Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 161 (2019) 475–483

Procedia Computer Science

www.elsevier.com/locate/procedia

The Fifth Information Systems International Conference 2019

Offline Signature Verification using Deep Learning Convolutional Neural Network (CNN) Architectures GoogLeNet Inception-v1 and Inception-v3

Jahandad[a], Suriani Mohd Sam[a,*], Kamilia Kamardin[b], Nilam Nur Amir Sjarif[b], Norliza Mohamed[a]

[a]*Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia*
[b]*Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia*

Abstract

Biometric systems such as signature verification are highly viable in order to identify individuals in organizations or in finance divisions. Advancement in classification of images using deep learning networks has opened an opportunity for this problem. In this study, the largest available handwritten signature dataset, namely, the GPDS Synthetic Signature Database, was employed for the classification of signatures of 1000 users, each of which having 24 original (or genuine) signatures, and 30 forged (or fake) signatures. Moreover, two popular GoogLeNet architecture versions of CNN, namely, Inception-v1 and Inception-v3, were used. Firstly, algorithms were trained on samples from 20 users, and achieved a validation accuracy of 83% for Inception-v1 and 75% for Inception-v3. In terms of Equal Error Rates (EER), Inception-v1 managed to obtain an EER as low as 17 for 20 users; while EER for Inception-v3 with 20 users obtained 24, which is a good measure compared to prior works in the literature. Although Inception-v3 has performed better in the ImageNet image classification challenge, in the case of 2D images of signatures, Inception-v1 has performed the classification task better than Inception-v3 It is also acknowledged in this study that Inception-v1 spent less time training, as it had a lower number of operations compared to Inception-v3.

* Corresponding author. Tel.: +60-03-261-547-04.
E-mail address: suriani.kl@utm.my

1877-0509
10.1016/j.procs.2019.11.147

Keywords: Convolutional Neural Networks; CNN; GPDS Synthetic Signature Database; Inception-v1; Inception-v3; GoogLeNet

1. Introduction and background

To identify personnel in an organization is critical and organizations depend on biometric systems for verification of individuals. These biometric systems include fingerprint, iris scanning and signature verification, among them signature verification remains socially and legally accepted for verification. Signature verification is the process of authenticating or identifying individuals based on their handwritten signatures as handwriting differs from person to person [1]. The process of signature verification starts from taking handwritten signatures from individuals on a paper and then scanning it to save it in digital form. The next time when same individual needs to be verified, the signature is taken again and compared with previously saved digital image of signature. At time of taking first signature for database, the velocity, position or pressure is not measured thus this dynamic information is not used in further verification process that's why it is called as offline signature verification process, and the non-dynamic verification process becomes challenging [2].

Many methods have been introduced and successfully applied for making decision whether a signature is genuine or forged. In reference [3] Dynamic time warping (DTW), an algorithm for measuring similarity between two temporal sequences is used to verify between genuine and forged handwritten signatures. Due to the capabilities of self-learning of properties Neural Networks (NNs) have also been used for verification of offline signatures. In reference [4] a neural network is designed and proposed for verification of handwritten signatures. In [5] researchers have applied Wavelet Based Approach for classification of Persian handwritten signatures. Structural approach is another method in which signatures are represented using trees and graphs. In reference [6] a structural approach is proposed based on graph matching cross validation for classification of handwritten signatures. Support Vector Machines (SVM), a machine learning algorithm used for classification and regression problems has also been proposed and used for classification of handwritten signatures. In [7] researchers have successfully applied SVM for classification of offline signature.

In machine learning, a Convolutional Neural Network (CNN) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multilayer perceptron designed to require minimal preprocessing [8]. These biologically inspired computational models are able to far exceed the performance of previous forms of artificial intelligence in common machine learning tasks.

One of the large-scale problems is named Large Scale Visual Recognition Challenge (LSVRC). Deep Learning Convolution Neural Networks (CNN) based techniques show state-of-the-art accuracy on the ImageNet task [8]. Russakovsky et al. recently published a paper on the ImageNet dataset and the state-of-the-art accuracies achieved during last few years [9]. The paper shows the story of deep learning techniques overtime on this challenge from 2012. ResNet-152 shows only 3.57% error, which is better than human error for this task at 5%.

