Advanced Deep Learning Models for Discriminating Real

Jiaguo Lin *

Seaver College, Pepperdine University, Malibu, 90263, United States

* Corresponding Author Email: jiaguo.lin@pepperdine.edu

Abstract. The rapid development of the field of artificial intelligence has brought many significant innovations across many domains. Nevertheless, it has also led to an increasing number of online crimes, particularly those involving the misuse of synthetic images. This study focuses on the critical issue of discriminating between real and fake faces using several advanced deep learning models. The models that are used are Convolutional Neural Networks (CNN), ResNet-50, and InceptionV3, and these models are used on a dataset containing both real and synthetically generated face images. Based on the result, the improved InceptionV3 model achieves the highest accuracy rate with 92.08%, followed by the improved ResNet-50 model at 89.37%. Additionally, the second CNN model eventually reaches an accuracy of 85.99%. This study emphasizes the importance of using advanced deep learning models to fight against online discrimination and fraud. In the future, data augmentation and model refinement will be added to further enhance the accuracy and robustness of real and fake faces discrimination tasks. Moreover, this research contributes valuable insights for the future development of deep learning in order to improve more on the field of network security and personal privacy protection.

Keywords: Convolutional Neural Networks (CNN), ResNet-50, and InceptionV3, real and fake faces discrimination, Deep Learning Models.

1. Introduction

Artificial intelligence technologies are developing at an unprecedented speed in recent decades. Artificial intelligence technologies have already provided people with the ability to create more innovations in many fields, such as medical diagnosis, self-driving cars, and NLP (natural language processing). Although these developments have offered a large number of conveniences to the human society, it also makes some online crimes easier to happen. The numbers of fake information and online scams on the internet are also increasing rapidly while artificial intelligence is developing. Some tech developers are starting to come out that artificial intelligence systems will not only enable some undesirable policing practices, but it will also open up numerous chances for serious criminal implementation [1]. The misuse of artificial intelligence technology has caused a large range of problems in society, and people are starting to care more about the information's authenticity and online security. In particular, there has been a great advancement in the field of image generation using deep generative models over the past few years, and some of the state-of-the-art methods can generate images that cannot even be distinguished from real images [2]. The rapid rise of deep learning and synthetic image generation technology allows people to create fake faces easily. The misuse of fake faces not only influences the accuracy and reliability of the face discrimination system, but also leads to more serious crime. Some criminals may use these technologies to generate fake faces which are the faces of people's relatives or friends to engage in fraudulent and deceptive activities. For example, romance fraud would happen when criminals use fake information and generate fake human images to deceive people into an intimate relationship for financial advantage [3]. In order to reduce this kind of exploitation of fake faces, finding a safe and efficient way to distinguish the differences between generated faces and real faces is extremely important. The main goal of this paper is to explore the most efficient way for real and fake face discrimination technology by using modern deep learning methods.

To address the real and fake face discrimination, researchers have proposed many deep learning-based models. Among these models, convolutional neural networks (CNN) perform really well on image tasks. CNN already had remarkable success in the field of image classification since it not only can do end-to-end trainable networks but also can extract features accurately and quickly [4]. In the architecture of CNN, ResNet-50 made by Microsoft and InceptionV3 made by Google both have their own unique design and features, which both have powerful performance. ResNet-50 is a refined model of residual network (ResNets) which is deep convolutional networks, and its basic idea is utilizing shortcut connections to skip the blocks of convolutional layers [5]. This idea effectively addresses the vanishing gradient and information flow in Deep Neural Networks. InceptionV3 is the third version of inception series that is a deep convolutional neural network developed by Google, and it has a better computational efficiency and performance on the network. InceptionV3's architecture is composed of some interconnected inception modules, and these inception modules all have different types of convolutional layers and pooling layers [6]. The goal of this study is to delve into the real and fake face discrimination task, and offer practical and better methods to solve the task. Firstly, this study will use the regular CNN model, ResNet-50 model, and InceptionV3 model to test their performance on real and fake face discrimination tasks based on the accuracy and loss value. Subsequently, this study will work on improving these three models by introducing additional layers and techniques to optimize the performance. According to these three improved models and their performance, this study will find the best models that fit this task and provide practical and feasible methods. Finally, this study will summarize the experimental gains and limitations, which will be helpful for possible future research direction in this field.

