A Proposed Framework: Face Recognition With Deep Learning

Michael Farayola, Aman Dureja

Abstract : Face recognition is the capability to ascertain the identification of a person solitary or amidst multitudes of individuals. In lieu to this, deep learning has dominated and it has been used in recent years due to its momentous performance to solve the face recognition challenges using convolutional neural networks (CNN). It is a technology with enormous capabilities and diversities used in computer vison problems such as modelling and saliency detection, semantic segmentation, handwriting digital recognition, emotion recognition and many more. CNN architectures such has Alex Net, VGG are the practically known architectures that have immensely prompt new dataset for CNN model designs. This paper contributes to actualization of a propose CNN based on a pre-trained VGG Face for face recognition from set of faces tracked in video or image capture achieving a 97% accuracy. Also, implementing the use of metric learning to actualized a discriminative feature from our instances.

Keywords : Deep Learning, Face Recognition, Metric Learning, TensorFlow

1 INTRODUCTION

Face represents a major part of the human used for recognition which helps in the transfer of information or communication, interpreting and understanding feelings. Facial images have helped to ascertain lots of information such as age, emotion, race, gender and others [1,3,4,6,12,19,34]. All these information and recognition of a person are effortlessly processed by the human brain, yet a challenge posed in the area of computer vision notwithstanding the different utilizations of several biometrics' applications. In regards to this, variations like occlusion, pose, illumination, expression are common reasons why face recognition is still a challenge faced in the fields of Artificial Intelligence, Computer Vision and many more [13]. Majorly, there are two applications of Face recognition: (i) identification and (ii) verification. These applications procure the diverse ways images can be classified and identified through extraction of discriminative features out of the images. These applications serve as the pre-processing phase for dimensionality reduction, noise cleaning and information packing [35]. Face identification can be regarded as a process by which a face recognition system feedbacks the detecting and identifying of an individual in a facial image input while face verification relies on the footing of a computational work done by identification application process on the face image input. It therefore ensures face recognition system to validate the identification of an individual to be true or false. Hence, with the condition that all pattern matching is accurate with a specified and desired output, the system stores individual's facial qualities data in a database each time a person is tagged in an image. Hence, adequate information is gathered such that the data are utilized to differentiate a similar face in several image. It is recommended to label those photos with people's name or the person's name. Furthermore, face recognition has proliferated in computer vision over years. The initial effort was exerted in the year 1966. A time where Bledsoe illustrated the demands relating to automatic facial recognition tools [8]: This recognition problem is characterized with great variability in facial expression, head rotation and tilt, aging, lighting intensity and angle etc. Several attempts at facial recognition by machine have allowed for little or no variability in these quantities. Yet the method of pattern matching of unprocessed optical data, which is mostly used is certain to fail in cases where the variability is great. Especially when pattern matching result to be low between two pictures of the same person with two different head rotations. To researchers' surprise, face recognition still face the challenges Bledsoe notified in the year 1966 and which can be corresponded with the problem of pose-illumination-expression. Although, avail of immerse training set, loss functions and different architectures useful for deep CNN has propagated towards limiting the PIE posed problems in face recognition and hereby improving the human-like performance accuracy positively in classification and verifications [9]. There exist two main ways of training deep CNN for face recognition is through a multi class classifier training to figure out different identities in a training set and likewise introducing a triplet loss by training directly on an embedding [5].

2 DEEP LEARNING

Deep learning is a division of Artificial Intelligence. Machines learn to think and process on their own like human, thus learn through different computational models of several processing layers and data representation with lots of levels of abstraction [11]. Deep learning can be called differential programming. It has been used in different areas such as audio recognition, natural language processing, computer vision which are modern technologies. Face recognition is not an exception in this regard. A subcategory of deep learning is the convolutional neural network which will be further discussed. Deep learning is an algorithm of multi-layer used to extract features, identifying edges like letters, digits, faces. It is objectively a proposed technology to mimic the workings of the brain. The word "deep" refers to numbers of layers through which data is transformed and each layer transforms it input into an outline representation.

