aims at investigating the deployment of GAN and cloud services in the identification of breast cancer. Providing a method of deep learning models integrated into cloud computing offers processing power and storage facilities on scalable and on-demand bases; it is the most critical solution when considering the huge bulk of medical data accumulated in healthcare systems and IoT-wearables. The research emphasizes that preprocessing methods, such as data normalization and augmentation, will improve the GAN model's efficiency. The features are extracted by Autoencoders and then classified via GANs for breast-cancer-image classification. This model has features like easy accessibility, hosted on-cloud, and scalable for deployment. Accuracy, precision, recall, and F1 score are among the used performance metrics to evaluate the effectiveness of the model. The results indicate high classification accuracy and robustness of the model, suggesting potential of employing cloud-based GANs for the breast cancer detection clinical settings.

**Keywords:** Breast Cancer Detection, Generative Adversarial Networks, Autoencoders, Cloud Computing, Feature Extraction, Medical Image Classification

## 1. INTRODUCTION

With the introduction of cloud computing to the healthcare sector, institutions can rent computing power without heavy upfront investments into infrastructure [1]. It has transformed IT resource provisioning by permitting scalable, on-demand access to storage, processing, and applications over the Internet [2]. Cloud computing provides the backbone to accommodate the voluminous data generated by healthcare systems, associated wearable devices, and IoT sensors, particularly patient records [3]. Cloud technologies enhance collaboration among interdisciplinary teams in healthcare by providing remote access, storage, and analysis of data [4]. That convenience has made cloud computing a major focus area for professionals across all sorts of fields: geologists, social scientists, astrophysicists, business practitioners, and industrialists [5]. It also highlights new avenues for Molecular Dynamics simulations in AMBER [6]. Increasingly rapid progress in technology and the digitization of patient records have facilitated the flow of data among healthcare workers [7]. Secure data sharing and processing across multiple clouds and sources backend of interconnected healthcare environments is a mandate nowadays [8]. Any breach in these critical mechanisms, whether user interfaces, data storage, hardware, or software, could severely damage health care services [9]. Altogether, these technologies work in the framework of the strong ecosystem [10] to optimize patient care and resource utilization and contribute to sustainable healthcare solutions [11], eventually giving patients and healthcare providers more reliable, efficient, and accessible medical services [12].

## PROBLEM STATEMENT

The early detection of breast cancer is instrumental to increasing patient outcomes; however, the long uptake period of diagnosis methods and the susceptibility to human errors is a core problem [13]. Integrating deep learning models with cloud computing technologies tends to be a good remedy for the fact that healthcare systems generate millions of data [14]. Nevertheless, although being challenged by how to process massive data efficiently, preserve privacy, and provide results to be scalable and accurate for the application in the clinics, these research areas will develop a breast cancer detection model that uses GAN for data generation and cloud services for scalable deployment [15]. This study will focus on using Autoencoders for feature extraction and subsequently classifies them with GANs to make the accurate automated identification of patients' breast absences while tapping cloud infrastructure for their accessibility and scalability facilitation [16].



## Objective

- GAN can be analysed as potential methods for generating synthetic data to improve breast cancer classification results. Develop infrastructure cloud-laden in the Healthcare for scalable data storage, processing, and the model deployment.
- Analyse whether using Autoencoders is effective in feature extraction for breast cancer pictures to develop improved performance models.
- Design an effective classifier differentiating malignant masses from a breast affecting benign growths using GAN and deep learning approaches.
- Measure the performance of developed models using different evaluation metrics: accuracy, precision, recall, and F1-score.
- Integrate cloud services for scaling, access, and treatment of high volume data in healthcare applications with the breast cancer detection capabilities.

## 2. LITERATURE SURVEY

Deep learning has come to play an increasingly important role in disease prediction and diagnostics as healthcare undergoes transformation via cloud computing [17]. They go on to describe the possibilities for cloud-based deep learning frameworks to use the very large medical datasets that are needed for the diagnosis [18]. They noted that cloud computing acts as an enhancer to computational effectiveness; AI models like AlexNet can work with large amounts of medical data. Also, it was demonstrated that the application of CNNs for disease detection thus corroborated the promise of deep learning architectures for predictive healthcare initiatives [19].