In this paper, our aim is to use GoogLeNet Inception-v1 and Inception-v3 to classify genuine and skilled forged signatures of 1000 users. Each user having 24 genuine and 30 forged signatures. This paper is organized as follows: first it states the problem, then discuss the details of dataset that it must evaluate. Then it describes the procedure of selected architectures and finally discusses the obtained results.

The problem of automatic handwritten signature verification is commonly modelled as a verification task: we have a learning set of genuine signatures (belonging to the individual) from 4000 users, each having 24 signatures to train the two models, Inception-v1 and inception v3. The models are than trained for skilled forged signatures (created by others) for 1000 users, for each user we have 30 forged signatures. Both models must classify between genuine and skilled forged signatures. To evaluate the classification performance of the system, Equal Error Rate (EER), Receiver Operating Characteristic Curve (ROC curve), Area Under the ROC Curve (AUC), validation loss and validation accuracy are measured.

2. Dataset

Most signature verification researched have been conducted on private collection of signatures. But there are few commonly used standard signature datasets that are used to compare results between different approaches. In this research GPDS Synthetic Signature Database is acquired from Grupo de Procesado Digital de Señales GPDS, University of Velancia Spain. GPDS Synthetic Signature Database contains data from 4000 synthetic individuals: 24 genuine signatures for each individual, plus 30 forgeries of his/her signature. All the signatures were generated with different modeled pens. The signatures are in "jpg" format and equivalent resolution of 600 dpi [10].

3. Selected algorithms

Deep Convolutional Neural Network architecture GoogLeNet with its two different versions codenamed Inception-v1 and Inception-v3 for classification of GPDS Synthetic Signature Database is used in this work and as in Table 1. GoogLeNet, which is the winner of the ILSVRC (ImageNet Large Scale Visual Recognition Competition) 2014, an image classification competition, which has significant improvement over ZFNet (The winner in 2013) and AlexNet (The winner in 2012) and has relatively lower error rate compared with the VGGNet (1st runner-up in 2014) [11].

Table 1: Comparison table of Inception-v1 and Inception-v3.

Parameters/Architectures	Inception-v1	Inception-v3
Architecture Details	Winner of ILSVRC 2014	1st Runner-Up of ILSVRC 2015
Number of Operations	4 Million operations	12 Million operations
Convolution Filter Size	1x1, 3x3 and 5x5 [12]	5 x5 filter size is replaced with two 3 x 3 filters [12]
Pooling	In GoogLeNet, global average pooling is used nearly at the end of the network by averaging each feature map from 7X7 to 1X1. Research shows that Global Average Pooling improves top-1 accuracy by 0.6% then fully connected layer [14]	In inception-v3, an efficient grid size reduction is proposed, feature maps are done by convolutions and by pooling separately and in last both set of feature maps are concatenated and forwarded to the next inception module [12].
Layers	22	42
Feature Map Downsizing	Max pooling	Efficient Grid Size Reduction
Auxiliary Classifier	2	1
Factorizing convolutions	Not applied	Filters factorized
Top 1 accuracy ImageNet	74.8%	78.8%
Inception Module	Naïve Inception module [12]. In Inception-v1 1X1convolution, 3X3 convolution, 5X5 convolution, and 3X3 max pooling are all done altogether as shown in Fig. 1.	Inception module with factorization of n x n convolutions [12]. In Inception-v3 as convolutions are factorized to smaller and into asymmetric convolutions a new inception module is formed shown in Fig. 2.
Overall Architectures	With 22 layers in total, it is very deep model compared with previous AlexNet, ZFNet and VGGNet [17].	With 42 layers deep, the computation cost is only about 2.5 higher than that of Inception-v1, and much more efficient than that of VGGNet [17].

The network architectures of GoogLeNet versions are different from VGGNet, ZFNet and AlexNet. The techniques used by GoogLeNet is called inception module, which means to have different sizes of convolutions for the same input and staking all the outputs. GoogLeNet architecture, winning in ILSVRC 2014 used 12 times fewer parameters than the previous winning architectures [12] while being more accurate. GoogLeNet goes deep in the sense that new level is introduced called inception module which is actually an increase in network depth.

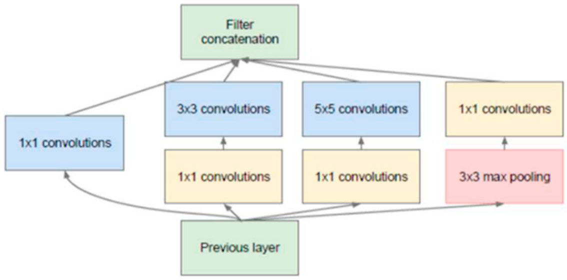

Fig. 1. Inception module with dimensionality reduction [12].

