International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 13, No.1 (Mar-23)

http://dx.doi.org/10.12785/ijcds/130150

GAN-Based One-Class Classification SVM for Real time Medical Image Intrusion Detection

Fadheela Hussain[1], Khaoula Tbarki[2], Riadh Ksantini [3]

1,3 Department of Computer Science, College of Information Technology, Manama, Bahrain
2 Artificial Intelligence at Private Higher School of Technology and Engineering, Tek-Up University, Tunisia

E-mail address: fadheelaali15@hotmail.com, rksantini@uob.edu.bh@hotmail.com, khaoula.tbarki@gmail.com

Received 12 Sep. 2022, Revised 16 Nov. 2022, Accepted 5 Mar. 2023, Published 16 Mar. 2023

Abstract: Medical data attack and detection technology has been a hot topic in the past few decades precisely as numerous attacks on hospitals and clinics led to the loss of data. Although many methods have been developed for detection and discrimination of fake images, the problem has not yet been properly solved. Classifying the data method tends to produce higher error when compare to other methods due to the large variance directions. One-class classification is a fairly competitive method for detecting fake medical images due to the data's unbalanced nature. However, it can also produce higher error when compare to other methods. One of the most effective ways to improve the accuracy of one-class classification is by implementing covariance-guided support vector machine (iCOSVM) especially with a real time system. Therefore, in this paper, we present a case study that uses incremental covariance-guided support vector machine to build suitable detection system. The results of the study showed that the proposed detection system is very accurate and efficient. It utilizes the training data to improve its accuracy and minimize its error. The iCOSVM supports incremental projections, improves significantly the performance of the one-class support vector machine. Additionally, our proposed detection system is very accurate and efficient comparing to other incremental one class classifications algorithms, outperforming the batch learning system as well.

Keywords: Fake, Tumor detection; incremental learning; VGG-16, GAN, MRI scans, Medical images, synthesis images; one-class classification; multiclass classification, iCOSVM, iMOSVM, and iOSVM, detection application, supervised learning, DCNN, incremental Covariance-guided One-Class Support Vector Machine (iCOSVM).

1. INTRODUCTION

Due to the increasing amount of medical data collected and stored electronically, it has become more important than ever that the data is protected from unauthorized access. This can lead to various issues when it comes to diagnosing and treating patients. In order to protect the privacy of your medical data, it is important that you only access it through a secure internet connection. Medical images are commonly used for diagnosing and treating various conditions. Some of these include CT scans, MRI, and X-ray. Due to the nature of the data collected and stored electronically, it is important that the security measures are taken to protect it. Attacks that can alter the image of a medical device can easily fool a skilled doctor. Medical images are very sensitive and essential to a patient's health. Each pixel in the image is necessary for a diagnosis [1], and any deformation could cause a faulty diagnosis. To ensure the security of the data transmitted, there are various techniques that are designed to protect against unauthorized access. Besides protecting patients' privacy, securing medical images and other sensitive data is also important to prevent unauthorized access. Consequently, artificial intelligence, machine and deep learning are becoming more prevalent in medical domains due to the evolution of techniques and the increasing expectations of patients. Machine learning has gained widespread recognition in the security industry due to its ability to build highly accurate security applications. For instance, it can perform various tasks such as identifying and securing malware and monitoring health prediction systems. Several studies have shown that the use of machine learning in medical image security can help prevent unauthorized access and exploitation. Machine learning can achieve exceptional accuracy when handling data-driven problems such as security audits and classification of malware. This technology can also be used to secure health prediction systems and other data-based systems from any potential risks. Due to these threats, various techniques have been proposed in medical domain to address the issue. For instance, DeepFake detection tools concerning the medical data security [2] use multimodal detection techniques to

analyze data and detect the inconsistencies. More work concerning the medical security[3] attempt to evaluate the capabilities of deep neural networks and machine learning algorithms to distinguish authentic and tampered data. Another study [4] that focused on medical fraud and abuse detection system based on machine Learning proposed a model that can detect abnormal records. Moreover, a study[5] proposed an algorithm to address the authenticity of the medical image problem by which one can detect and localize tampering in a digital medical image uses discrete wavelet transform method. Furthermore, a paper [6] presented a hybrid watermarking technique that combines the Singular Value Decomposition and the Discrete Wavelet transform. It allowed to verify the modifications that were carried out through the receiver's actions. Another publisher[7] proposed a hybrid medical image watermarking scheme that combines the functions of Singular Value Decomposition and the DCT. It achieved high security and efficiency while reducing the cost of doing so. Most of the published current works and methods are implemented to study or designed to detect medical tampering in medical data or images using machine learning techniques. They are usually implemented in combination with other methods to improve the detection of specific forms of tampering. Unfortunately, the current generation of online image detection tools is not designed to effectively detect medical tampering. This paper aims to develop a novel framework that will allow medical image tampering detection in real time. It will also help secure the medical data in a secure and timely manner to detect fake medical images. Our recommends system uses an incremental Covariance-guided One-Class Support Vector Machine (iCOSVM) method approach in medical domain which is the first work using this method in medical imagery field. The objective of the strategy is to take advantage of the various features of the machine to improve its classification. Before implementing iCOSVM , we first had to develop a deep learning CNN for extracting the features from MRI tumor samples. In this work, we first showed how to generate fake images using the framework Generative Adversarial Network (GAN). After extracting the features, a CNN-VGG-16 was then used to perform further analysis. In order to classify the images, we first introduced a real-time method that can be used to check if the image is fake or real. It then performed a robust analysis to identify the system's robustness. We then performed a comparison between the iCOSVM and the existing relevant incremental boundary-based methods. We found that the iCOSVM method performed better. The outline of this paper is as follows. Section II is the related work and background, section III is the proposed Novel MRI Altering and Detection System. Section IV is the Experimental protocol, Section V is the result and discussion section, the conclusions are drawn and explained in Section VI.

2. RELATED WORK AND BACKGROUND

Due to the weak security protocols in medical facilities [8], sensitive information about patients was able to be accessed and stored on the Internet. A few study focused on developing systems to detect a fraud encountered the medical data. For instance, DeepFake detection tools [2] use multimodal detection techniques to analyze data and detect the inconsistencies. The study evaluated eight different machine learning methods. These include three support vector machine, five deep learning models, and a decision tree. It was able to identify untampered and tampered images. Deep learning models are mainly used for feature extraction. They are then trained to detect abnormalities in images, such as tumor removals and injections. The results of the study show that these models are very accurate in detecting these abnormalities. Another study[3] aimed to investigate the capabilities of deep neural networks and machine learning algorithms to distinguish authentic and tampered data. The goal of this research was to develop a method that can classify cancer scans using deep neural networks. The results of the study showed that the proposed system can improve the performance of cancer scans by up to 90%. It also reduced the number of false alarms. Another work [4] build a model that can identify abnormal records in the healthcare system using a multi-label prediction method. The main advantages of their model was its simplicity and its ability to perform well in terms of accuracy and recall. It eliminates the need for data analysts to perform complicated calculations. A paper [5] presented a method that aims to avoid medical images being modified by means of equivalence checking. The goal of this process is to check if the images have been subjected to illegal modifications. The method is focused on comparing the generated automaton from the sender and the generated one from the receiver. It allows to verify if the modifications were carried out through the receiver's actions. Further work [6] presented a hybrid watermarking technique that combines the functions of Singular Value Decomposition and Discrete Wavelet Transform. It achieved high security and efficiency while reducing the cost of doing so. The paper tested the efficiency of the hybrid technique proposed against various attacks such as salt and pepper noise, filtering attack, and Gaussian noise. The performance of this technique is evaluated by taking into account the NC, SSIM, and PSNR. The results of the simulation showed that the PSNR is above 37 dB, which significantly improved the imperceptibility of the technique. It also shows better robustness against image attacks. Also another work [7] aims to learn a representation that can enable the recognition of these queries during inference. The results of the proposed scheme show that it is very robust. Most of the extracted watermarks exhibited NC values of 0.9 and above after these attacks.

This study proposes a robust strategy for detecting adversarial images that can be used against deep learning systems that classify medical images. We start with pre-trained CNN classifier to perform the detection. Then adapting fake MRI images using GAN framework. Also a CNN used to extract useful features. Furthermore a real-time classification method used to identify if the image is real or fake.

A. Generation of realistic images using Generative Adversarial Networks (GANs) in medical domain

Generative adversarial networks, or GANs, are a promising technique for creating artificial intelligence systems that can alter medical images introduced in 2014[9]. A new study [10] shows how easy it is to use deep learning to alter images and fool the best radiologists. GAN can be used in various positive ways, such as improving the quality of medical images and overcoming the scarcity of patient data. On the other hand, it can also be harmful. Although GANs are known for making realistic media possible, they're also being used in the medical field to create more realistic images. One study shows how they can help doctors identify skin lesions[11] that they're not able to see in real photos. Other studies explore how they can be used to create liver lesions[12]. GANs are a clever technique for training a generative model by taking advantage of the two sub-models in its learning framework. These allow the model to generate realistic examples of a given problem in various domains, such as image-to-image translation and rendering realistic images of people. The field is rapidly becoming more popular due to its ability to generate realistic and relatable images. The model has been able to perform well in various image generation tasks, such as the medical image synthesis. Papers presented in various academic institutions revealed how doctors and radiologists are exposed to various types of GAN attacks. There are also several methods that can be used to detect if an image is GAN-generated or not. One of the main challenges in storing large datasets of images is finding a way to make them work efficiently while keeping them in a secure environment. Several GAN-based image detection architectures have been proposed[13], [14] and they show good accuracy even after compression. However, these are not ideal for generating synthetic data. Due to the increasing number of proposed architectures, the next generation of these tools require the training of more sophisticated models. Recent advancements in the field of artificial intelligence (AI) Generative Adversarial Networks (GANs) [15] have allowed the generation of realistic images by implementing several generation methods, such as single-shot or few-shot learning. These methods can also reduce the visible artifacts and patterns in images. For instance, by reducing the number of strange and blurred objects, GANs can also improve the performance of their models. Due to the lack of distinguishing features between GAN images and real images, they are not widely considered to be accurate representations of reality. One of the most common methods of detecting GAN images is by training a Convolutional Neural Network (CNN) and a binary classifier. However, recent research [16] has shown that this method can be improved by analyzing the patterns and artifacts in the images. A binary classifier and a Networks (CNNs) that can be used to analyze large numbers of images from GANs is known to perform better than a network. Some researchers[17] have shown that it can also improve the detection performance by analyzing the patterns and artifacts in the images. Current methods for detecting GAN-images [15]. have performed well when compared with the training dataset. They can also be used to develop a well-structured binary classifier[18] that can be used on existing CNN architectures.