2. Data and Method

2.1. Data

The dataset implemented in this study is sourced from Kaggle [6]. This dataset is derived from the Zenodo, which is the original dataset [7]. It includes a large number of real human faces and fake human faces that are generated in multiple different ways. In this way, the model would not be limited in discrimination to one single type of fake face. The image in the dataset is all 256 x 256 color JPEG files. The dataset has a training set, validation set, and testing set, which can help to increase and evaluate the performance of real and fake faces discrimination. The distribution of the dataset is as follows: training set contains 70,001 fake face images and 70,001 real face images used for the training of the model; validation set contains 19,641 fake images and 19,787 real images used to monitor the model's performance and tune hyperparameters during training; testing set contains 5,489 fake face images and 5413 real face images used to evaluate the performance of the model on unseen data.

Before starting to train the model, data preprocessing is an unavoidable process because it helps to enhance the quality and effectiveness of the training process for the models, which can also boost the performance of the models. The first stage for data preprocessing is to do face detection and alignment on the images, which guarantees that the face parts of all images are aligned and centered. The second stage is to normalize the images by dividing them by 255.0. Through these steps, these ensure the data's consistency and comparability which is beneficial to evaluate and find the best model.

2.2. Model and Evaluation

Convolutional Neural Network (CNN) is a really classic deep learning model, and it is widely used in many fields of tasks. CNN already has countless achievements and becomes one of the most representative neural networks in deep learning, and it not only helps humans accomplish many tasks that were considered as impossible but also its basic principle behind is not complicated [8]. The input layer of CNN is for images. Then it has the convolutional layer which allows the filter and kernel to do the procedure of convolution. This procedure can extract features from the original images. After the features are extracted, the pooling layers are able to simplify the output of the convolutional layer. In the last, there would be fully connected layers that can connect pooling layers to the output neuron. This study built a classic CNN to do real and fake face discrimination. The model applies several convolutional layers, pooling layers, batch normalization layers, and dropout layers, and fully connected layers. The input layer accepts 256 x 256-pixel color images as input. There are four convolutional layers and each convolutional layer uses ReLu activation function to learn features of the image. Followed by each convolutional layer, there is a maximum pooling layer which is for reducing the dimension of the feature map and batch normalization layers which is beneficial to speed up training and increase the robustness of the model. After the first two convolutional layers, there also is dropout layers for each with rate 0.2, which can prevent the occurrence of over-fitting. At the end of the model, there is a flatten layer to flatten the feature map into a vector to feed the fully connected layers. There are three fully connected layers, each with a softmax activation function, and this is used to output the probability of real and fake faces. This model might not be the best CNN model that has the best performance. For this reason, a second CNN model is built which is modified based on the first model. There is an additional convolutional layer before the flatten layer of the first model. For this extra convolutional layer, maximum pooling layer, batch normalization layer, and dropout layer with the rate of 0.2 is utilized, too. These additional layers might provide better performance on the task of real and fake face discrimination, but it also has the possibility to cause overfitting and reduce the accuracy. Hence, these two models will be tested on their accuracy and loss value. Both models use the Adam optimizer and binary cross-entropy loss function. They will be trained for 50 epochs while performance will be monitored by using the validation set of the dataset.