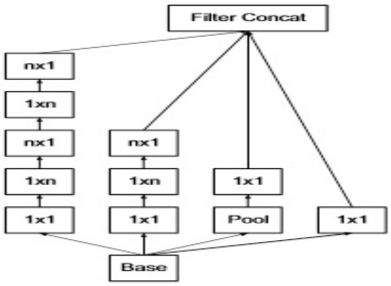

Fig. 2. Inception modules after the factorization of the n × n convolutions [12].

4. Training methodology

The architectures were coded and trained using TensorFlow, Google's Deep Learning System. Models were imported from keras, a high-level neural networks API capable of running on top of TensorFlow. Everything was performed on a local machine in Jupyter Notebook, an open-source web application that allows to create code, equations, visualization and narrative text, popularly used for machine learning projects.

In this work, Inception-v1 and Inception-v3 models are trained on GPDS Synthetic Signature Database on CPU with batch size 50. Epochs are set to 100 and early stopping callback is used to early stop the training if when accuracy percentage has stopped improving. ReduceLROnPlateau callback function is used to reduce learning rate when model has stopped improving. Dataset is divided with ration of 70 and 30, 70 used to train the model and 30 to test the results. Dropout of 0.25 is used with the sigmoid as the activation function. For early testing, models were trained and validated on subsamples of dataset such as for 10 users, 20 users, 100 users and 1000 users. But once all the 216,000 signatures of all 4000 users were used, the training characteristics of the architecture changed (accuracies, losses, speed of training). These changes are explained in experimental results and comparison section.

4.1. Signature classification system based on deep learning

This project refers to the design of an offline handwritten signature classification system that can classify genuine and skilled forged signatures of personnel in an organization. This project uses GPDS synthetic Signature Database to train two types of Convolutional Neural Architectures, Inception-v1 and Inception-v3. The networks used are callback and Reduce Learning Rate functions to fine tune network during the training and testing epochs. Finally, the graphs were generated based on the results obtained from both architectures. Proposed procedure of the system is summarized in Fig. 3 below. Fig. 4 and Fig. 5 show the inception history for 20-users data.

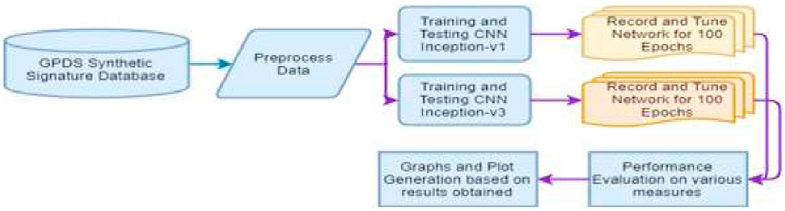

Fig. 3. Model diagram of proposed system.

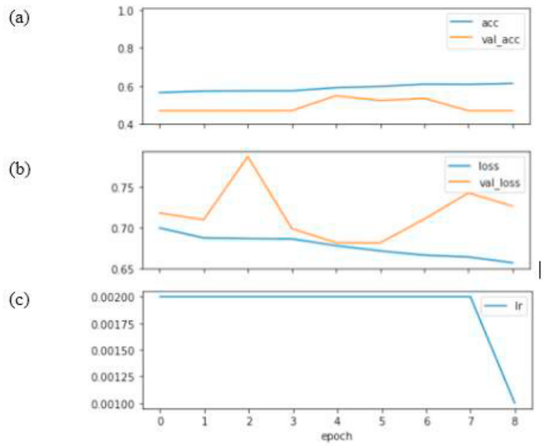

Fig. 4. Inception-v1 training histroy for 20 users data: (a) training and validation accuracies (b) training and validation loss (c) learning rates.