One of the main reasons why transfer learning performance is not improved by using methods such as CNN is because it trains on one dataset. In 2018, Forensic Transfer [19] introduced an autoencoder for detecting GAN-image images. This method works by reconstructing GAN-images using an error-free reconstruction algorithm. The advantages of using Forensic Transfer are its low data usage and its ability to transfer knowledge about GAN-image detection to other models. However, its performance remains mediocre. In previous research, various artifacts and patterns were used to enhance the performance of the model, but this method can also be combined with other methods to improve performance. Several methods are available to detect GAN-images, such as the addition of a learning method or the transformation of the image. A more recent technique, which involves implementing multi-task incremental learning[20] can provide transferability between different types of GAN-image. This method, which is based on the loss function and autoencoder, can be used to improve the performance of existing methods. A proposed technique[21] for transforming GAN-images into data augmentation is based on the combination of JPEG compression and co-occurrence matrices. The increasing number of threats that the medical imaging industry is facing has prompted many companies to rethink their security training. Unfortunately, many of them do not understand the importance of maintaining a secure environment. According to a study conducted by Kaspersky Lab in 2019, only 29% of healthcare workers are aware of the Health Insurance Portability and Accountability Act of 1996 (HIPAA).Researchers were able to alter the 3D images created by CT and MRI Scan machine by taking advantage of a flaw in the imaging technology. Both medical devices use powerful magnetic fields to create 3D images of the body. These scans are commonly used to diagnose various conditions such as arthritis, bone, cartilage, and joint problems.

In order to diagnose various diseases, such as cancer, MRI or CT scans are used. Today, these medical devices are managed through a system known as a PACS, which stands for Picture Archiving and Communication System. This network allows users to access and store images from multiple imaging devices. It then retrieves the data from the

scans and allows radiologists to analyze and interpret them. The medical scans are sent and stored in a digital format known as DICOM. This is a standardized communication and imaging format used in medicine. In 2019, Mirsky and et al [22]showed that GAN technology can be used to create convincing fake medical images. They trained an unsupervised learning system known as GAN to either add or remove a lung cancer from a computed tomography scan. The training dataset was derived from a database containing over 800 images of public research images. According to the researchers [23], they trained the CT-GAN to remove lung cancer and then recruited three radiologists to examine the tampered scans. The results of the experiment were then analyzed in two trials: one blind and one open. In the blind trial, the radiologists were able to identify 99% of the injected patients with cancer, while 94% of them were healthy. After learning about the attack, the radiologists failed to identify almost 80% of the patients who had been injected with cancer and 87% of those who had been removed. Artificial intelligence is being studied to supplement the human workforce, with many models surpassing the capabilities of doctors. A study reveals [24] how to use a machine learning technique to alter or delete cancer-related images from a patient's medical records using a PACS infrastructure to change MRI or CT scan Images by implementing GAN algorithms. The researchers use this method to perform various tasks, such as generating fake tumors and removing real ones.

GANs are being used in the healthcare industry to address the various challenges faced by the industry in image analysis and labeling. They are widely considered to be useful in both cybersecurity and medical industry. GAN can help improve the performance of machine learning models by generating new samples that closely resemble the data they're based on. It can also be used to develop adversarial models [23]. to attack them. In a study published in 2013, Li and et al[25] used a modified GAN to perform attacks against the Android cloud firewall. The attack they created consisted of two discriminators. The goal of the generator is to generate adversarial examples that are different from normal ones. The two discriminators' objective is to distinguish between malicious and benign examples. This allow the generator to be resistant to malware detection systems. In a previous work, Zhang and et al [25] presented a method to detect cross-site scripting attacks. They proposed a new Monte Carlo Tree Search algorithm that takes into account the different stages of the tree's evolution. They also improved it by implementing a GAN-based model to detect tree-based adversarial examples. In order to improve the detection rate of discriminators, they added additional adversarial examples. The researchers found that deep learning could help improve the performance of models[26]. Despite the negative publicity surrounding GANs, they have many commercial applications. For instance, they can create highly realistic audio and images [27] for the $135Bn online gaming industry. One of the most important applications of GANs is in malware detection[28]. Because of their ability to create new types of malware, which are indistinguishable from real code, they are considered a threat to businesses. Through their work, scientists can also create new labels for their data, which will allow them to train their AI systems even better. Through the use of GANs, we can quickly identify new types of attacks and prevent them from happening. In cyber-security, we can use them to predict and prevent the spread of malware. Cyber-security professionals are faced with a daunting task [29] when it comes to identifying and defeating sophisticated attackers using GANs. These tools are a powerful new tool that can help users improve their detection capabilities and protect their customers. Therefore, in this work we face this challenging problem and propose a method specifically, using GAN to generate the fake images.

B. Feature selection using Deep Convolutional Neural Network Models DCNN in medical domain

A deep learning algorithm was able to infiltrate a healthcare organization and trick an AI system into believing that it was medical images. Although the exact reasons behind the vulnerability of deep neural networks against adversarial attacks are still not known, there are various defensive mechanisms that can be used to prevent attacks. A number of machine learning-based tools that can be used to detect evasion attacks are being proposed [30] [31]. Also, the progress in detecting adversarial malware is being studied [32]. Existing methods for tackling linear classifiers are either poorly accurate or only target one type of target. The authors [33] argue that the model's blind spots are the factors that determine the success of an attack. A paper [34] presented a new approach to measure the effectiveness of various models against four different types of adversarial attacks, namely, dFGSMk, rFGSMk, BGAk, and BCAk. Two methods are proposed to improve the security of systems against multiple types of attacks, namely, SecureDroid and SecMD. Although these methods can be used to secure machine learning models, securing them against malware remains a challenge. Due to the increasing number of attacks against deep learning-based malware, both defensive and adversarial mechanisms have gained momentum. A robust adversarial attack is a valuable tool for evaluating the models that are built using deep learning techniques, such as neural networks. It can also help us understand how these models work and why they fail. Generative adversarial networks are an example of how deep learning can be used to develop models [34]. One of the most popular types of networks used for image recognition is a Convolutional Neural Network. This type of network has multiple layers and is mainly used for performing various tasks such as classification and segmentation or auto correlated data due to its high accuracy[35]. CNN model is a type of medical image analysis that uses convolutional filters to learn and extract various features from medical images. Extracting grayscale images from the raw files of malware allows analysis of its

various features. These images can be used to visualize the various features of the malware. In a study conducted by Nataraj and et al [36], they presented the first use of a byte plot visualization tool to classify different types of malware. The researchers collected 9,342 samples of malware, which were all belonging to 25 different classes. The researchers extracted the various features of the malware using the GIST feature from the images. They then classified them using the K-nearest neighbor classification. In addition, they were able to extract the features from the files of the malware. using the combined decision trees and support vector machines. A proposed method for malware detection was presented by Tobiyama and et al [37], who trained a recurrent neural network RNN to extract various features of a process. They then trained a CNN to classify these features extracted by RNN. Another researchers [38] presented a deep learning model that was based on the LSTM and CNN methods to classify different types of malware on Malimg dataset. They found that the model was able to achieve an accuracy of 96.3%. To classify one-channel grayscale images from two families, the researchers used a Convolutional Neural Network. They were able to achieve an accuracy of 96.3% and 94.0% for goodware and malware, respectively. In order to classify brain tumor data, the researchers Saxena et al.[39] used three different models: the Inception V3, ResNet-50, and the VGG-16. The ResNet-50 model had the highest accuracy rate at 95%. In terms of COVID-19 detection, CNNs have dominated the field with their use of chest X-rays and CT scans. In conventional studies, the images were used to apply the CNN models to the study's overall learning method [5]. In addition to breast cancer images, there are also various applications of these images in other pathological procedures. For instance, they can be used to visualize multiple cancer cells and their gastrointestinal tract. The use of deep learning algorithms such as CNN has been widely used in the field of computer vision. Recently, several researchers have been trying to develop AI systems that can be used for intrusion detection of medical data. One of the main features of these systems is the ability to extract and capture the hidden malicious behavior features. Our work is proposing an intrusion detection model based on convolutional neural network CNN that extracted the deep features.

C. One-Class classification in medical domain

In this section of the paper, we introduce the concept of imbalanced data classification, which can be performed using one-class or binary classification models. The main difference between a one class classification and a multiclass classification is the amount of effort required to produce it (their training data) and the amount of effort needed to produce it. Multiclass and binary classification [40] require a lot of effort to produce. This is because they require a large training set and the labels of classes that are not relevant to the user's work. Due to the complexity of the task involved in injecting and removing tumor from MRI scans, the extraction and processing of images features has some limitations. For instance, while extracting various features from a wide range of images, the extraction and processing algorithms often have a hard time identifying defective products. Therefore a one-class classification system is proposed in this part of our work to be used for detecting any breaching in MRI images. It can be built on a deep convolutional neural network. In various studies, it has been shown that this model is very robust in detecting defects. One-Class Classification (OCC) is a type of multi-class classification that focuses on the recognition of positive queries in training data [41]. The goal is to learn a representation that can enable the recognition of these queries during inference. Due to the increasing interest in the field of medical data security using machine learning and artificial intelligence, this topic has attracted a lot of attention from both the academic and commercial communities. One of the most successful techniques that can be used in the development of classification applications is the support vector machine (SVM) [42]. This type of machine learning is capable of performing at least as well as other methods when it comes to the generalization error. Due to the increasing popularity of this technology, many factors have contributed to its success. One of these is the deep understanding of its theoretical foundations. Through a convex optimization procedure, the machine learning techniques can achieve the global optimal. One of the most important factors that has contributed to the success of this technology is the efficient implementation of the solution comparison to other kernel-based approaches. The OSVM method [43] is a boundary-based approach that only considers the data points that are related to the training data distribution. Although the small variance projectional directions are not considered in the OSVM method, they can still improve classification performance. It has been shown that the method tends to separate classes depending on the high variance direction. In order to maintain the robustness of the OSVM classification, we introduce the covariance matrix. This component helps in the optimization of the OSVM problem. We provide an experiment of using the top recent deep learning-based OCC methods iCOSVM to investigate the effectiveness of the system for malicious detection and medical image altering and recognition.