2.1 Convolutional Neural Network

Convolutional is a part of deep neural networks which has been introduced to evaluate visual images. It can be otherwise called a SIANN (space invariant artificial neural networks). It is majorly implemented as a trained multi-layer network to perform classification decision. It is a network embedded with an advantage against other conventional approach, hence making it a more preferred approach. It is an approach built to perform different tasks simultaneously such as features extractions, reducing data dimension, classification. In addition, it has been specialized to

separate and detect patterns in an input image, thereby making this approach useful in the area of face recognition. As explained, the modern CNN encapsulated all classification decision in a single system and hence makes room for a more processing capability of the face recognition system.

2.2 CNN Architecture

CNN are presented with layers called convolutional in their architecture. These layers receive image input, detect patterns and transform the input to the next layer. The present of filters gives these layers the necessary capabilities to detect and differentiate patterns. Also, each layer is associated with a needed filter for a specific function. Furthermore, the CNN consist of progression of layers but that can be sub grouped into part: (i) feature learning part and (ii) classification part. In the initial part, convolution and pooling are the operations performed and it is a point where image features are extracted and detected. Furthermore, the latter part stands as a classifier on the said extracted features. Nevertheless, there are different layers used for building a CNN but they are streamed down to (i) Convolutional layer, (ii) Pooling layer and (iii) Fully connected layer.

2.3 Convolutional Layer

This is a layer where the input images are convolved. This layer is the brain behind the CNN architecture. Convolution is regarded as a situation whereby a filter is applied on an input data (image) and gives an activation result. Also, it can be said as a linear operation involving multiplication performed between set of weights and input. It is the layer required for the extraction of features from an input image.

2.4 Pooling Layer

This layer is commonly and periodically used in a CNN for reducing the size of the inputs data to enhance the computational speed of the network. It works on each feature map independently. Hence, anytime situation of excessive image input arises, the pool layer section will reduce the number of parameters. Moreover, pooling can be of different types. We have: max pooling, average pooling, and sum pooling. Overtime, the most commonly used is the max pooling. Max pooling is a process called the sample-based discretization. It down-sample an input data, reducing the input dimensionality and create space for an assumption to be made regarding the sub-regions where the feature is.

2.5 Fully Connected Layer

This layer is regarded as the final stage of a CNN. Every input neuron has a full connection to every activation unit. Every process (output) done in the feature learning part are passed to this layer for proper classification decision or to predict the best label describing an image. This process is flattened to make a single vector of values such that each of it will represent a certain feature belonging to a label. This layer has its own backpropagation process it uses to determine an accurate weight. In addition, Convolutional neural network (CNN) is the most popular technique used in face recognition. It is used to tackle problems in computer vision [15]. CNN can be trained by two methods: (i) based on classification layer and (ii) based on metric learning. Metric learning minimizes intraclass variance and helps to maximizes the interclass variance [16,18]. Object detection, image classification, face recognition are numerous vision tasks that have greatly benefitted from discriminative representation learned via CNNs. Over time, lots of the scale of training dataset for CNN has emerged to obtain optimal accuracy such as CelebFaces+ [26], VGG face dataset, MS-Celeb-1M, VGGFace2, CASIA-Web Face and UMDFace and many more [30]. These are large scale face datasets but the issue associated with them is that they often constitute noisy signals [2,7,10,31]. This happens when they are collected from movies or image search engine. Hence, Max-Feature-Map (MFM) is an operation and alternative of ReLU which can be regarded as an implementation of maxout activation for separating the enormous noisy and informative signals through low-activation neurons suppression in each layer. Lot of large-scale network input in this approach are raw data used for detecting and classifying underlying patterns through convolve filters in multiple levels. Lightened CNN showcase a computational efficiency using MFM for extracting theoretical illustrations in face recognition and VGG-Face constitute it [5,10]. Both are major Deep CNN models that have shown high successful rate in face recognition. The essence of CNN is to by-pass understanding our images and selecting appropriate features manually like color histograms, binarization, histogram of gradients and many more which pose a tedious work to do and which is extensively dependent on our original dataset, that is, feature engineering done on one image dataset cannot always be applied to another image dataset [9].

3 METRIC LEARNING

It is a contemporary development in deep learning which is applicable to visualizing similar images in an embedding manifold. This is a part of supervised learning relating to classification and regression. Through this, models learn distance functions at the preprocessing stage between instances. An illustration of this can be seen in [29,30] where triplet loss is applied to lessen image feature to a low dimension.