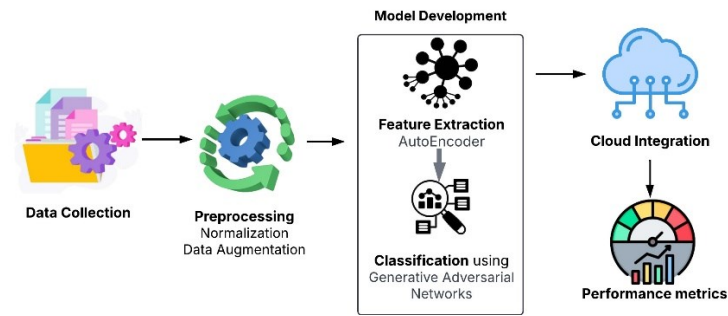
Meanwhile, much promise has been attributed to CNNs and deep learning systems for medical image analysis for diabetes prediction. They compared various machine learning methods to arrive at one model for the analysis of retinal pictures in detecting diabetic retinopathy [20]. As per, AlexNet performs much better. Analogous arguments have been advanced by [21], citing CNNs for cloud-based diagnostic frameworks for accuracy and computational efficiency perspectives. Both these researchers seem to have unequivocally proved the viability and effectiveness of deep learning models for application in medical imaging and diagnosis prediction integrated with cloud computing [22].

Surely, the pros of cloud computing provide a cutting-edge system of healthcare delivery through AI; however, it is the security and privacy factors that remain the prime concern. [23] detailed the potential risks of hosting medical-sensitive data on a cloud platform, elaborating on secure operational ways for transmitting such data. [24] further pointed out the computational constraints noted in AI-based diagnostic systems and proposed using GPU acceleration on the cloud instead, thereby optimizing speed-of-model performance. The above works also call for striking some balance between optimization for performance and securing medical applications hosted on the cloud [25].

To uplift the accreditation levels of cloud AI models, privacy-preserving methods have to be sought for healthcare applications. [26] declared that the cloud would host AI architectures optimized for speedier and more accurate predictive applications in health for diagnosis. [27] proposed the implementation of federated learning in conjunction with cloud computing, in a move that would further enhance the security of the data by decentralizing its model training across several devices. Such investigations unveil the benefits and potential of the integration of AlexNet with the cloud towards predictive diabetes prediction based on alleviation of computationally intensive limitations [28].

## 3. PROPOSED METHDOLOGY

The workflow of developing a breast cancer detection model using GANs is represented in the diagram. The workflow begins with Data Collection, where relevant healthcare data is collected. The second step is Preprocessing, which contains Normalization (for scaling the data) and Data Augmentation (for creating more data through transformations such as rotation and scaling). In a subsequent stage after preprocessing, Model Development commences, which contains Feature Extraction using Autoencoders to learn compact arrangement of data. Which is then classified using GANs represented through processed features. Integrated with Cloud Services for efficient handling of data and scaling, the model is evaluated performance-wise through Performance Metrics to assess accuracy, precision and recall, and overall effectiveness of the model. In this way, the model is ensured to be robust and scalable for the deployment with a high level of accuracy.



**Figure 1:** Cloud-Enabled GAN Model Development

### 3.1 Data collection

Collection of data by obtaining digitized pictures of fine needle aspirates of breast masses and thereafter features computed to represent the shape of cell nuclei. This collection often referred to as the Breast Cancer Wisconsin Dataset, contains 357 benign and 212 malignant cases each with 30 real-valued features such as radius, texture, perimeter, area, smoothness, compactness, concavity, and symmetry in addition to the diagnosis being benign or malignant. The database for this has numerical attributes, as well as associations with a given label, which are essential for training machine learning models, particularly GANs for synthesizing data, alongside classifiers in detecting cancers. The data can either be obtained from popular public repositories such as the UCI Machine Learning Repository or directly from health institutions that have access to FNA diagnostic datasets with an assurance that the dataset is well curated and validated for cancer-related tasks.

### 3.2 Preprocessing

The first portion of the dataset preparation process mainly deals with preprocessing before feature engineering for effective model training. Cleaned data devoid of any form of missing or inconsistent values are then normalized for all the numerical features (for instance, radius, area, texture)-using Z-score normalization or min-max scaling. Further, categorical variables, such as the diagnosis, are converted into binary values (malignant or benign). If image data are to be contained, resizing all pictures to a standard size followed by applying data augmentation methods, like rotation, flipping, and image cropping, will also help create model robustness and reduce overfitting. In addition, feature selection methods may be applied to choose the most relevant attributes, as fine as splitting the dataset into training, validation, and testing sets to ensure model could generalize well against unseen data. These preprocessing procedures ensure the data set is clean, imbalanced, and providing the state-of-art scenarios for deep learning models and machine learning algorithms to improve performance and accuracy.