D. Incremental Learning Method

Different kinds of abnormality detection tasks require different approaches to classify and interpret data [43] such as medical diagnostic or intrusion detection. For instance, in network intrusion detection, the use of a different type of regression method is required. One of the most important factors that machine learning community members need to consider when it comes to developing effective classification systems is the ability to cope with the data streams. This is done through the development of incremental learning. An artificial intelligence system that learns continuously[44] from new data is called incremental learning. Incremental learning is a well-known machine learning method that continuously improves its

capabilities by learning from new data. It allows it to perform new tasks without forgetting the previous knowledge. In the detection domain, incremental learning is very important due to the increasing number of attacks on medical devices and platforms. Due to the advancements in artificial intelligence, the use of incremental learning techniques has gained increasing popularity. Despite the widespread use of deep learning in IDSs, most studies are focused on improving the performance of existing models by implementing deep learning algorithms. On the contrary, incremental learning is flourishing in image processing fields. In a paper, Roy et al. [25] present a hierarchical model for learning incremental images using deep convolutional neural networks. It takes into account various features to create super classes. The model takes into account the addition of new classes of images to the hierarchy. These classes are then trained using a limited number of retraining processes. Sarwar et al [45] also proposed a method for learning incremental images that uses CNNs. In their incremental learning method, the authors used a partial network sharing technique, unlike Roy et al.'s system. This method is similar to transfer learning techniques. It splits CNN layers into classification and shared layers. The first few layers are labeled as shared layers, while the others are used to classify images. When the learning model is retrained, the classification layers are then cloned to improve the efficiency of the model. The resulting tree structure model is a representation of the two layers. The two methods used for generating an incremental learning model differ in their approach to structure the model. Although they have the same goals, the methods limit the number of branches that can be reconstructed in the model. As per incremental learning was never used for securing medical images, this work is the first to do that by using iCOSVM. The iCOSVM has the advantage of incrementally projecting the data onto low-variance directions, thereby improving detection performance.

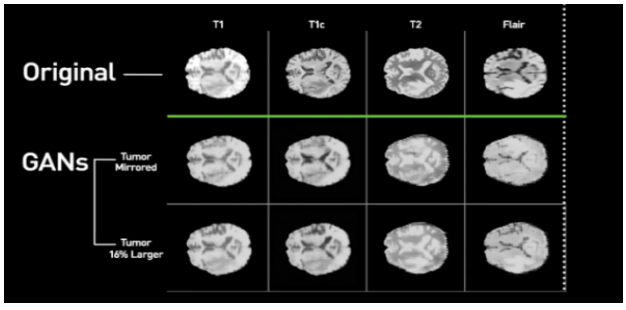


Figure. 1. Synthetic versions of MRI generated by a GAN. [Image: courtesy Shin et al.]

3. Proposed Novel MRI Altering and Detection System

This section describes, in detail, the proposed adaptive MRI Fake detection system, which consists of generate a synthesized fake images using the GAN framework. Then, a CNN - VGG-16 used to extract useful features. Furthermore, a real-time one-class iCOSVM classification method used to identify if the image is real or fake and finally to identify the robustness of the system. The source code for the MRI fake detection system is available on Github, each experiment was conducted on Google Colaboratory platform.

A. GANs' Mathematical based principle

A proposed model for generating a synthetic MRI (Fig. 1) using GAN system. The system takes into account an input vector (Z), a generator (G), and a discriminator (D). The latter two systems that are the generator and discriminator are implicit function expressions that are commonly implemented in deep neural networks. The concept of GANs is that the Discriminator tries to minimize its reward V(D,G) while the Generator tries to maximize its loss. This is done through a combination of strategies. The formula for this game is shown in mathematical way below: minG maxV(D,G)

$$V(D,G)= \mathbb{E}_{\chi\sim Pdata(\chi)} [\text{Log } D(x)]+ \mathbb{E}_{z}\sim P_z (z) [\text{Log}(1-D(G(z)))] \quad (1)$$

where,

G	=Generator
D	=Discriminator
Pdata(x)	=distribution of real data
P(z)	=distribution of generator
x	= sample from Pdata(x)
z	= sample from P(z)
D(x)	= Discriminator network
G(z)	= Generator network In particular,

fixing G and optimizing for D in (1.1.1), the optimal discriminator would be

$$D *_G(x) = \frac{P_r (x)}{P_r (x) + P_0 (x)}, \quad (2)$$

where

pr and pθ are density functions of Pr and Pθ respectively. Plugging the above D∗ G back to Equation (1.1.1), the following equation holds,

$$\min_G \{\mathbb{E}\, x\sim \mathbb{p}r \left[\log \frac{P_r (x)}{P_r (x) + P_r (x)}\right] + \{\mathbb{E}\, Y\sim \mathbb{p}_0 \left[\log \frac{P_0 (Y)}{P_0 (x) + P_0 (Y)}\right]\}$$

$$= -\log 4 + 2JS(P_r, P_0). \quad (3)$$

The goal of training GANs with the help of the equation (1.1.1) is to minimize the Jensen-Shannon -JS divergence between Pr and Pθ. Through optimization techniques, GANs can minimize the proper divergences between the generated distribution and the true distribution in a given space X. The first part of training a GAN is to train the Discriminator. When the generator is idle, the network is only propagated. This phase does not involve back-propagation. The Discriminator is trained to analyze real data for n epochs and try to predict its accuracy. In this phase, it also trains on the fake data generated by the

generator. If it can correctly predict the fake data, it will continue to train on it. The second part is the generator is training, the Discriminator is idle. After it has trained itself on the fake data, it can then perform better and fool the other GANs. For each subsequent epoch, the Discriminator checks the data generated by the generator to make sure that it is genuine. If it is acceptable, it stops the training.

B. Convolutional Networks - VGG16

A paper presented by K. Simonyan and A. Zisserman [46]of Oxford discussed the development of a very deep convolutional network model known as VGG16 for large-scale image recognition. The model achieves an accuracy of 92.7% in ImageNet, which is a database of over 14 million images. It was one of the most popular models submitted to the ILSVRC 2014 for AlexNet. It takes into account the different sizes of the various filter types used in the first and second layers of the database and automatically adds multiple 3x3 kernel-sized filters to each layer after another. The VGG-16 model is shown in Fig. 2. The network's input is an image of dimensions(224, 224, 3)., and the first two layers have 64 channels of 3*3 filter size and padding. After a max pool layer of stride(2, 2), two layers have 128 filter size and 3*3 padding layers. The next layer is a max-pooling layer(2, 2) , which is the same as the previous one. There are also two more convolution layers, each with its own filter size (3, 3) and 256 filters. After that, there are two sets of three more sets of 3 convolution layers. Each of these has 512 filters of (3, 3) size, and the image is then passed to the two sets of 2 convolution layers. Max-pooling and convolution layers use different sizes for their filters. For instance, in the former, the size 3*3 can be used instead of the 11*11 in AlexNet and the 7*7 in ZF-Net. In some of the layers, the size 1*1 pixel is used to manipulate the multiple input channels. It also has a padding of 1-pixel to prevent the spatial feature of an image.

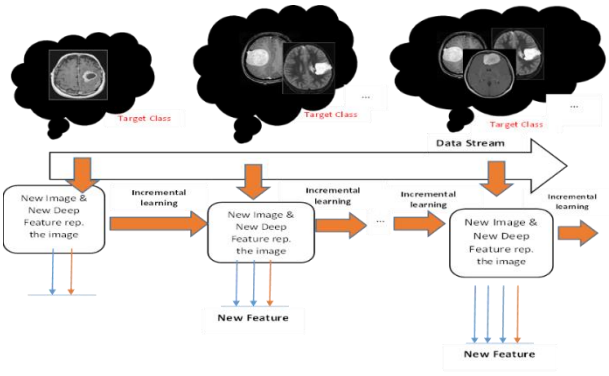


Figure. 2 Incremental learning model: the network needs to grow its capacity with arrival of images and Features Rep. the images

C. Incremental Covariance-guided One-Class Support Vector Machine (iCOSVM)

Despite its advantages, such as focusing on the low variance directions, the COSVM can have some limitations in real applications. In large training sets, the complexity of the computational complexity problem associated with the use of the COSVM increases significantly. One of the main limitations of the COSVM[47] is its ability to not handle dynamic data. This is because it cannot efficiently process the data samples that are required for the learning process. Also, the data collected from medical images datasets can be hard to train at the very beginning. An incremental learning strategy (Fig. 3) is needed to access the data collected from medical images datasets. In this paper, Kefi et al. [43] present the incremental COSVM, which can handle dynamic data. The incremental version of the COSVM can easily train and update new data without relearning the existing datasets. The advantages of implementing the iCOSVM over the older version are its lower training complexity and its ability to handle large datasets. As the OSVM method does not get special consideration in the selection of projectional directions, it can improve the classification performance. In this step, we introduce the iCOSVM, which has the high accuracy to distinguish between real and fake images. The dual formulation of this method is used to improve the performance of the OSVM. The COSVM optimization problem can be written as follows

$$\begin{cases} \tilde{\alpha} = \arg\min \mathrm{W}[\alpha, \mathrm{b}] = \tfrac{1}{2} \arg\min[\alpha \mathrm{T}\, \Gamma\alpha + (1 - \sum_{i=1}^{N} \alpha i) b] \\ \text{s.t } 0 \le \alpha i \le 1/\nu N; \sum_{i=1}^{N} \alpha i = 1) \end{cases} \quad (4)$$

The iCOSVM can update and train new data without affecting the previously trained dataset. This eliminates the need for relearning the data while still keeping the Karush–Kuhn–Tucker (KKT) conditions. It also controls the changes in the data samples after they have been acquired. The iCOSVM's control procedure is based on the KKT conditions of the training dataset and the distribution of it. The KKT conditions are used to control the distribution and