Residual Network (ResNet) is a model that combines the convolutional neural network and residual learning. Since artificial intelligence has been developing for several years, in order to solve more complex problems, making the neural network deeper is the most common way. Nevertheless, training a deep neural network is extremely difficult because of many problems like degradation problems and vanishing gradient problems, and the emergence of residual networks overcomes these problems [9]. ResNet has great performance in computer vision tasks, and it has become a popular and important model for tasks like image classification, discrimination, and object detection. The core of its principle is the idea of residual learning, and it adds residual blocks in the model. This special block allows the information to be passed directly between different layers of the neural network without being lost when it gets deeper. In this way, the neural network is able to learn the difference between the input and output, which is the residual map, and this would make it easier to train very deep neural networks. The residual network used for the task of real and fake face discrimination is ResNet-50. It is a famous residual network because of its high performance on image classification tasks. ResNet-50 is composed of several residual blocks, convolutional layers, batch normalization layers, pooling layers, and fully connected layers. Firstly, the model accepts 256 x 256-pixel color images as input. Then, the model contains multiple residual blocks, which enable the gradients to be propagated efficiently and model to be trained stably when the model is really deep, and each residual block is followed by a series of convolutional layers and batch normalization layers. Afterward, a global average pooling layer is applied to reduce the size of the feature map to a fixed-size vector. Finally, the fully connected layer is used to output the probability of real and fake faces by softmax activation function. This model is initialized with a pretrained weight on the ImageNet dataset. Based on this regular ResNet-50 model, further additional layers are added in order to try to improve the performance of the model. A global average pooling layer is added which is good for reducing the number of parameters and preventing overfitting. A fully connected layer and batch normalization layer are also added before the final output layer because they help the model to better get the abstract features in the images. After all, these extra layers still need to be tested in order to determine whether they help the performance of the model.

Inception version 3 (InceptionV3) also is a deep convolutional neural network. It is developed by Google and it is a 48 layered convolutional neural network architecture [10]. The aim of this model is to solve computer vision problems such as image classification, which is similar to ResNet. Nonetheless, InceptionV3 has its own unique characteristics. The main principle of InceptionV3 is using the multi-scale convolutions and parallel convolution strategy. It introduces the inception module. The Inception module contains multiple convolutional kernels which have different sizes, and this allows the model to extract features of different scales at the same time. Another characteristic of InceptionV3 is that it uses an auxiliary classifier to enhance gradient propagation and optimize deeper layers training. For these features, InceptionV3 should have a nice performance on real and fake faces discrimination. The InceptionV3 model for this real and fake face discrimination tasks have an input layer with size 256 x 256. Afterwards, it has multiple convolutional layers and pooling layers with different sizes of kernel, which is effective to extract features from images and decrease the dimension of the feature maps. Then, it includes the inception module, which is the core of InceptionV3, and the inception module consists of a series of convolutional kernels and pooling operations of different sizes. Moreover, each inception module has several sub-module, and each also uses convolutional kernels of different sizes to get features of different scales. These sub-modules not only enable the model to learn multi-scale feature representations at the same time but helps to improve the performance of the model. In the last, there is a global average pooling layer and some fully connected layers. Similar to ResNet-50, this study will add additional global average pooling layers and fully connected layers to test whether these could provide a higher performance on the real and fake faces discrimination task. For the InceptionV3 model's weight, the pretrain weight on the ImageNet dataset will be used for this task. This model will be trained with the Adam optimizer with learning rate 0.001 and categorical cross entropy loss function by 10 epochs.

3. Result and Discussion

After every model is built, each of them is tested separately in order to find their performance on the task of real and fake faces discrimination tasks. Accuracy and loss value are going to be the metrics to test the performance of the model in this study. First of all, the accuracy for the CNN model is 85.44 %, and the loss value is 0.3943. The CNN model with extra layers has a slightly higher accuracy and lower loss value which are 85.99 % and 0.3694. Figure 1 shows the comparison between these two models. Secondly, the accuracy of the standard ResNet-50 model on this task is 86.46 %, and the loss value is 0.5946. Then, the accuracy of the refined ResNet-50 for this task is 89.37 %, and the loss value is 0.3235. Figure 2 displays the comparison between these two models. Thirdly, the standard InceptionV3 model provides a 91.36 % accuracy and a 0.2985 loss value on real and fake faces discrimination tasks. After that, the modified InceptionV3 model has an accuracy of 92.08 % and a loss value of 0.2643. Figure 3 presents the comparison between these two models.