#### 3.2.1 Normalization

Normalization is the preprocessing technique which is used to scale the numerical features to the standard range that will be one more way in which you can say that no feature would be more dominant than the other features due to scale differences among them. One among the many schemes to perform normalization will be Min-Max normalization. This helps to convert the actual feature values into a value that normally falls between 0 to 1. The equation for min-max normalization in eqn:1

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (1)$$

Where X is the original feature value,  $X_{\text{min}}$  and  $X_{\text{max}}$  are the smallest and maximum values of the feature, respectively, and  $X_{\text{norm}}$  is the normalized value. Alternatively, Z-Score Normalization, or standardization, transforms data to have a mean of 0 and a standard deviation of 1. The equation for Z -score normalization in eqn:2

$$X_{\text{std}} = \frac{X - \mu}{\sigma} \quad (2)$$



Where  $x$  is the original feature value,  $\mu$  is the mean,  $\sigma$  is the ordinary deviation of the feature, and  $X_{std}$  is the standardized value. This method is especially useful when the feature distribution is not uniform or when the model assumes zero mean and unit variance.

### 3.2.2 Data Augmentation

Image Augmentation results from performing a variety of transformations on one image and generating many different forms out of it. The most popular of these transformations is rotation, and mathematically this transformation is given in eqn:3

$$\text{Image}_{\text{rotated}} = R(\theta) \cdot \text{Image} \quad (3)$$

where  $R(\theta)$  denotes the rotation matrix which rotates Image by an angle  $\theta$ , while Image represents the original Image. Hence, new samples are generated by changing the orientation of the image and creating a diversification of the training data.

### 3.3 Feature extraction

Here, the Autoencoders are applied in feature extraction applications, where they learn the compressed lower-dimensional representation of the input data with important features that are significant for classification. An Autoencoder is thus a neural network divided basically into two parts: encoder and decoder. The encoder directs the input data to a latent space while the decoder tries to recreate the original data from that compressed form. The features in that latent space are a result of the transformations learned by the encoder illustrating the features that were extracted. Usually, these features are more informative and more compact, which ultimately improves the speed and accuracy of next classification models. The autoencoder has learnt how to concentrate itself on the most important parts of the data, which turns out to be quite useful in the case of malignant benign classification or detection.

#### 3.3.1 Autoencoder

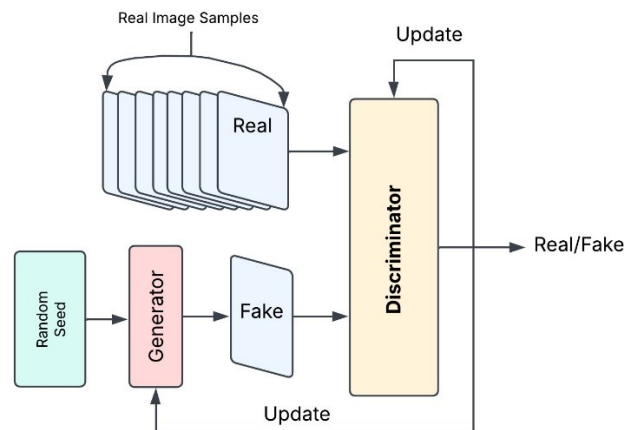
An autoencoder is simply a set of functions that creates a compressed representation of the data in a low-dimensional space and reconstructs that data back. The aim is to find the representation by minimizing reconstruction error. An equation for the autoencoder can thus be expressed eqn:4

$$\hat{x} = g_{\phi}(f_{\theta}(x)) \quad (4)$$

where  $x$  is the input data,  $f_{\theta}$  is the encoder function,  $g_{\phi}$  is the decoder function,  $\hat{x}$  is the reconstructed data after decoding. The concept of autoencoding is to minimize the reconstruction error between the input  $x$  and the output  $\hat{x}$  reconstructed.