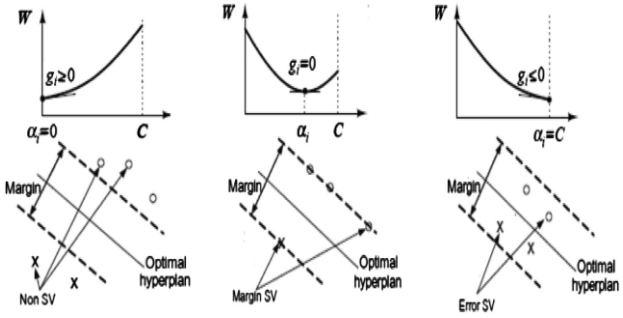


Figure. 3 The set O of non-support vectors within the margin, and the set S of margin vectors strictly on the margin, are respectively shown in 4. Finally, the set E of margin SV exceeds the margin, which is not necessarily misclassified. As the number of samples in each incremental stage increases (increasing in xc), the equilibrium of the KKT forms should be maintained. [47].

update the training dataset. The partial derivatives gi of the objective function W of Eq. 3 are shown in Fig. 4. *By allowing* $\Gamma \coloneqq \eta K + (1 - \eta)\, \Delta$, one has

$$\text{gi}= \partial w/\partial \alpha i = \sum_j \rho \Gamma i,j\alpha - b \begin{cases} \geq 0; & \alpha i = 0, \\ = 0; & 0 < \alpha i < C, \\ \leq 0; & \alpha i < C \end{cases} \quad (5)$$

Therefore, The KKT state should be assured by using limited alternatives to the following tow Eqs. (4) and (5) [47] as follows:

$$\partial W/\partial b = 1 \sum_{j=1}^{N} \alpha j = 0 \quad (6)$$

With the Γs, s and Γs, c function, we can represent the matrix entries of the SVs aa vector of kernels, between the new data sample xc, and the margin SVs. Adiabatic increment can be generated by these two Eqs., (7) and (8):

$$\begin{bmatrix} \Delta g_c \\ \Delta g_s \\ \Delta \alpha g_r \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & \Gamma_{c,s} \\ 1 & \Gamma_{s,s} \\ 1 & \Gamma_{r,s} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta b \\ \Delta \alpha_c \end{bmatrix} + \Delta \alpha_c \begin{bmatrix} \Gamma_{c,c} \\ \Gamma_{s,c} \\ \Gamma_{r,c} \\ 1 \end{bmatrix} \quad (7),$$

where the margin sensitivities γi are:

$$\begin{cases} \gamma i = \Gamma_{i,c} + \sum_{j \in S} \Gamma_{i,j}, \beta_j & i \in S, \\ \gamma i = 0, & i \in S. \end{cases} \quad (8)$$

In addition to removing and adding vectors from the set S, xc is also added to the set S. The matrix R is expanded using the Woodbury formula 8,9 as follows:

$$R_N = \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix} + 1/\gamma_c [\beta/1][\beta^T \quad 1] \quad (9)$$

A sample Xk leaves the set S and the matrix R contracts as follows:

$$R_N = R_{i,j} - R_{k,k}^{-1} R_{i,k} R_{k,j} \quad \forall i,j \in S \cup \{0\}; i,j \neq k, \quad (10)$$

where the index 0 refers to the b-term.

4. The Architecture of the Proposed System and Experimental Assessment

In this section, we present the architecture of incremental tumor detection method as it shown in Fig. 4 that was performed on an MRI database in a comparative mean. The evaluation was performed on datasets described in the next paragraph. This section aims to provide an overview of the various aspects of the paper's development, including the data collected and the experimental protocol used. It also explores the results of the study.

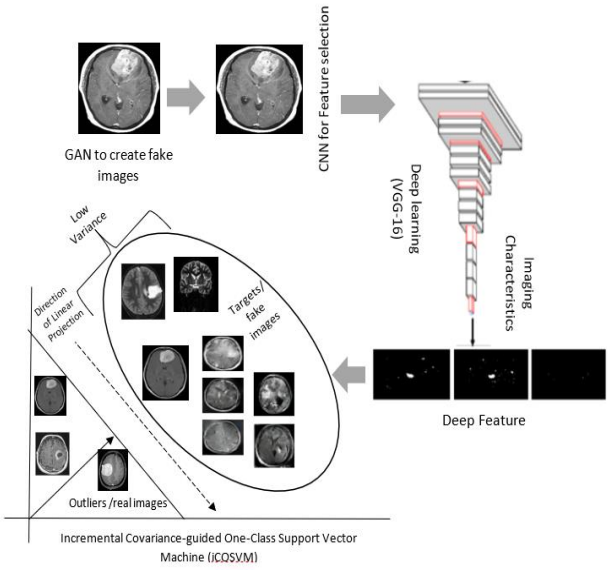


Figure.4. The block diagram of proposed real time Fake MRI tumor detected system

A. Experimental Evaluation

In this section, we present the results of our experimental and an evaluation of an incremental fake tumor detection method that was performed on the MRI database in a comparative mean. This part of the paper are organized into four main subsections. These include the datasets used, the experimental protocol, the description of the classifiers, and finally results discussion section.

B. Dataset Used

Images used in our experiment was downloaded from the Kaggle website (Brain MRI dataset). Two folders existed that are labeled "normal" and "tumor" [48]. There are two folders one represents the normal brain image and the other represents the tumor images. The WHO has classified brain tumors into four categories[49]. These include grade 1, 2, 3, and 4. The lower-level tumors that is with grade 1 and 2, such as meningioma, are usually treated with antibiotics and are usually detected through a clinical evaluation. The severe level that are grade 3 and 4 such as glioma, where Magnetic resonance imaging (MRI) is the most common method use to diagnose glioblastomas. The incidence rates of various types of tumors, such as meningioma, glioma, and pituitary, are as follows: 15%,15%, and 45% respectively. In our dataset [50], Out of the all images collected, 155 contain tumors. The remaining 98 images were without any tumors. There are over 250 images in both the brain tumor and normal MRI folders (253 images). Fig. 5 shows the samples of these two images. Fig. 5 shows the sample normal MRI Images and brain tumor image. The images in the MRI

dataset folder vary in size (eg.630X630,225X225) so these image are resized to 80x80.

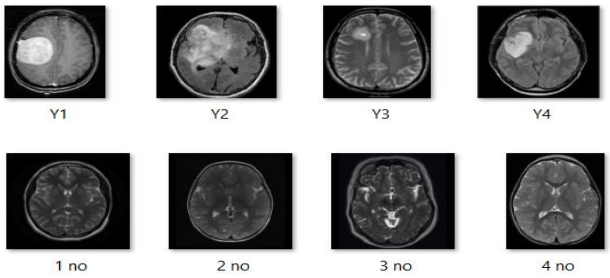


Figure. 5. Normal Brain (Y1 to Y4) and Brain with Tumor (1no to 4no)

C. Experimental Protocol

Generally speaking, our experiment composed of four main sets steps. At first set, the brain MRI image is taken as the input image in GAN network so to generate a fake tumor in the MRI brain images. Next, In order to normalize data, image thresholding and dilations are applied. This process is performed in order to remove noise. After that After that, the images were re-sized using the model's input and a pre-trained CNN. VGG-16 then classifies them into target classes that are fake and outliers (real images). The convolutional neural network model (VGG-16) is used in our experiment as the extraction of features in our study. Then an incremental learning method deployed based on OC-SVM to identify if the tumor is fake or real and prior final step of the experiments, the iCOSVM method compared with other four incremental one-class classifiers. [i.e. incremental one-class support vector machine (iOSVM), incremental support vector data description (iSVDD), and incremental Mahalanobis one-class SVM (iMOSVM)], in order to validate the superiority of iCOSVM as incrementally emphasizing low-variance directions, while classifying data. Finally, incremental one-class systems have been compared to the batch classifier for one set of the GAN,(GAN 7 used), so to prove the superiority of incremental learning in fake MRI image detection.

1) Pre-processing-Proposed GAN-based synthetic MRI Tumor Generation Approach

Our novel GAN-based approach for medical data augmentation adopts Deep Convolutional GAN (DCGAN) to generate realistic images, and an expert physician validates them. We started with the first step in this study where to create fake images using a Generative Adversarial Network (GAN). We perform this by taking into account the various features of the real MRI dataset. For better GAN training, (DCGAN) architecture results, (DCGAN [9] is a standard GAN with a convolutional architecture), the images are scaled to 80 × 80 from different sizes i.e. 240 × 155-500 for each GANs (400 images for each GANs). The total number of GAN generated were 11 (Fig. 7-11 sample) GANs where even an expert surgical physician was unable to distinguish the fake images from the real ones. This ensures that the training is stable on a high-resolution device. Fig. 6 shows some of the real images used (part of the 155 images) for the training purpose.

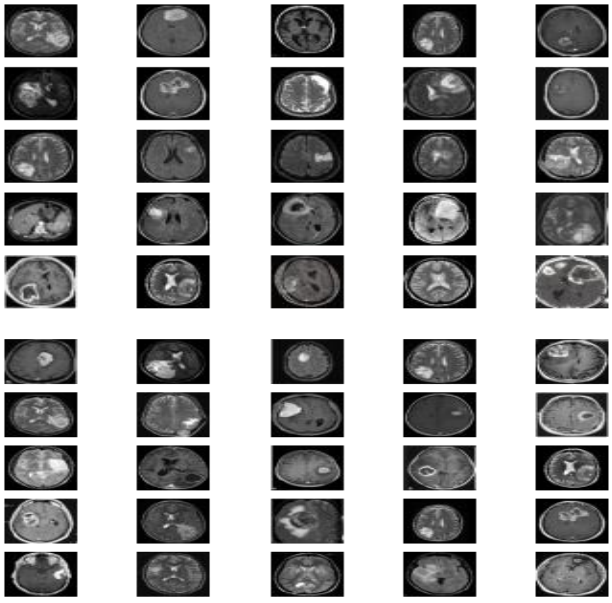


Figure**. 6.** The original MR images

a) Proposed GAN Composition

GANs are very expensive. They require high-end GPUs and a lot of time to perform properly. Table 1 and 2 shows a detailed representation of our suggested GAN design, which consists of a generator and discriminator.

- *Discriminator*

The discriminator in Table 1 is composed of, 2 Convolutional Layer (number of filter or kernel =256, size of the kernel=(3,3), stride=(2,2), pading = 'same' activation function='leaky RELU'), Dropout(=0.4), a Flatten, and fully Connected Neural Network (FCNN) (composed of one Hidden layer (Nodes=1, activation function='sigmoid'), Table 1 shows the architecture of Discriminator for our experiment.