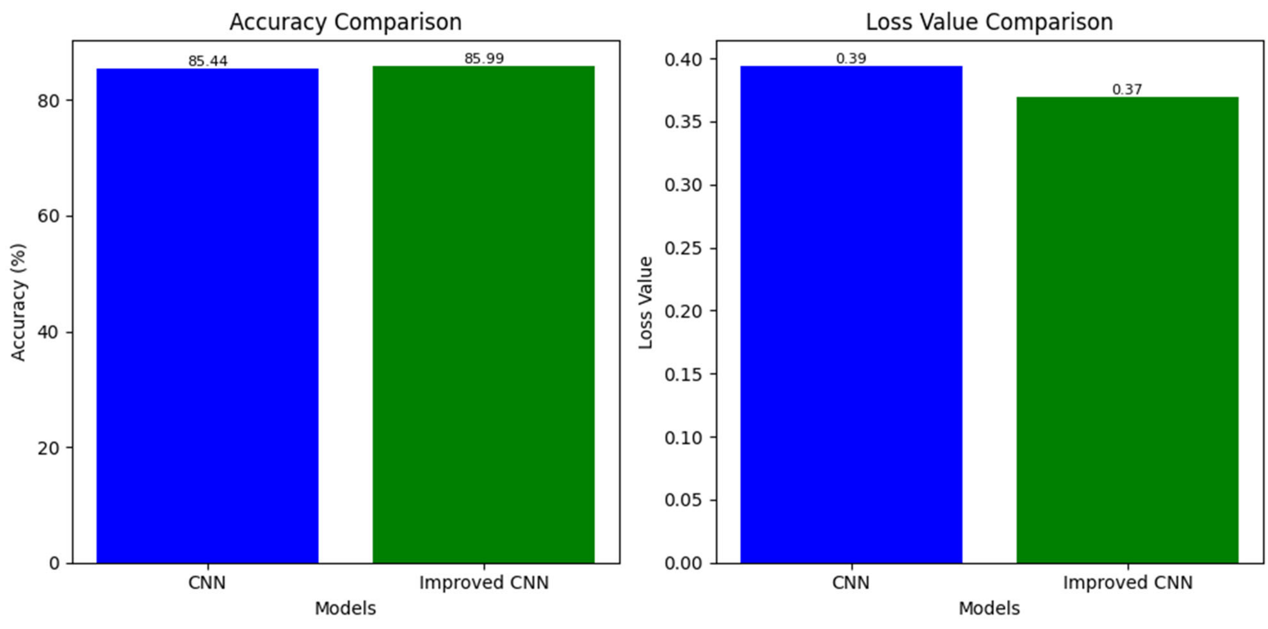

Figure 1. Performances of CNN (Photo/Picture credit: Original).