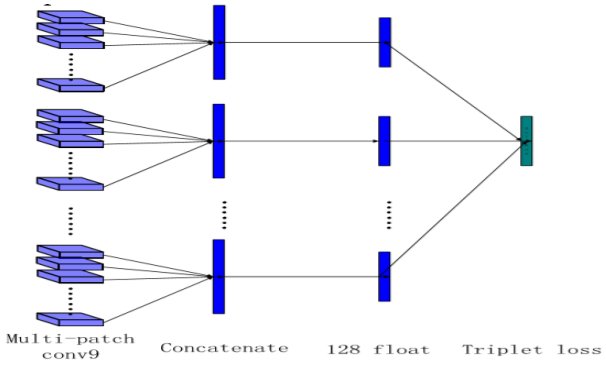

Fig 1: *Metric Learning with triplet loss*

Metric learning focus shortening the distances between samples or classes belonging to an identity and expanding the distances between samples of different identity. It thereby tackles difficulties associating with computing similarity between two faces with intra-personal variations. Over time, use of metric learning has resulted into a well discriminative features for a better visualization.

4 RELATED WORK

CNN has successively proven to achieve remarkable results and drawn attentions of researchers in recent years in the field of face recognition. This section contains brief review on CNNs. The work [33] presented the performance of network fusion using three single layers (Pool4, FC, and SoftMax) and one combination of FC and SoftMax (FC+SoftMax), which is a fusion of 2 layers. It was shown that network fusion can greatly improve face recognition accuracy and performance by using separate networks to capture information from several regions and scaling to a structure of powerful face representation. Also, [20] proposed a two novel CNN architecture which is done by fusing the intermediate layer of the network, two post-processing methods and a study on the R-CNN-based approaches for individual and multi-task approach without intermediate layer fusion. The network achieved the capability to understand more discriminative features and post-processing methods leveraged on landmarks and detection scores for a region. This experiment was done on an unconstrained dataset to demonstrate its effectiveness. The research on face recognition via deep embedding used a two-stage approach combining deep metric learning and a multi-patch deep CNN to achieves a high discriminative feature for face verification and recognition [29]. It was claimed that it outperforms other state-of-the-art methods on LFW dataset which thus obtain 99.77% pair-wise verification accuracy. This method proved to handle occlusions, variant poses and expressions. An extensive DL framework to learn face representation with a multimodal information was established in [26]. It is a design of a comprehensive convolution neural networks (CNN) with a layered stacked auto-encoder. It works to actively extracts correlative facial features and then get concatenated to structure a high dimensional feature vector, whose dimensions are compressed by stacked auto-encoder (SAE). CNN recognition ability was enhanced such as filters, deep structures, data augmentation, ReLU nonlinearity, L2 normalization and several losses on multi-stage training. Lastly, a feature level fusion approach was presented using SAE to merge the extracted features from the CNNs to achieve an advantage of learning non-linear dimension reduction. In addition, combining numerous biometric methods for multimodal face and iris recognition to improve system performance was introduced in [24]. A system which verify the attainability of a framework multimodal biometric by the fusion of two biometric modalities face and the iris. Multimodal biometric framework performance depends on the technical normalization choices. Based on FER-2013 dataset, [32] proposed two novel convolutional neural networks which are unique and simple for selecting hyper-parameters across the layers in a network. It was shown that significant impact can be caused by number of filters and the kernel size on network accuracy. A proposed work showcases a deep learning-based face recognition attendance system where a CNN for creating face embedding and CNN cascade for face detection is used in [17]. It was shown that face images in smaller number with a method of augmentation achieved a 95.02% accuracy.

5 METHODOLOGY

It is not a new thing in the field of machine learning and deep learning that face recognition has lots of challenges associated with it. The major task of deep learning is feeding it with lots of images to learn upon. Most at times, it is quite appropriate to train a model with few numbers of instances. Also, the reproduction of the model with new added classes should not be required at all times. Fig 2 shows the different component of the proposed face recognition model. This illustrates the different stages at which several instances pass through before reaching the output stage.