TABLE. 1. THE ARCHITECTURE DISCRIMINATOR

Model: "sequential"		
Layer (type)	Output Shape	Param #
Conv2d (CONV2D)	(None, 40,40,256)	7168
Leaky_re_lu (LeakyReLU)	(None, 40,40,256)	0
dropout (Dropout)	(None, 40,40,256)	0
Conv2d (CONV2D)	(None, 40,40,256)	590080
Leaky_re_lu _1 (LeakyReLU)	(None, 40,40,256)	0

dropout	(None, 40,40,256)	0
Flatten (Flatten)	(None, 102400)	0
Dense (Dense)	(None, 1)	102401
Total Params: 699,649 Trainable Params: 699,649 Non-trainable Params: 0		

- *Generator*

The generator composed of one Fully Convolutional Neural Network (FCNN), (Nodes= Number of filters* Size of Image* Size of Image, activation function= 'LeakyReLU'), 2 transpose convolutional layer (Conv2DTranspose), (number of filter or kernel =256, size of the kernel=(4,4), stride=(2,2), pading = 'same' activation function='leaky RELU'), and 1 transpose convolutional layer (Conv2D), (number of filter or kernel =3, size of the kernel=(8,8), stride=(2,2), pading = 'same' activation function='sigmoid'), Table 2 showed generator architecture proposed in our experiment.

TABLE.2. THE ARCHITECTURE OF GENERATOR

Model: "sequential"		
Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 102400)	10342400
Leaky_re_lu_2 (LeakyReLU)	(None, 102400)	0
reshape (Reshape)	(None, 20,20,256)	0
Conv2d _transpose (Conv2DTranspose)	(None, 40,40,256)	1048832
Leaky_re_lu_3 (LeakyReLU)	(None, 40,40,256)	0
Conv2d _transpose_1 (Conv2DTranspose)	(None, 80,80,256)	1048832
Leaky_re_lu _4 (LeakyReLU)	(None, 80,80,256)	0
Conv2d_2 (CONV2D)	(None, 80,80,3)	307203
Total Params: 12,747,267 Trainable Params: 12,747,267 Non-trainable Params: 0		

The output images produced by the proposed GAN by the time difference of an hour (100 epochs for a batch size 50), are shown in Fig. 7 to 11.

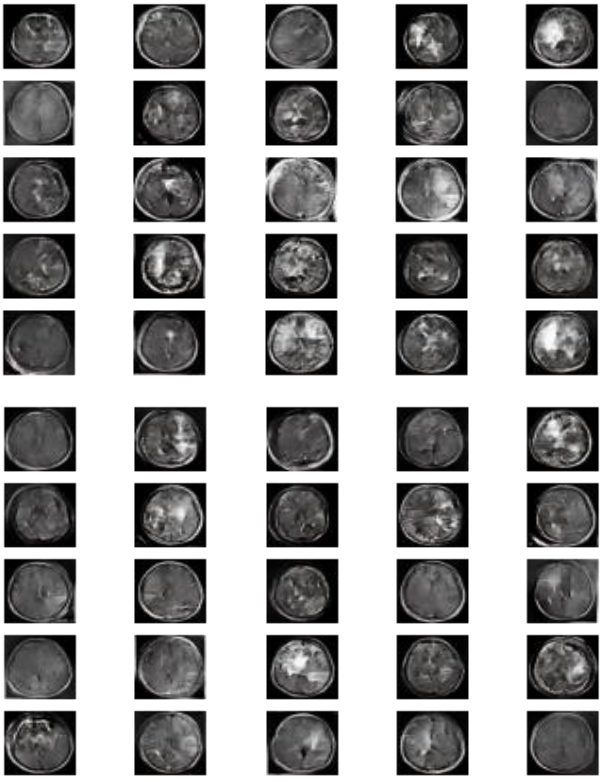


Figure.7. Gan 6 - Output images produced by the GAN by the time difference of an hour (100 epochs for a batch size 50)

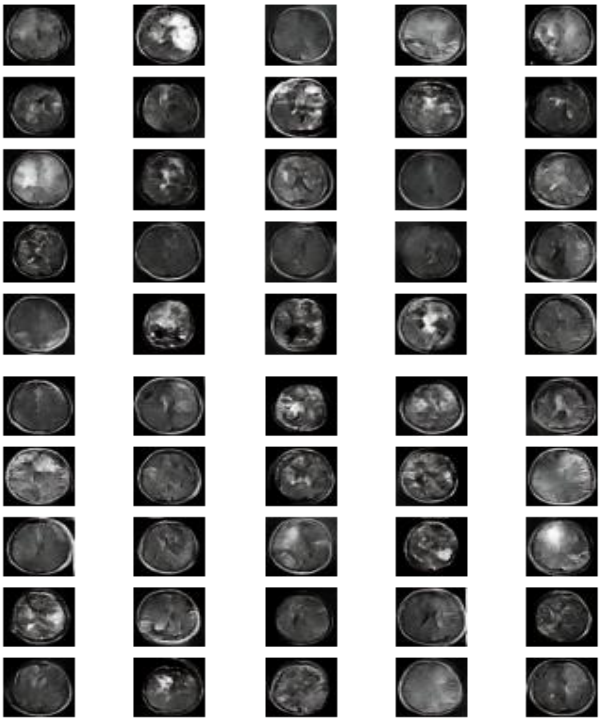


Figure.8.: GAN 7 - Output images produced by the GAN by the time difference of an hour (100 epochs for a batch size 50).

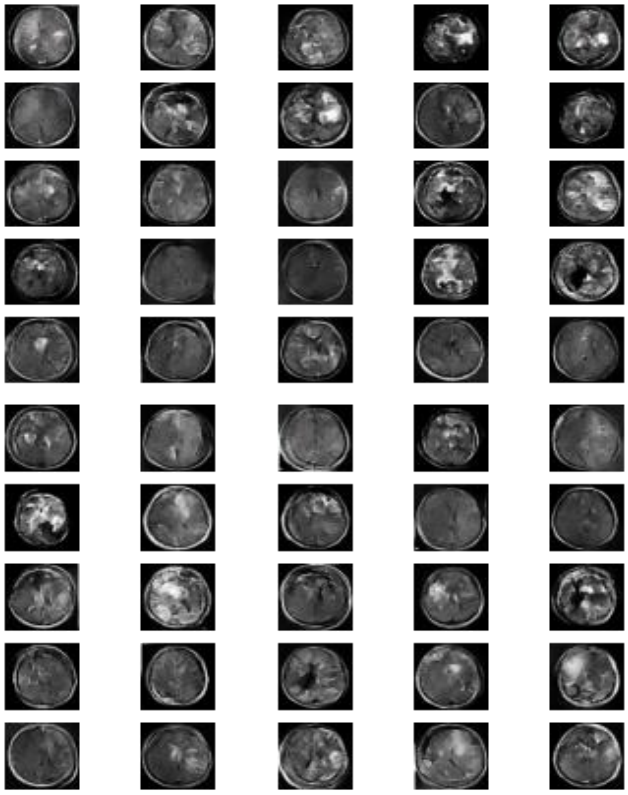


Figure.9. GAN 8 Output images produced by the GAN by the time difference of an hour (100 epochs for a batch size 50)

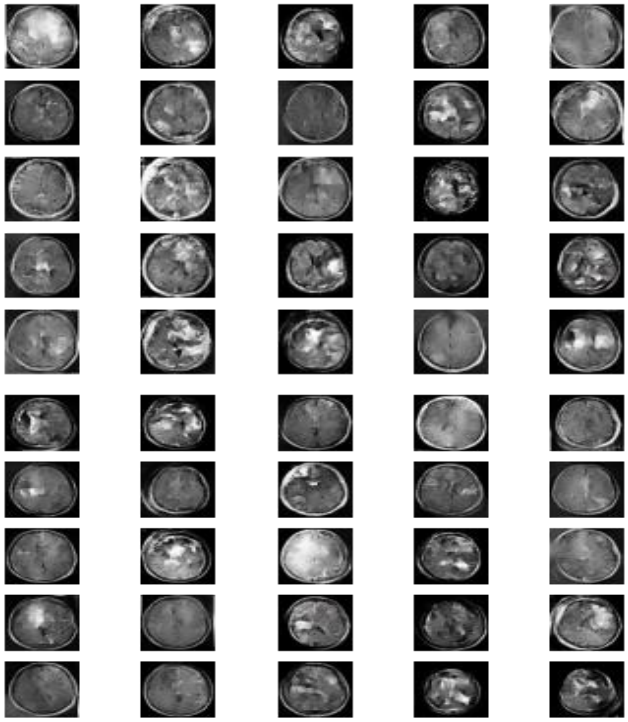


Figure.10. GAN 9 Output images produced by the GAN by the time difference of an hour (100 epochs for a batch size 50)

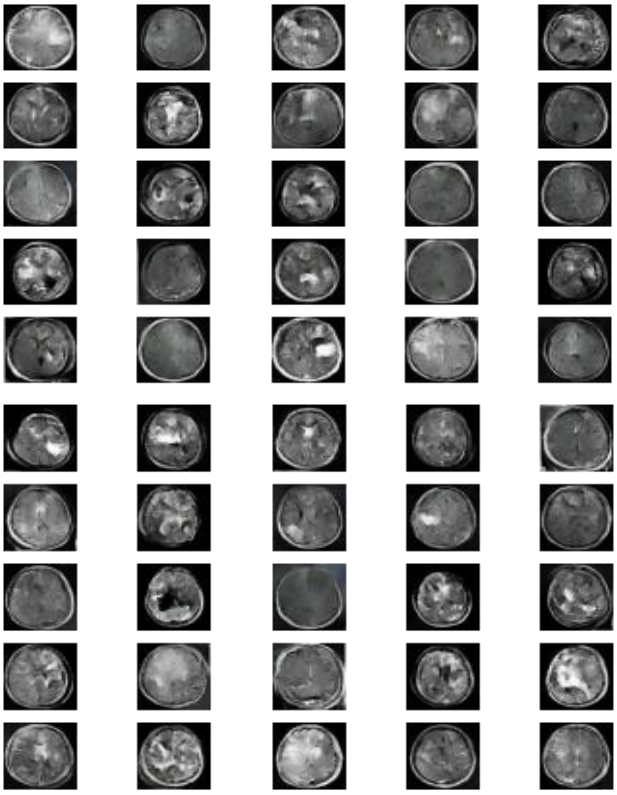


Figure.11. GAN 10 -Output images produced by the GAN by the time difference of an hour (100 epochs for a batch size 50)

b) Proposed VGG-16 -DCNN Model for feature extractuion

The second goal of this experiment is to classify the brain tumor using the VGG-16 layer. The VGG-16 classification method performed well in terms of feature selection [51]. It was able to extract deep learning features from two pre-trained models. The VGG-16 architecture was first proposed by Zisserman and Simonyan during the 2014 ImageNet Competition [52]. They were able to secure first and second places in the classification and localization categories, respectively. The main objective of the feature selection process was to remove the redundancy among the features. It was also focused on selecting those that were robust enough to be used in the classification. The goal of this step was to minimize the number of predictors. It also helped in the fast execution of the testing process. Fig 12 showed an Example of synthesis images created using GAN for feature extraction.