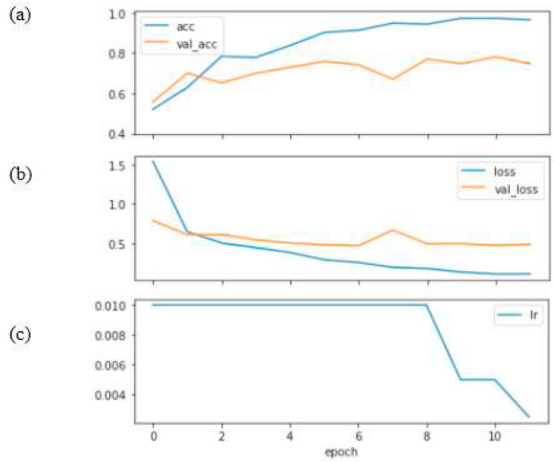

Fig. 5. Inception-v3 training histroy for 20 users data (a) training and validation accuracies (b) training and validation loss (c) learning rate.

4.2. Accuracy validation, EER and F1 score

Table 2 shows the experimental results validation accuracy, Equal Error Rate (EER) and F1 Score for the classification of GPDS Synthetic Signature Database using our proposed two CNN models; Inception-v1 and Inception-v3 as described in the previous section. Fig. 6 shows a comparison bar graph for performance of different models.

Table 2. Experimental Results with different configurations.

Network	No. of Users	No. of Samples for Training	Number of Samples for Testing	Equal Error Rate (EER) %	F1 Score	Validation Accuracy (%)
Inception-v1	10	370	170	37	0.7976	78
Inception-v3	10	370	170	41	0.7411	88
Inception-v1	20	730	330	17	0.8275	83
Inception-v3	20	730	330	24	0.7076	75
Inception-v1	100	3700	1700	22	0.7862	78
Inception-v3	100	3700	1700	26	0.70	72
Inception-v1	1000	36000	17000	22	0.7532	77
Inception-v3	1000	36000	17000	26	0.7266	73

Inception-v1 performed better in case of GPDS Synthetic Signature Database which contained a large dataset of 2-Dimensional images with resolution of 224X224X1. In ImageNet classification Inception-v3 achieved a higher accuracy percentage and lower EER but in this case, Inception-v1 outperformed Inception-v3 as Inception-v1 had higher validation accuracy and lower EER for all samples tested.

Fig. 6. Validation accuracy in percentage for Inception-v1 and Inception-v3 models with different number of users.

Comparing the results in Table 2, these two models have performed better than other results available in literature. As the same dataset is used [18] and achieved lowest 26.74% EER while this present study achieved as low as 26% EER. Not many publications include study on GPDS Synthetic Signature Database, so results achieved in this work can be used for future enhancements and comparisons.

Firstly, algorithms were trained on samples from 20 users and achieved validation accuracy of 83% for Inception-v1 and 75% for Inception-v3. Furthermore, both architectures with samples of 10, 100, and 1000 users were trained and evaluated achieving validation accuracy of 78%, 78% and 77% for 10, 100 and 1000 users in Inception-v1 and 88%, 72% and 73% for 10, 100 and 1000 users in Inception-v3 model. Although Inception-v3 performed better in ImageNet image classification challenge, but in the case of 2D images of signatures Inception-v1 performed the task better then Inception-v3. In terms of Equal Error Rates (EER), Inception-v1 managed to get lower EERs i.e. 37, 17, 22 and 22 for 10, 20, 100 and 1000 users respectively compared to Inception-v3 EERs 41, 24, 26 and 26 for 10, 20, 100 and 1000 user's data respectively. Architectural details of Inception-v3 focus on factorizing of convolutions into smaller sizes and grid size reduction for reducing computational cost. This project's experimental results on GPDS

Synthetic Signature Database of low-resolution images hint that Inception-v3 can outperform Inception-v1 of high resolution 3- dimensional images such as ImageNet. The authors of Inception-v3 claimed that models employing higher resolution receptive fields tend to result in significantly improved recognition performance. For already low-resolution input images Inception-v1 with 22 layers deep network performed better compared to 42 layers deep Inception-v3 network.

4.3. Area under the Receiver Operating Characteristic Curve (AUC ROC) from prediction scores

AUC ROC curve is a performance measure for classification problem at various thresholds settings. ROC is a probability curve and AUC represent degree or measure of separability. It tells how much model is capable of distinguishing between classes. Higher the value of AUC, the better will the model at predicting 0s as 0s and 1s as 1s [16]. Table 3 contains AUC scores showing the performance of classification models; Inception-v1 and Inception-v3 with different number of users and samples for both networks with different users.

Table 3. ROC AUC scores for Inception v1 and Inception-v3 with different number of samples.