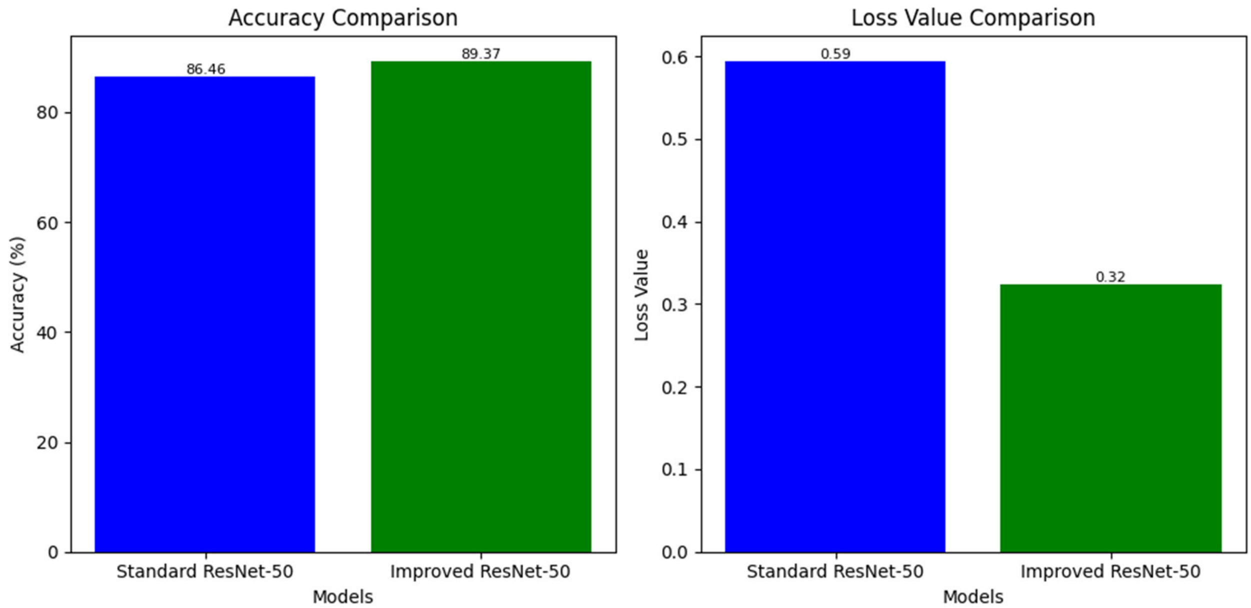

Figure 2. Performances of ResNet-50 (Photo/Picture credit: Original).

According to the results of these various models, all of these models have achieved great success on the task of real and fake face discrimination. Among these models, the InceptionV3 model has the highest accuracy and lowest loss value, which means that it has the best performance on this specific task. Follow up is ResNet-50 model, it also achieves an accuracy that is higher than 90%, and this tells that it also can handle this task and reach a high standard performance. The CNN model obtains the lowest accuracy among these three models, but it still has an accuracy that is higher than 85%. The first CNN model achieved certain success in the task of discrimination of real and fake faces, and it reached an accuracy of 85.44% and got a loss value of 0.3943. This model uses a relatively simple convolutional neural network architecture to extract image features by using several convolutional layers, pooling layers, and fully connected layers. However, the reason it got relatively low performance is that the low model complexity might not capture the high-level features of the images. For the second CNN model, more layers are added, which can extract and learn higher level features of the input images more efficiently, and these layers improve the accuracy to 85.99 % and reduce the loss value to 0.3694. This shows that the improvements on the first CNN model are effective to a certain extent, and the performance can be further optimized by adjusting the model structure or tuning hyperparameters. The ResNet-50 model has a deeper neural network structure than the CNN models, and it introduces residual connection to allow the training on a deep neural network. In this way, it is able to extract and learn complex feature representations of the images better. This is reflected in its higher accuracy rate of 86.46% and lower loss value of 0.5946. Even so, because of the increment in model complexity, it has the possibility of overfitting or difficulties in the training process. The second ResNet-50 model, which is the improved ResNet-50 model, significantly enhances the performance of the ResNet-50 model, with the accuracy increasing from 86.46% to 89.37%, and the loss value also reduced from 0.5946 to 0.3235. This better performance might due to the layers that are added provides a better training and complex feature learning. This improvement on the model's accuracy and loss value is considerable and successful, and this proves that deep models can achieve significant performance improvements in tasks with appropriate architectures refining and hyperparameter tuning. The initial InceptionV3 model showed an excellent performance, with a high accuracy rate of 91.36% and low loss value of 0.2985. The pretrained InceptionV3 model performs a strong advantage in this real and fake faces discrimination task, which emphasizes the importance of transfer learning in image discrimination tasks. By utilizing pre-trained models that trained on large-scale datasets, the model would obtain better powerful feature extraction capabilities for this specific task. Using pre-trained models trained on large-scale datasets can provide powerful feature extraction capabilities for specific tasks. After that, the improved InceptionV3 model provides the highest accuracy rate of 92.08% and the lowest loss value of 0.2643 in this research. This indicates that the complex multi-branch structure of the InceptionV3 model is very effective for the real and fake faces discrimination task.