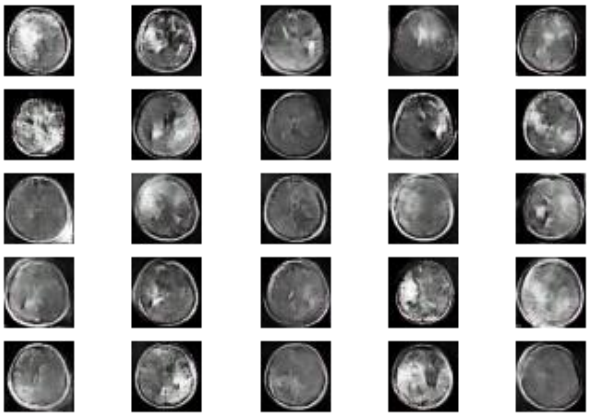


Figure.12. An Example of proposed synthesis images created using GAN for feature extraction

The VGG-16 network structure is shown in Fig 13. The first two layers are composed of 64 feature kernel filters each with a size of 3x3. As the input image where RGB image with depth 3 is passed into the second and third layers, the dimensions of these layers change with a width of 224x224x64 then the resulting output is then passed to the max pooling layer, stride of 2.The third and fourth layers of the convolutional layer are composed of 124 feature kernel filters. The size of these filters is 3x3. They are followed by a Max pooling layer, which reduced the output to 56x56x128. The three followed layers of the convolutional layer system are the fifth, sixth, and seventh. They use 256 feature maps. Max pooling is also used in these layers with stride 2. Next, the Eighth to thirteen layers are two groups of convolutional layers with 512 kernel filters that have kernel sizes 3 x 3. These layers are followed by Max pooling layer with a stride of 1.Last part of the VGG-16 architecture the Fourteen and fifteen layers are fully connected hidden layers of 4096 units followed by a softmax output layer (Sixteenth layer) of 1000 units.

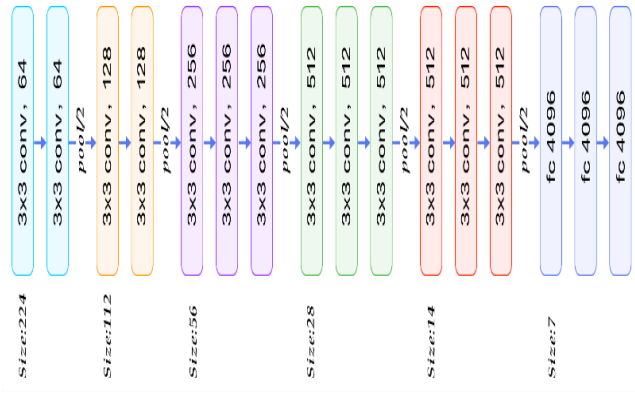


Figure.13. VGG16: For feature reduction used, the Convolutional Layer, pooling layer used, we delete from the VGG13 the fully

c) Evaluation of iCOSVM on MRI Images

The third step in the classification process is the training and testing of the iOSVM method. This method mainly uses target data collected during the training stage. However, it also uses the data from the testing step. Due to the imbalance problem in medical image datasets, it is difficult to identify the rare health care events that happen in the MRI images. This is why it is important that classification methods are able to identify these outliers. One-class classification is a promising technique to improve the efficiency of classification by learning a model from a single sample. As COSVM has its cons therefore, we discuss the advantages of implementing an incremental covariance-guided support vector machine for building a novel incremental detecting fake MRI tumor system. Usually, this type of modeling uses feature mapping or fitting to enforce the learning process. Unfortunately, deep learning techniques are not widely used for medical images due to the complexity of the features. In this paper, we present a novel method that enables deep learning models to learn single-class-relevant imaging features. The paper presents a method that combines the effects of perturbing operations and feature learning to improve the efficiency of deep learning models in capturing complex imaging features. In this part of the experiment the k-fold cross-validation technique (k = 10) has been used. The purpose of using the k-fold cross-validation method is to evaluate the performance of the iOSVM method, also to assure that the collected output is not biased or coincidental, and to generalize the classification of hidden data and finally to minimize the overfitting issues. An overfitting problem is when a proposed method's performance on the trained dataset exceeds that of a tested one. It can also mean that the training data gets better while the performance of the test gets worse. The k-fold cross-validation method is used to train and test the model's performance on different sets of training data. It can also estimate the model's performance on unseen data. We have created 10 training and testing subsets in our real MRI database using leave-one-out (LOO) cross-validation the LOO cross-validation method. The LOO method takes into account the subset that's out of the training step and then tests it in the testing step. This experiment, each training set consists of 90% of the target class. However, the remaining 10% of the target class was not included in the training set and was added to the testing data. This strategy was repeated 10 times to create different training and testing sets. The final results of each fold are the averages of the various evaluation measures. We then continuously change the parameters of the classifier to find the optimal fit for the given testing subset. This ensures that the classification accuracy is maintained. A comparative evaluation of the iCOSVM to relevant state-of-the-art SVM-based incremental models was performed to compare their performance in detecting fake MRI images. In the second set of experiments, incremental

one-class classifiers were compared to the bath SVM to show the superiority of the latter. Receiver Operating Characteristic (ROC) curves are used to measure the performance of the system. In this way, they represent a graphical representation of the variable's true-positive rate and its false-negative rate. The Area Under the ROC curve (AUC) is also computed by taking into account the variation of the true-positive and the false-negative rates. Moreover, we have taken into account the training time (in seconds) of the SVM-based classifiers.

5. RESULTS AND DISCUSSION

In this section, we discuss the accuracy of detecting fake tumor detection in real-world scenarios. The AUC average values for GAN's No. 3 using the incremental one class classifiers (iMOSVM, iCOSVM, iSVDD and iOSVM) are shown in Table 3 to 7. They were obtained from the training and testing datasets. In order to minimize the overfitting problem, we used a cross-validation technique to test the different parameters of each GAN. This method led to the highest accuracy of testing of our datasets used (MRI Dataset). The AUC average values for training and testing datasets are shown in Table 3 - 7. They correspond to the highest average accuracy values for both tests and training. One-class classification is preferred over batch one-class classification when it comes to handling the problem of unbalanced data. On the other hand, the difference between batch and incremental classification can be noticed in the Tables. It is clearly shows from the result that the iCOSVM, iMOSVM, and iOSVM and iSVDD are some of the leading one-class classification platforms that provide large AUCs. The ability to train and update new classification rules with large amounts of data is expected to be very useful for various applications. Since the iCOSVM does not require access to the previous data, it can easily perform new tasks without requiring relearning the training data. It is also faster than iSVDD and iMOSVM. In this study, we propose an online fake detection system (Fig. 4), first of all a dataset collection step performed, which aims to collect a dataset containing Brain MRI real tumor images. Then a framework for generating brain MRI images based on GAN is conducted as a second procedure and generate a fake tumor, then the MRI images were input to a pretrained VGG-16 deep transfer learning models, CNN, and each image is represented by deep feature vector which extracted using feed forward through VGG16 then features were selected as output values of the last convolutional layer and extracted as characteristic features. Subsequently, the selected image deep features, and preoperative and intraoperative parameters were given to a plurality of COSVM algorithms to find fake images. The framework is shown in Fig. 4.

Table 3. Shows the time and variable of different one-class classifiers. True Positive Rate (TPR), False Positive Rate (FPR) and Area Under the receiver operating characteristic Curve (AUC) MRI dataset - GAN 6

	iCOSVM	iMOSVM	iSVDD	iOSVM
AUC (Test)	99.06	98.50	97.14	96.64
Std (AUC test)	0.0217	0.0261	0.0516	0.0729
AUC (Train)	100	100	100	100
Std (AUC train)	0	0	0	0
Time (s) (in seconds)	2.23	2.46	4.57	2.29
FPR(%)	0	0	0	0
TPR(%)	100	92.5	0	52.5%

Table 4. Shows the time and variable of different one-class classifiers. True Positive Rate (TPR), False Positive Rate (FPR) and Area Under the receiver operating characteristic Curve (AUC) MRI dataset GAN 7

	iCOSVM	iMOSVM	iSVDD	iOSVM
AUC (Test)	99.25	98.88	98.07	98.63
Std (AUC test)	0.016	0.019	0.032	0.023
AUC (Train)	100	100	100	100
Std (AUC train)	0	0	0	0
Time (s) (in seconds)	2.74	2.8	4.75	2.53
FPR (%)	0	0	0	0
TPR (%)	100	95	0	85

Table 5. Shows the time and variable of different one-class classifiers. True Positive Rate (TPR), False Positive Rate (FPR) and Area Under the receiver operating characteristic Curve (AUC) MRI dataset GAN 8

	iCOSVM	iMOSVM	iSVDD	iOSVM
AUC (Test)	99	99.25	98.16	98.38
Std (AUC test)	0.0217	0.0168	0.0404	0.0281
AUC (Train)	100%	100%	100%	100%
Std (AUC train)	0	0	0	0
Time (s) (in seconds)	2.84	2.92	5.04	2.71
FPR	0	0	0	0
TPR	100	95	0	92.5