### 3.4 Classification using GAN

This diagram illustrates how Generative Adversarial Network (GAN) works, with its two major mechanisms, the Creator and the Discriminator. The Producer takes a random seed and generates fake image samples. Fake pictures are then fed into the Discriminator, which is trained to distinguish real from fake pictures. The Discriminator, therefore, gets real pictures as well as fake pictures from the Producer, which it then classifies as either real or fake. In this manner, both the Producer and Discriminator are updated according to the response from the Discriminator. Producer learns to produce realistic pictures, and the Discriminator improves its ability to distinguish fake and real pictures. This is done in an iterative procedure. Ultimately, the competition between the two networks has caused them to improve in tandem, and the end result has been a Producer producing high-quality realistic pictures.



**Figure 2:** Generative Adversarial Network

### 3.4.1 Random Seed

The Producer will take the random noise vector  $z$  which is sampled from some kind of uniform or normal distribution. The idea is to map this random seed to a synthetic image  $G(z)$ . The general equation for the Producer in eqn:5

$$G(z) \rightarrow \text{Fake Image} \quad (5)$$

where,  $z$  is the random seed (input),  $G(z)$  is the generated fake image.

### 3.4.2 Discriminator

The Discriminator takes pictures which are real image entries from the dataset along with all the fake ones which are generated by the Producer. The mapping of the output in the Discriminator is supposed to give a probability from 0 to 1 being an indication of the authenticity of the image. Accordingly, the output of the Discriminator on a particular image  $x$ , is represented eqn:6

$$D(x) = \sigma(W_d \cdot x + b_d) \quad (6)$$

Where,  $D(x)$  is the Discriminator's output,  $\sigma$  is the sigmoid function,  $W_d$  and  $b_d$  are the weights and biases of the Discriminator network,  $x$  is the input image.

### 3.4.3 Update Producer

The Producer is updated in accordance with the identification of the fake pictures by the Discriminator. Thus, the Producer is motivated to deceive the Discriminator into identifying the pictures generated as real, attained by the minimization of loss given in eqn:7

$$\mathcal{L}_G = -\mathbb{E}_{z \sim p_z(z)}[\log(D(G(z)))] \quad (7)$$

Where,  $G(z)$  is the generated fake image,  $D(G(z))$  is the Discriminator's probability of the generated image being real,  $\mathbb{E}$  denotes the expected value.

### 3.4.4 Update Discriminator

The Discriminator is updated in a normal manner to improve the ability to distinguish between real and fake pictures. In this context, the Discriminator aims at maximizing the likelihoods of correct classification of both real and fake pictures. The Discriminator loss function in eqn:8

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(D(x))] - \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (8)$$



were.  $D(x)$  is the Discriminator's probability of the real image being real,  $D(G(z))$  is the Discriminator's probability of the generated image being factual.

### 3.5 Cloud integration

The GAN was given access to the cloud to achieve maximum efficiency for its hosting, storage, and deployment of models. Its training data had real and synthetic pictures generated by the GAN stored on the local cloud platforms that enable scalability and usage of large datasets easily-accessible, such as AWS S3 or Google Cloud Storage. Once the GAN is trained, it is hosted on AWS or Google cloud services for computation over new access data using AWS Sage Maker or Google AI Platform. A cloud-based application programming interface (API) can also be made for healthcare professionals to perform the predictions by uploading new pictures for either malignant or benign class labels without any necessity for local infrastructure. This integration promotes collaboration, flexibility, and accessibility of the breast cancer detection model, making it resource-efficient and ready for deployment into real-life settings within healthcare.

## 4. RESULT AND DISCUSSION

The diagram under consideration is a bar chart that pictures the various performance measures of the model used for breast cancer classification. These measures contain Accuracy, Precision, Recall, and F1-Score, all expressed in percentages. The values for each of the metrics were presented as follows: Accuracy-98.52%, Precision-98.20%, Recall-97.75%, and F1-Score-98.33%. These measures highlight the high performance of the model since all the values are above 97%. Thus, indicating the model's effectiveness in classifying breast cancer cases. The chart exemplifies a visual comparison among these performance measures, with F1-Score appearing higher than the others in ranking.