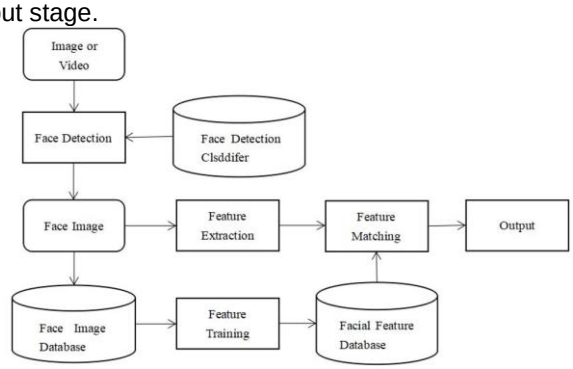

Fig 2: *The Face Recognition component*

The face recognition model is created with the use of pre-trained VGG weights and applying transfer learning. Also, an auto-encoder is used to represent images as vectors. This model is carried out on a NVIDIA GeForce GTX 1060 of 6GB VRAM. This is due to the fact that the processing speed of images or videos is a lot faster on GPU than using the CPU. Although, it is possible on CPU but it is a lot relieving working on a high GPU. For easy model building, TensorFlow framework is incorporated for its flexible architecture that aid excellent computational deployment. The model uses a CNN classifier to successfully achieve face detection and thereby making an appropriate contrast on the instance.

5.1 TensorFlow

This is a framework used for machine learning implementation especially for neural networks. It is an open source that takes in set of input as a multidimensional array. That is, there is a flow of operation from the input stage to the output stage. It is a versatile platform and hence has been used to implement face recognition based on convolutional neural networks.

5.2 Autoencoder

This is a tool used in neural networks that works to learn and create different representation of an input. Therefore, it has been used for an image compression and dimensionality reduction

5.3 Vector Similarity

This vector similarity is used to compare between pictures to know if the person in input images resembles another by finding the distances of the vectors(images). Sometimes, it can be called a Siamese network. Euclidean distance and Cosine distance are common methods to find distances

between vector. Here, we made use of cosine distance. A threshold is set at a lower value to 0.5.

5.4 Testing the model

Fig 3 shows that the model successfully verifies the instances that both person in the image are similar by giving a cosine similarity of 0.25 thereby loading a pre-trained model of VGGFace in conjunction with a similarity metric to ascertain two faces are similar or not similar. The loaded similarity metric creates a similarity score between the faces under comparison. A threshold is also set at a lower value to make the face matching strict in detecting and comparing between the faces.

```
verifyFace("angelina.jpg", "angelina2.jpg")
```

```
Cosine similarity:  0.2544904947280884
They are same person
```

Fig 3: *Testing output of two instances*

Fig 4 shows the comparison of two images and our models actualized that the person in each image are totally different with a cosine similarity of 0.688.

```
verifyFace("angelina.jpg", "Pheobe.jpg")
```

```
Cosine similarity:  0.6882018446922302
They are not same person!
```

Fig 4: *Output of different Instances*

Also, this proves that the higher the cosine similarity, the less in similarities between images. In addition to this model, we apply this model on camera and also using cosine similarity as our vector distance. This is done by training our model with set of instances (images). The model extracts the instance feature and make comparison with any images or person that passes through the camera. The extracted features are saved in a database. Hence, the camera captures and recognize any person or image face by making a comparison with feature extracted and images save in the database.

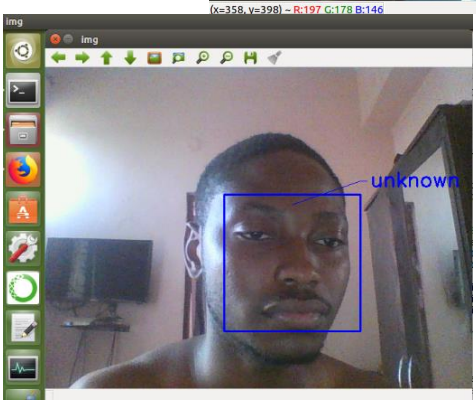

Fig 5: *Face Recognition Output on Cam*

This model is actually a one-shot learning and it attained a 97% accuracy.