Table 6 Shows the time and variable of different one-class classifiers. True Positive Rate (TPR), False Positive Rate (FPR) and Area Under the receiver operating characteristic Curve (AUC) MRI dataset - GAN 9

	iCOSVM	iMOSVM	iSVDD	iOSVM
AUC (Test)	98.51	96.89	96.20	95.28
Std (AUC test)	0.0320	0.0547	0.0673	0.0774

AUC (Train)	100	100	100	100
Std (AUC train)	0	0	0	0.
Time (s) (in seconds)	1.99	2.15	3.96	2.05
FPR	0	0	0	0
TPR	100	90	0	85

Table 7. Shows the time and variable of different one-class classifiers. True Positive Rate (TPR), False Positive Rate (FPR) and Area Under the receiver operating characteristic Curve (AUC)
MRI dataset - GAN 10

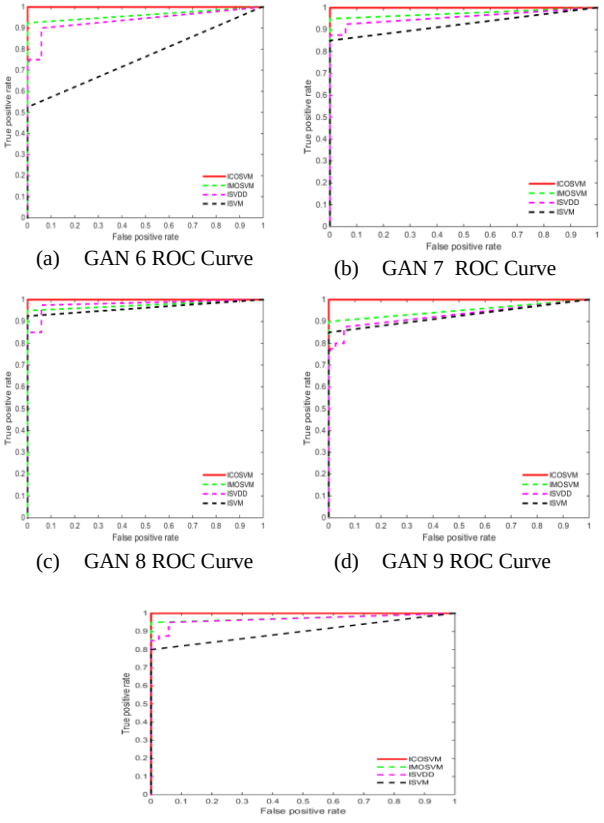
	iCOSVM	**iMOSVM**	**iSVDD**	**iOSVM**
AUC (Test)	98.75	98.01	97.44	97.27
Std (AUC test)	0.0315	0.0367	0.0493	0.0460
AUC (Train)	100	100	100	100
Std (AUC train)	0	0	0	0
Time (s) (in seconds)	2.07	2.18	3.86	2.09
FPR	0	0	0	0
TPR	100	95	0	80

Moreover, the performance of the iOSVM obtained less efficiency than other two classification tools: the iCOSVM and the iMOSVM. This is because the radius for the hypersphere-SVM (HS-SVM) cannot be flexibly chosen. The iMOSVM outperforms the iOSVM when it comes to performing discriminative distance classification. This is because the latter uses a more discriminative distance, which is called the Mahalanobis distance. In this paper, we present a set of graphical results for the dataset, which are based on the ROC curves of the MRI database. At as a final step, the paper shows that incremental learning is superior to batch methods when it comes to detecting fake MRI images. In terms of absolute performance, the results of our study show that the incremental SVM-based one-class classifiers perform better than the batch methods. Table 6 exhibit the result by examining the accuracy of GAN 7 when compared the four incremental classifiers (iMOSVM, iCOSVM, iSVDD, and iOSVM) against the batch SVM model. Finally, the incremental learning outperform the batch learning in which that incremental learning methodologies provide cleaner solutions as they estimate support vectors incrementally.

Table 8. Shows the time and variable of different one-class classifiers. True Positive Rate (TPR), False Positive Rate (FPR) and Area Under the receiver operating characteristic Curve (AUC)
MRI dataset GAN 7

	iCOSVM	**iMOSVM**	**iSVDD**	**iOSVM**	**SVM batch**
AUC (Test)	99.25	98.88	98.07	98.63	85.25
Std (AUC test)	0.016	0.019	0.032	0.023	0.0577
Time (s) (in seconds)	2.74	2.8	4.75	2.53	1.18

To conclude this experiment, the ROC curves [53] of various one-class classifiers are shown in Fig. 14. The rule-of-thumb when it comes to assessing the performance of a classifier is that the best one has the largest area under curve. Comparatively to the iMOSVM, iOSVM and iSVDD methods, the iCOSVM consistently leads to the best ROC curves. We are aware that according to Table 8 that SVM batch has provided lower running time than the incremental model. This was expected as the estimated training times are of the classifiers on the whole dataset. However, as new observation is added to the dataset the SVM batch would provide much higher training time since it should be re-trained on the previous training data from scratch.



(a) GAN 6 ROC Curve (b) GAN 7 ROC Curve

(c) GAN 8 ROC Curve (d) GAN 9 ROC Curve

(e) GAN 10 ROC Curve

Figure. 14. ROC Curve of the 4 classifiers for the two fold from the MRI Fake images

6. CONCLUSION

In this paper, we presented an overview of the incremental learning algorithm in change detection of medical images more precisely the iCOSVM 's capabilities to build a novel system detecting fake MRI images and provides cleaner solutions as it estimate support vector

incrementally. The system is composed of four main parts. The first part of the system is to create a fake images using generative adversarial network model the second part of the system is focused on feature selection model using the most capable deep learning model that is the CNN-VGG16. The third part of the system we emphasized on the use of incremental learning as most of the healthcare system nowadays exchanging thee data using online system. The last part of the system there were a need to compare our proposed method iCOSVM method with others incremental one class classifications algorithms, to prove of the highest accuracy we gain from our proposed method. With the use of incremental learning in securing the medical images field, (iCOSVM) and as per the results of the studies have shown that the system can effectively and efficiently solve the medical deepfake images problem. Moreover, the results of the studies have shown that the incremental one-class classification performed significantly better than the batch classification. This method allows our proposed system to perform both the detection tasks with high accuracy. One-class classification has also been shown to perform better when dealing with the issue of unbalanced data. For future work, it is notable that the suggested scheme can be applied to other classification circumstances, as considering distinct Medical imaginary databases. Likewise, other networks may be utilized, allowing this method to be adapted to unsupervised classification system.

References

[1] A. V. Sokolova and T. I. Buldakova, "Security of the telemedicine system information infrastructure," CEUR Workshop Proc., vol. 3035, pp. 183–192, 2021.

[2] S. Solaiyappan and Y. Wen, "Machine learning based medical image deepfake detection: A comparative study," Mach. Learn. with Appl., vol. 8, p. 100298, 2022, doi: 10.1016/j.mlwa.2022.100298.

[3] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Radiomic features for medical images tamper detection by equivalence checking," Procedia Comput. Sci., vol. 159, pp. 1795–1802, 2019, doi: 10.1016/j.procs.2019.09.351.

[4] C. Zhang, X. Xiao, and C. Wu, "Medical fraud and abuse detection system based on machine learning," Int. J. Environ. Res. Public Health, vol. 17, no. 19, pp. 1–11, 2020, doi: 10.3390/ijerph17197265.

[5] T. D. Gadhiya, A. K. Roy, S. K. Mitra, and V. Mall, "Use of discrete wavelet transform method for detection and localization of tampering in a digital medical image," *TENSYMP 2017 - IEEE Int. Symp. Technol. Smart Cities*, 2017, doi: 10.1109/TENCONSpring.2017.8070082.

[6] M. Pravin Savaridass, R. Deepika, R. Aarnika, V. Maniraj, P. Gokilanandhi, and K. Kowsika, "Digital Watermarking For Medical Images Using Dwt And Svd Technique," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1084, no. 1, p. 012034, 2021, doi: 10.1088/1757-899x/1084/1/012034.

[7] I. Assini, A. Badri, K. S. A. Sahel, and A. Baghdad, "A robust hybrid watermarking technique for securing medical image," *Int. J. Intell. Eng. Syst.*, vol. 11, no. 3, pp. 169–176, 2018, doi: 10.22266/IJIES2018.0630.18.

[8] J. Shahid, R. Ahmad, A. K. Kiani, T. Ahmad, S. Saeed, and A. M. Almuhaideb, "Data Protection and Privacy of the Internet of Healthcare Things (IoHTs)," *Appl. Sci.*, vol. 12, no. 4, 2022, doi: 10.3390/app12041927.

[9] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F. Y. Wang, "Generative adversarial networks: Introduction and outlook," *IEEE/CAA J. Autom. Sin.*, vol. 4, no. 4, pp. 588–598, 2017, doi: 10.1109/JAS.2017.7510583.

[10] M. Levy, G. Amit, Y. Elovici, and Y. Mirsky, "The Security of Deep Learning Defences for Medical Imaging," 2022, [Online]. Available: http://arxiv.org/abs/2201.08661.

[11] A. Shahsavari, S. Ranjbari, and T. Khatibi, "Proposing a novel Cascade Ensemble Super Resolution Generative Adversarial Network (CESR-GAN) method for the reconstruction of super-resolution skin lesion images," *Informatics Med. Unlocked*, vol. 24, no. June, p. 100628, 2021, doi: 10.1016/j.imu.2021.100628.

[12] Y. Liu, L. Meng, and J. Zhong, "MAGAN: Mask Attention Generative Adversarial Network for Liver Tumor CT Image Synthesis," *J. Healthc. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6675259.

[13] F. Marra, C. Saltori, G. Boato, and L. Verdoliva, "Incremental learning for the detection and classification of GAN-generated images," *2019 IEEE Int. Work. Inf. Forensics Secur. WIFS 2019*, pp. 13–18, 2019, doi: 10.1109/WIFS47025.2019.9035099.

[14] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[15] H. Jeon, Y. Bang, J. Kim, and S. S. Woo, "T-GD: Transferable GAN-generated Images Detection Framework," *37th Int. Conf. Mach. Learn. ICML 2020*, vol. PartF16814, no. August, pp. 4696–4711, 2020.

[16] S. Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-Generated Images Are Surprisingly Easy to Spot.. For Now," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 8692–8701, 2020, doi: 10.1109/CVPR42600.2020.00872.