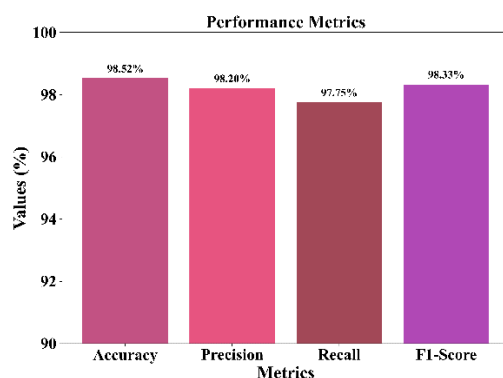
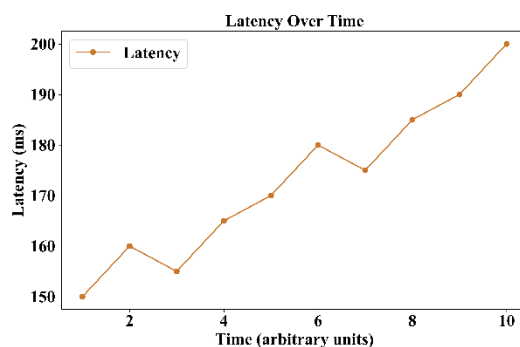


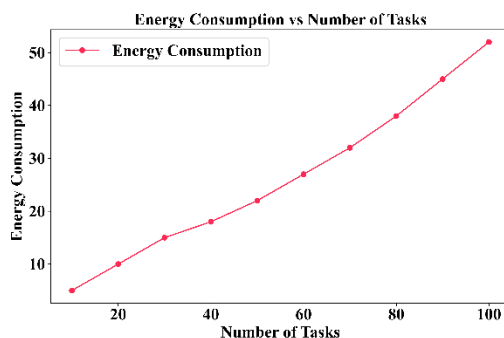
Figure 3: Performance Metrics

A chart depicts the connection of time in arbitrary unit measurement with latency. The latency measured in milliseconds increased over time. Initially, the values were lower than 150 and increased steadily until they were 200 or slightly more. The line connects data points in accordance with how much latency is measured at a specific time unit. It is noted that, as time goes by, an ongoing increase in latency continues, pointing to what seems to be the linear increase of latency over the observed time period.



**Figure 4:** Latency Growth Over Time

The figure represents a line chart comparing energy with the number of tasks accomplished. As the number of tasks increases from zero up until a hundred, so does the energy, which has directly proportional values with the other variable. This energy consumption is given in umbra dimensions-most commonly watts or joules in such a case-and the number of tasks totaled within the chart is from ten up to a hundred. The chart illustrates how task performances cause a gradual increase in energy consumption, thus creating awareness that more tasks will need energy to be completed.

**Figure 5:** Energy Consumption vs. Number of Tasks

## CONCLUSION

The method proposed can be very beneficial in using GAN along with cloud services for breast cancer classification. With the Autoencoders for feature extraction and cloud computing for infinite scaling of the model coupled with the predictions, this solution can be considered very strong and useful for healthcare institutions. All these properties make it a reliable device for differentiating malignant from benign cases, as shown by the positive results from high-performance metrics like accuracy, precision, recall, and F1-measures. The system captures the large amount of data available in the cloud while making it accessible to healthcare professionals to be able to make remote diagnoses. Future works should therefore contain improvements in the area of security and privacy on cloud-based AI applications and enlargement of this model to have even more diagnostic variables for enhanced clinical utility.

## REFERENCE

- [1] Sultan, N. (2014). Making use of cloud computing for healthcare provision: Opportunities and challenges. *International Journal of Information Management*, 34(2), 177-184.
- [2] Moreno-Vozmediano, R., Montero, R. S., & Llorente, I. M. (2012). Key challenges in cloud computing: Enabling the future internet of services. *IEEE internet Computing*, 17(4), 18-25.
- [3] Sakr, S., & Elgammal, A. (2016). Towards a comprehensive data analytics framework for smart healthcare services. *Big Data Research*, 4, 44-58.
- [4] Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, 10(1), 13-53.
- [5] Yang, C., Xu, Y., & Nebert, D. (2013). Redefining the possibility of digital Earth and geosciences with spatial cloud computing. *International Journal of Digital Earth*, 6(4), 297-312.
- [6] Salomon-Ferrer, R., Case, D. A., & Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(2), 198-210.
- [7] Tresp, V., Overhage, J. M., Bundschuh, M., Rabizadeh, S., Fasching, P. A., & Yu, S. (2016). Going digital: a survey on digitalization and large-scale data analytics in healthcare. *Proceedings of the IEEE*, 104(11), 2180-2206.
- [8] Fabian, B., Ermakova, T., & Junghanns, P. (2015). Collaborative and secure sharing of healthcare data in multi-clouds. *Information Systems*, 48, 132-150.