Network	Inception-v1	Inception-v3	Inception-v1	Inception-v3	Inception-v1	Inception-v3	Inception-v1	Inception-v3
Number of Users	10	10	20	20	100	100	1000	1000
ROC AUC Score	0.87	0.83	0.91	0.83	0.86	0.81	0.86	0.82

ROC AUC scores for different number of users in the Table 3 showed the classification performance. In this work, Inception-v1 with samples of 20 users has total ROC AUC score of 0.91 which means it has performed well compared to other configurations. Given Fig. 7 and Fig. 8, ROC AUC Curve graphs are plotted for Inception-v1 with 20 users and Inception-v3 20 users. The curves plotted out two parameters; True Positive Rate (TPR) and False Positive Rate (FPR).

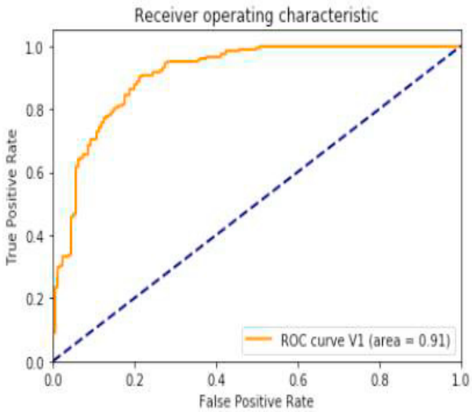

Fig. 7. ROC AUC Curve of Inception-v1 trained and tested samples of 20 users signatures.

In terms of AUC ROC score, Inception-v1 with 20 users achieved the highest score while Inception-v3 with 100 users was the lowest. The graph in Fig. 9 below summarizes the AUC ROC score value performance for each configuration.

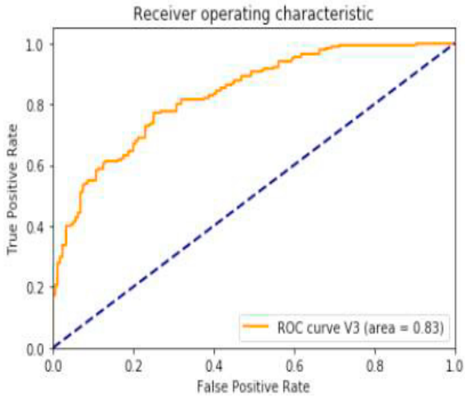

Fig. 8. ROC AUC Curve of Inception-v3 trained and tested on samples of 20 users signatures.

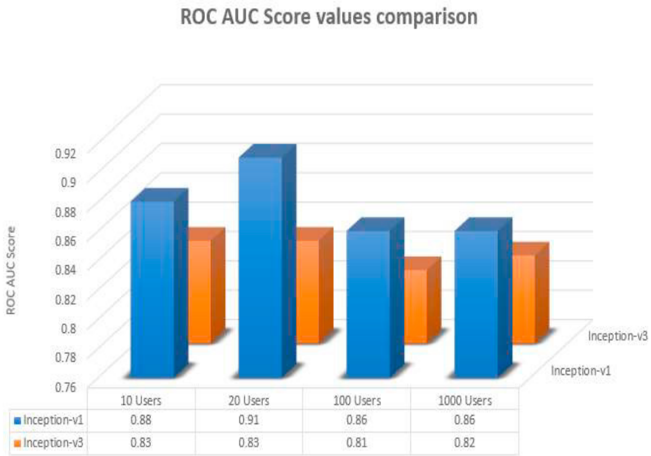

Fig. 9. AUC ROC scores comparison graphs.

5. Conclusions

This paper has worked on two Convolutional Neural Network Architectures; Inception-v1 and Inception-v3 and proved that these two algorithms can be successfully used to verify individuals in an organization using handwritten signatures. These two models have performed better than other results available in literature. As the same dataset is used in [19] and achieved lowest 26.74% EER while this present study achieved the lowest at 26% EER. Not many publications include study on GPDS Synthetic Signature Database, so results achieved in this project can be used for future enhancements and comparisons.

This project's experimental results on GPDS Synthetic Signature Database of low-resolution images hint that Inception-v3 can outperform Inception-v1 of high resolution 3-dimensional images such as ImageNet. The authors of Inception-v3 claimed that models employing higher resolution receptive fields tend to result in significantly improved recognition performance. For already low-resolution input images Inception-v1 with 22 layers deep network performed better compared to 42 layers deep Inception-v3 network.