[17] Y. He, N. Yu, M. Keuper, and M. Fritz, "Beyond the Spectrum: Detecting Deepfakes via Re-Synthesis," pp. 2534–2541, 2021, doi: 10.24963/ijcai.2021/349.

[18] M. Sethi, S. Ahuja, S. Rani, P. Bawa, and A. Zaguia, "Classification of Alzheimer's Disease Using Gaussian-Based Bayesian Parameter Optimization for Deep Convolutional LSTM Network," *Comput. Math. Methods Med.*, vol. 2021, 2021, doi: 10.1155/2021/4186666.

[19] X. Zhang, S. Karaman, and S. F. Chang, "Detecting and Simulating Artifacts in GAN Fake Images," *2019 IEEE Int. Work. Inf. Forensics Secur. WIFS 2019*, 2019, doi: 10.1109/WIFS47025.2019.9035107.

[20] T. De, D. Université, and U. D. U. Paris-saclay, "Enabling real-world EEG applications with deep learning Apprentissage profond pour la mise en application de l ' EEG en conditions réelles Thèse de doctorat," 2022.

[21] T. Yang, Z. Huang, J. Cao, L. Li, and X. Li, "Deepfake Network Architecture Attribution," 2022, [Online]. Available: http://arxiv.org/abs/2202.13843.

[22] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, "CT-GAN: Malicious tampering of 3D medical imagery using deep learning," *Proc. 28th USENIX Secur. Symp.*, pp. 461–478, 2019.

[23] V. Sorin, Y. Barash, E. Konen, and E. Klang, "Creating Artificial Images for Radiology Applications Using Generative Adversarial Networks (GANs) – A Systematic Review," *Acad. Radiol.*, vol. 27, no. 8, pp. 1175–1185, 2020, doi: 10.1016/j.acra.2019.12.024.

[24] L. C. Chu, A. Anandkumar, H. C. Shin, and E. K. Fishman, "The Potential Dangers of Artificial Intelligence for Radiology

and Radiologists," *J. Am. Coll. Radiol.*, vol. 17, no. 10, pp. 1309–1311, 2020, doi: 10.1016/j.jacr.2020.04.010.

[25] S. Gupta and B. B. Gupta, "Cross-Site Scripting (XSS) attacks and defense mechanisms: classification and state-of-the-art," *Int. J. Syst. Assur. Eng. Manag.*, vol. 8, pp. 512–530, 2017, doi: 10.1007/s13198-015-0376-0.

[26] K. Choudhary *et al.*, "Recent advances and applications of deep learning methods in materials science," *npj Comput. Mater.*, vol. 8, no. 1, 2022, doi: 10.1038/s41524-022-00734-6.

[27] S. Minaee and A. Abdolrashidi, "Iris-GAN: Learning to Generate Realistic Iris Images Using Convolutional GAN," 2018, [Online]. Available: http://arxiv.org/abs/1812.04822.

[28] A. Cheng, "PAC-GAN: Packet Generation of Network Traffic using Generative Adversarial Networks," *2019 IEEE 10th Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2019*, pp. 728–734, 2019, doi: 10.1109/IEMCON.2019.8936224.

[29] I. Rosenberg, A. Shabtai, Y. Elovici, and L. Rokach, "Adversarial Machine Learning Attacks and Defense Methods in the Cyber Security Domain," *ACM Comput. Surv.*, vol. 54, no. 5, 2021, doi: 10.1145/3453158.

[30] I. Semanjski, *The Seventh International Conference on Data Analytics*. 2018.

[31] N. Pitropakis, E. Panaousis, T. Giannetsos, E. Anastasiadis, and G. Loukas, "A taxonomy and survey of attacks against machine learning," *Comput. Sci. Rev.*, vol. 34, p. 100199, 2019, doi: 10.1016/j.cosrev.2019.100199.

[32] J. Yan, Y. Qi, and Q. Rao, "Detecting Malware with an Ensemble Method Based on Deep Neural Network," *Secur. Commun. Networks*, vol. 2018, 2018, doi: 10.1155/2018/7247095.

[33] K. He and D. S. Kim, "Malware detection with malware images using deep learning techniques," *Proc. - 2019 18th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. IEEE Int. Conf. Big Data Sci. Eng. Trust. 2019*, pp. 95–102, 2019, doi: 10.1109/TrustCom/BigDataSE.2019.00022.

[34] L. Gonog and Y. Zhou, "9-ICIEA.2019.8833686.pdf," *2019 14th IEEE Conf. Ind. Electron. Appl.*, pp. 505–510, 2019.

[35] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, *A survey of the recent architectures of deep convolutional neural networks*, vol. 53, no. 8. Springer Netherlands, 2020.

[36] T. M. Mohammed, L. Nataraj, S. Chikkagoudar, S. Chandrasekaran, and B. S. Manjunath, "Malware detection using frequency domain-based image visualization and deep learning," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 2020-Janua, pp. 7132–7141, 2021, doi: 10.24251/hicss.2021.858.

[37] A. Bensaoud, N. Abudawaood, and J. Kalita, "Classifying Malware Images with Convolutional Neural Network Models," 2020, doi: 10.6633/IJNS.202011_22(6).17.

[38] R. H. Hwang, M. C. Peng, V. L. Nguyen, and Y. L. Chang, "An LSTM-based deep learning approach for classifying malicious traffic at the packet level," *Appl. Sci.*, vol. 9, no. 16, 2019, doi: 10.3390/app9163414.

[39] N. Ullah *et al.*, "An Effective Approach to Detect and Identify Brain Tumors Using Transfer Learning," *Appl. Sci.*, vol. 12, no. 11, p. 5645, 2022, doi: 10.3390/app12115645.

[40] X. Deng, W. Li, X. Liu, Q. Guo, and S. Newsam, "One-class remote sensing classification: One-class vs. Binary classifiers," *Int. J. Remote Sens.*, vol. 39, no. 6, pp. 1890–1910, 2018, doi: 10.1080/01431161.2017.1416697.

[41] S. S. Khan and M. G. Madden, "One-class classification: Taxonomy of study and review of techniques," *Knowl. Eng. Rev.*, vol. 29, no. 3, pp. 345–374, 2014, doi: 10.1017/S026988891300043X.

[42] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, no. xxxx, 2020, doi: 10.1016/j.neucom.2019.10.118.

[43] G. Jaiswal, "Performance analysis of incremental learning strategy in image classification," *Proc. Conflu. 2021 11th Int. Conf. Cloud Comput. Data Sci. Eng.*, no. January, pp. 427–432, 2021, doi: 10.1109/Confluence51648.2021.9377034.

[44] A. Gepperth and B. Hammer, "Incremental learning algorithms and applications," *ESANN 2016 - 24th Eur. Symp. Artif. Neural Networks*, pp. 357–368, 2016.

[45] S. S. Sarwar, A. Ankit, and K. Roy, "Incremental Learning in Deep Convolutional Neural Networks Using Partial Network Sharing," *IEEE Access*, vol. 8, pp. 4615–4628, 2020, doi: 10.1109/ACCESS.2019.2963056.

[46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[47] T. Kefi, R. Ksantini, M. B. Kaâniche, and A. Bouhoula, "A novel incremental covariance-guided one-class support vector machine," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9852 LNAI, pp. 17–32, 2016, doi: 10.1007/978-3-319-46227-1_2.

[48] M. F. Alanazi *et al.*, "Brain Tumor/Mass Classification Framework Using Magnetic-Resonance-Imaging-Based Isolated and Developed Transfer Deep-Learning Model," *Sensors*, vol. 22, no. 1, 2022, doi: 10.3390/s22010372.

[49] D. N. Louis *et al.*, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathol.*, vol. 131, no. 6, pp. 803–820, 2016, doi: 10.1007/s00401-016-1545-1.

[50] A. Mehmood *et al.*, "A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images," *Neuroscience*, vol. 460, pp. 43–52, 2021, doi: 10.1016/j.neuroscience.2021.01.002.

[51] H. A. Khan, W. Jue, M. Mushtaq, and M. U. Mushtaq, "Brain tumor classification in MRI image using convolutional neural network," *Math. Biosci. Eng.*, vol. 17, no. 5, pp. 6203–6216, 2020, doi: 10.3934/MBE.2020328.

[52] S. Ghosh, A. Chaki, and K. Santosh, "Improved U-Net architecture with VGG-16 for brain tumor segmentation," *Phys. Eng. Sci. Med.*, vol. 44, no. 3, pp. 703–712, 2021, doi: 10.1007/s13246-021-01019-w.

[53] T. Yang, "Deep AUC Maximization for Medical Image Classification: Challenges and Opportunities," pp. 1–7, 2021, [Online]. Available: http://arxiv.org/abs/2111.02400.

Fadheela Hussain, a PhD Candidate in Computing and Information Sciences, M.Sc. in information technology, both received from University of Bahrain , currently a part time lecture in University of Bahrain (UOB), University of technology Bahrain (UTB) and a full time senior health specialist in ministry of health - MOH, in information and health planning directorate.

Riadh Ksantini received the M.Sc. and Ph.D. degrees in Computer Science from the Université de Sherbrooke, Sherbrooke, QC, Canada, in 2003 and 2007, respectively. Presently, he is Associate Professor at the Department of Computer Science, College of IT, University of Bahrain,

Adjunct Associate Professor at the School of Computer Science, within the Faculty of Science of the University of Windsor, Windsor, Ontario, Canada, and Adjunct Professor at the Department of Computer Science, Université du Québec à Montréal (UQAM). He has also served as Visiting Fellow Research Scientist at the Canadian Space agency. In 2008, he was awarded a fellowship (of excellence) for postdoctoral research from the granting agency ¨Fonds quebecois de la recherche sur la nature et les technologies ¨ (FQRNT). His PhD was evaluated and ranked third Ph.D. in Quebec for 2007 by the committee of Information Technology and Communications. His research interests include Artificial Intelligence, Machine/Deep Learning, Pattern Recognition and Computer Vision.

Dr. Khaoula TBARKI received her PhD degree in Electrical Engineering from National School of Engineers of Tunis/Tunisia (ENIT), 2018. She is currently an Assistant professor of Artificial Intelligence at Private Higher School of Technology and Engineering (Tek-Up University). Her research interests include image/signal processing, computer vision, machine learning, deep learning, landmine detection and pattern recognition, ground penetrating radar (GPR).