6 CONCLUSIONS

Face recognition has adversely impacted peoples' life through the implementation of deep learning. Deep learning techniques has the ability to learn and train on a very large dataset of data, learning on low level features from raw data with little processing and detecting complex interactions. Thereby, the use of these deep learning techniques will integrate for a more secure system and for building accurate model, helping to curb the criminates act which is on the rise every day. The face recognition will help using these techniques for making more secure system for society, academics, researchers and also will help for future scope in industries. In this regard, this paper has been able to established that an appropriate training is needed to achieve a deserving result. Also, the proposed work can be applied to any camera footage system and use by security sectors to fight against crimes.

REFERENCES

[1] M. M. Ghazi, H. K. Ekenel: A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition. arXiv: 1606.02894v1, pp. 34-41 (2016).

[2] X. Wu, R. He, Z. Sun, T. Tan: A Light CNN for Deep Face Representation with Noisy Labels. Journal of LATEX class files 14(8), arXiv:21511.02683v, pp. 1-13 (2017).

[3] A. Admane, A. Sheikh, S. Paunikar, S. Jawade, S. WAdbude, M. J. Sawaarkar: A Review on Different Face Recognition Techniques. IJSRCSEIT 5(1), pp. 207-213 (2019). DOI: https://doi.org/10.32628/CSEIT195159

[4] M. A. Abu, N. H. Indra, A. H. Abdrahman, N.A. Sapiee, I. Ahmad: A study of Image Classification based on Deep learning and TensorFlow. IJERT 12(4), pp. 563-569 (2019).
[5] J. Deng, J. Guo, N. Xue: Arc Face: Additive Angular Margin Loss for Deep Face Recognition. arXiv: 1802.07698v3, 9 Feb, 2019.
[6] M. Hashemi, M. Hall: Criminal tendency detection from facial images and the gender bias effect. Journal of Big data 7(2), pp. 1-16 (2020). DOI: https://doi.org/10.1186/s40537-019-0282-4
[7] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, F. Huang: Curricular Face: Adaptive Curriculum Learning Loss for Deep Face Recognition. arXiv:20004.00288v1, April 2020.
[8] I. Masi, Y. Wu, T. Hassner, P. Natarajan: Deep Face Recognition: A Survey. Sibgrapi Conference on Graphics, Patterns and Images, DOI: 10.1109/SIBGRAPI.2018.00067, pp. 471-478 (2018).
[9] S. Balaban: Deep learning and face recognition: the state of the art. arXiv: 1902.03524v1, May 2015.
[10] P. S. Prasad, R. Pathak, V. K Gunjan, H. V. Ramana: Deep Learning Based Representation for Face Recognition. DOI: https://doi.org/10.1007/978-981-13-8715-9_50, Jan 2020.
[11] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis: Deep Learning for Computer Vision: A Brief Review. DOI: https://doi.org/10.1155/2018/7068349, Feb 2018.
[12] A. Elmahmudi, H. Ugali: Deep face recognition using imperfect facial data. DOI: https://doi.org/10.1016/j.future.2019.04.025, pp. 213-225 (2019).
[13] Z. Zhu, P. Luo, X. Wang, X. Tang: Deep Learning Identity-Preserving Face Space. DOI: 201001DOI04648280, ICCV 2013.
[14] J. Mehta, E. Ramnani, S. Singh: Face Detection and Tagging using Deep Learning. International Conference on Computer, Communication and Signal Processing (ICCCSP), pp. 1-6 (2018). DOI: 10.1109/ICCCSP.2018.8452853
[15] U. Aiman, V. P. Vishwakarma: Face Recognition using modified Deep Learning neural network. ICCCNT, IEEE-40222, July 2017.
[16] Z. Pei, H. Xu, Y. Zhang, M. Guo, Y. Yang: Face Recognition via Deep Learning Using Data Augmentation Based on Orthogonal Experiments. DOI: 10.3390/electronics8101088, pp. 1-16 (2019).
[17] M. Arsenovic, S. Sladojevic, A. Anderla, D. Stefanovix: FaceTime - Deep Learning Based Face Recognition Attendance System. International Symposium on Intelligent Systems and Informatics pp. 53-58 (2017.)
[18] S. Cui, Y. Zhou, Y. Wang, L. Zhai: Fish Detection using Deep Learning. DOI: Applied Computational Intelligence and Soft Computing, https://doi.org/10.1155/2020/3738108 (2020).