- [9] Zeadally, S., Isaac, J. T., & Baig, Z. (2016). Security attacks and solutions in electronic health (e-health) systems. *Journal of medical systems*, 40, 1-12.
- [10] Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research policy*, 43(7), 1239-1249.
- [11] Proctor, E., Luke, D., Calhoun, A., McMillen, C., Brownson, R., McCrary, S., & Padek, M. (2015). Sustainability of evidence-based healthcare: research agenda, methodological advances, and infrastructure support. *Implementation Science*, 10, 1-13.
- [12] Chassin, M. R., & Loeb, J. M. (2013). High-reliability health care: getting there from here. *The Milbank Quarterly*, 91(3), 459-490.
- [13] for Research, E. O., & European Association for the Study of the Liver. (2012). EASL–EORTC clinical practice guidelines: management of hepatocellular carcinoma. *Journal of hepatology*, 56(4), 908-943.
- [14] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2015, June). Big data in healthcare: Challenges and opportunities. In *2015 International Conference on Cloud Technologies and Applications (CloudTech)* (pp. 1-7). IEEE.
- [15] Andreu-Perez, J., Poon, C. C., Merrifield, R. D., Wong, S. T., & Yang, G. Z. (2015). Big data for health. *IEEE journal of biomedical and health informatics*, 19(4), 1193-1208.
- [16] Yang, M. Q., Yu, S., Cruz-Neira, C., Yang, W., Tudoreanu, M. E., Li, D., ... & Zhao, W. (2017, December). Secure Privacy Preserving across Personal Health Data and Single Cell Genomics Research INSPIRE Academic Pedagogy—Merging Big Data Multiplatform with Deep Learning. In *2017 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1244-1251). IEEE.
- [17] Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., ... & Zanetti, G. (2016). Making sense of big data in health research: towards an EU action plan. *Genome medicine*, 8, 1-13.
- [18] Pakdel, R., & Herbert, J. (2016, June). Scalable cloud-based analysis framework for medical big-data. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 2, pp. 647-652). IEEE.
- [19] Ker, J., Wang, L., Rao, J., & Lim, T. (2017). Deep learning applications in medical image analysis. *Ieee Access*, 6, 9375-9389.
- [20] Roychowdhury, S., Koozekanani, D. D., & Parhi, K. K. (2013). DREAM: diabetic retinopathy analysis using machine learning. *IEEE journal of biomedical and health informatics*, 18(5), 1717-1728.
- [21] Obinikpo, A. A., & Kantarci, B. (2017). Big sensed data meets deep learning for smarter health care in smart cities. *Journal of Sensor and Actuator Networks*, 6(4), 26.
- [22] Fernandez, A., Lopez, V., del Jesus, M. J., & Herrera, F. (2015). Revisiting evolutionary fuzzy systems: Taxonomy, applications, new trends and challenges. *Knowledge-Based Systems*, 80, 109-121.
- [23] Latif, S., Qadir, J., Farooq, S., & Imran, M. A. (2017). How 5G wireless (and concomitant technologies) will revolutionize healthcare? *Future Internet*, 9(4), 93.
- [24] Bou-Harb, E., Fachkha, C., Pourzandi, M., Debbabi, M., & Assi, C. (2013). Communication security for smart grid distribution networks. *IEEE Communications Magazine*, 51(1), 42-49.
- [25] Lounis, A., Hadjidj, A., Bouabdallah, A., & Challal, Y. (2016). Healing on the cloud: Secure cloud architecture for medical wireless sensor networks. *Future Generation Computer Systems*, 55, 266-277.
- [26] Bacon, J., Eyers, D., Pasquier, T. F. M., Singh, J., Papagiannis, I., & Pietzuch, P. (2014). Information flow control for secure cloud computing. *IEEE Transactions on Network and Service Management*, 11(1), 76-89.
- [27] Suzic, B., Prünster, B., Ziegler, D., Marsalek, A., & Reiter, A. (2016). Balancing utility and security: Securing cloud federations of public entities. In *On the Move to Meaningful Internet Systems: OTM 2016 Conferences: Confederated International Conferences: CoopIS, C&TC, and*



*ODBASE 2016, Rhodes, Greece, October 24-28, 2016, Proceedings* (pp. 943-961). Springer International Publishing.

[28] Seneviratne, S., Hu, Y., Nguyen, T., Lan, G., Khalifa, S., Thilakarathna, K., ... & Seneviratne, A. (2017). A survey of wearable devices and challenges. *IEEE Communications Surveys & Tutorials*, 19(4), 2573-2620.