Our Inception-v1 achieved higher performance then Inception-v3, however Inception-v3 is deep with 42 layers and has performed better in ImageNet challenge but in our case of 2D images of signatures Inception-v1 has outperformed Inception-v3. This hints that as Inception-v3 being deep and large network can lead to overfitting if the image data has not been optimistically pre-processed. GPDS Synthetic Signature Database has not been used in literature as dataset to test different neural network algorithms but now it is the largest publicly available signature database which can be used in future to test for different CNNs and compare with the results we have achieved in this research.

Future improvements guided by this work are to use regularization techniques to improve models. Furthermore, data augmentation might be added as pre-processing step to each batch before feeding it to the inception modules. Resolving these two issues might improve the overall performance and avoid overfitting of data. Finally, as future work more advanced convolutional neural networks (such as Inception v4 and Inception-ResNet can be used to acquire better results for the similar problem.

Acknowledgements

The authors fully acknowledge Ministry of Education (MOE) and Universiti Teknologi Malaysia for the research grant (R.K130000.7740.4J283) that help in funding the research works. Special thanks to the author of GPDS Synthetic Signature Database, Mr. Miguel A. Ferrerer from Grupo de Procesado Digital de Señales GPDS, University of Velancia Spain to providing us with this useful dataset.

References

[1] Taherzadeh, G, R Karimi, A Ghobadi, and HM Beh. (2011) "Evaluation of Online Signature Verification Features".
[2] Hafemann L.G., R. Sabourin, and L.S. Oliveira. (2017) "Learning Features for Offline Handwritten Signature Verification using Deep Convolutional Neural Networks".
[3] Inglis, S., and I.H. Witten. "Compression-based Template Matching", in *Proceedings of IEEE Data Compression Conference (DCC'94)*. pp. 106–115.
[4] Mccabe, A., J. Trevathan, and W. Read. (2008) "Neural Network-based Handwritten Signature Verification." **3 (8)**: 9–22.
[5] Ghandali, S., and M.E. Moghaddam. (2009) "Off-Line Persian Signature Identification and Verification Based On Image Registration and Fusion." *J. Multimedia* **4 (3)**: 137–144.
[6] Ramachandra A.C., J. Ravi, K.B. Raja, K.R. Venugopal, and L.M. Patnaik. (2009) "Signature Verification using Graph Matching and Cross Validation Principle." **1 (1)**.
[7] Kiani, V., R. Pourreza, and H.R. Pourreza. (2009) "Offline Signature Verification Using Local Random Transform and Support Vector Machines." *Int. J. Image Process* **3 (5)**: 184–194.
[8] He, K., X. Zhang, S. Ren, and J. Sun. (2016) "Deep Residual Learning for Image Recognition." *Cvpr*. pp. 1–9.
[9] Russakovsky, O., et al. (2014) "ImageNet Large Scale Visual Recognition Challenge". **Sep**.
[10] Ferrer, M.A., M. DIaz-Cabrera, and A. Morales. (2013) "Synthetic Off-line Signature Image Generation", in *6th IAPR International conference on Biometrics*. pp. 667–680.
[11] Alom, A., et al., "The History Began from AlexNet : A Comprehensive Survey on Deep Learning Approaches."
[12] Szegedy, C., S. Reed, P. Sermanet, V. Vanhoucke, and A. Rabinovich. "Going Deeper with Convolutions". pp. 1–12.
[13] Lin, Q., Q. Chen, and S. Yan. (2013) "Network In Network." 1–10.
[14] Tsang, S. (2018) "Review: GoogLeNet (Inception-v1) — Winner of ILSVRC 2014 (Image Classification)". [Accessed: 24-Sep-2018].
[15] Juvonen, Antti. (2018) "Overfitting and Dropout in Neural Networks". [Accessed: 24-Sep-2018].
[16] Kumar, Rajeev, and Indrayan, Abhaya. (2011) "Receiever Operating Characteristic (ROC) Curve for Medical Researchers," in *Indian Paediatrics* **48 (4)**: 277-287.
[17] Tsang, S. (2015) "Review: Inception-v3 — 1st Runner Up (Image Classification)" in *ILSVRC*. [Accessed: 26-Nov-2018].
[18] Zehua Zhang Yan Cui, X. L. (2017) "Multi-Phase Offline Signature Verification System Using Deep Convolutional Generative Adversarial Networks", in *9th International Symposium on Computational Intelligence and Design*, IEEE.