[19] Y. Sun, X. Wang, X. Tang: Hybrid Deep Learning for Face Verification. IEEE, ICCV 2013.
[20] R. Ranjan, V. M. Patel, R. Chellappa: HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, arXiv: 1603.01249v3, 2016.
[21] M. M. Krishna, M. Neelima, M. Harshali, M. V. G. Rao: Image classification using Deep learning. International Journal of Engineering and Technology 7(2), 624-617 (2018). DOI: 10.14419/ijet. v7i2.7.10892.
[22] E. Zangeneh, M. Rahmati, Y. Mohsenzadeh: Low Resolution Face Recognition using a two-branch Deep Convolutional Neural Network architecture. DOI: https://doi.org/10/1016/j.eswa.2019.112854, 2019.
[23] S. Loussaief, A. Abdelkrim: Machine Learning framework for Image Classification. Advances in Science, Technology and Engineering Systems Journal 3(1), 1-10 (2018).
DOI: https://doi.org/10.2174/2213275911666181025143029
[24] L. Mansoura, A. Noureddine, O. Assas, A. Yassine: Multimodal Face and Iris Recognition with Adaptive Score Normalization using Several Comparative Methods. Indian Journal of Science and Technology 12(7), pp. 1-8 (2019). DOI: 10,17485/ijst/2019/v12i7/140755.
[25] X. Han, Q. Du: Research on Face Recognition Based on Deep Learning. pp. 53-58 (2018) ISBN: 978-1-5386-3903.
[26] C. Ding, D. Tao: Robust Face Recognition via Multimodal Deep Face Representation. IEEE Transactions on Multimedia, 17(11), pp. 2049-2058 (2015).
[27] S. Bussa, S. Bharuka, A. Mani, S. Kaushik: Smart Attendance System using OPENCV based on Facial Recognition. IJERT 9(3), pp. 54-59 (2020).
[28] K. Grm, V. Struc, A. Artiges, M. Caron, H. K. Ekenel: Strengths and weaknesses of deep learning models for face recognition against image degradations. The Institution of Engineering and Technology Journals 7(1), pp. 81-89, (2017). DOI: 10.1049/iet-bmt.2017.0083.
[29] J. Liu, Y. Deng, T. Bai, Z. Wei, C. Huang: Targeting Ultimate Accuracy: Face Recognition via Deep Embedding. arXiv:1506.07310v4, 2015.
[30] H. Kulkarni, G. Tofighi: Unconstrained Facial Recognition using Supervised Deep Learning on Video. ResearchGate 2018.
[31] G. Goswami, N. Ratha, A. Agarwal, R. Singh, M. Vatsa: Unravelling Robustness of Deep Learning Based Face Recognition against Adversarial Attacks. AAAI Conference 2018.
[32] A. Agrawal, N. Mittal: Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy. DOI: https://doi.org/10.1007/s00371-019-01630-9 (2019).
[33] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, T. Hospedale: When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition. IEEE International Conference on Computer Vision Workshop, DOI: 10.1109/ICCVW.2015.58, 2015.
[34] S. Sharma, A. Sodhi, N. Nagpal, A Dureja, A Dureja, Ajay: Practical Approaches: Machine Learning Techniques for Image Processing and Text Processing. International Journal of Information Systems & Management Science, Vol. 2, No. 2, 2019.

[35] A. Dureja, P. Pahwa: Analysis of Non-Linear Activation Functions for Classification Tasks Using Convolutional Neural Networks. Recent Patents on Computer Science 12(3), 2019.

[36] Michael Farayola is pursuing his Master of Technology degree in Computer Science and Engineering from PDM University, India. He received his Bachelor of Technology degree in Pure and Applied Mathematics in 2015 from FUTMINNA, Nigeria. He has published a paper in a reputed international journal. His research interests include data science, machine learning and face recognition. E.mail: mayostictos3@gmail.com .

[37] Aman Dureja is pursuing his PhD from GGSIPU, New Delhi. He received his Master of Technology degree from MDU University in year 2010 and Bachelor of Technology from Bhiwani Institute of Technology and Sciences in year 2007. Currently working as Assistant Professor in Department of Computer Science and Engineering, PDM University since 2010. He has published more than 20 research papers in reputed International Journals including Scopus Indexed and conferences including IEEE. His main research work focuses on machine learning and deep learning. Email: aman_engg@pdm.ac.in .