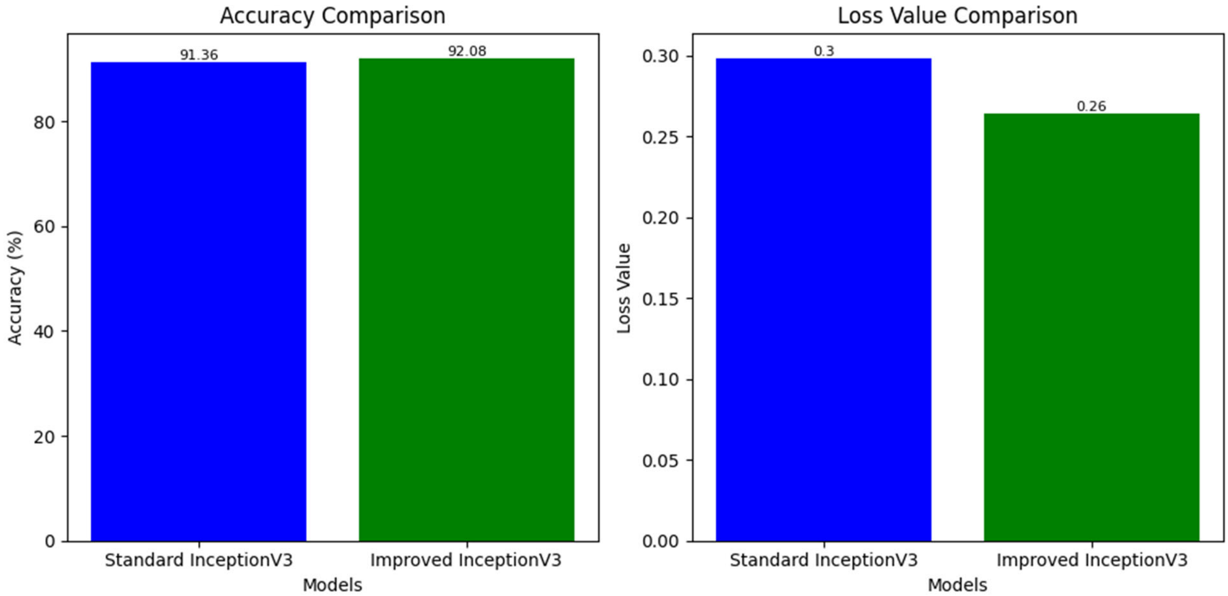

Figure 3. Performances of InceptionV3 (Photo/Picture credit: Original).

4. Limitations and Future outlooks

Despite the remarkable results of this study, there still are several limitations that need to be considered in order to improve the future research. Firstly, each model has its own limitations. Simple CNN models might provide relatively good performance but if the model needs more layers to enhance the performance, the difficulty of training such deep CNN models would increase, too. Secondly, although ResNet-50 can decrease the difficulty of training a deep neural network, when the computing resources are limited, it still may face slower training speeds. Thirdly, the complexity of the InceptionV3 model may make it difficult to deploy in an environment with resources that are constrained. Except for the limitations on each model, even though the dataset that is used in this study was processed, it may still contain some noise and errors, which may affect the performance and generalizability of the model. Additionally, the task of real and fake faces discrimination itself also has some challenges, especially when faced with highly realistic synthetic images, the accuracy of these models may decrease.

In the future, this research can continue to study in order to improve the accuracy of the real and fake faces discrimination task. Firstly, introducing more advanced data preprocessing techniques is a good direction to improve this research, such as using generative adversarial networks (GANs) to further improve the quality and diversity of training data. Moreover, considering more data augmentation techniques can expand the training set and improve the generalizability of the model. In addition, using reinforcement learning techniques allows the model to dynamically adjust its structure to adapt to more different and complex situations, which is a considerable research direction. Furthermore, the model can be applied to actual scenarios like detecting fake accounts on social media, which can verify the reliability and robustness of the model in real applications. With the development of deep learning technology, the real and fake faces discrimination can have more breakthroughs, which will provide stronger support for network security and personal privacy protection.

5. Conclusion

The main goal of this study is to test, compare, and improve the performance of different deep learning models on the real and fake faces discrimination task. Based on the result of this study, the improved InceptionV3 model performed the best among these models, with an accuracy rate of 92.08%. Improved ResNet-50 also achieved significant progress after tuning and making the neural network deeper, and it reaches an accuracy of 89.37%. In addition, the second CNN model also had some great improvements on the performance, and its accuracy increased to 85.99%. However, this study also has its limitations. The model selection and dataset preprocessing may affect the generalizability and robustness of the model. In the future research, exploring more aspects such as model improvement, data augmentation, and hyperparameter tuning can improve the accuracy and robustness of this real and fake faces discrimination tasks further. This study is really important and helpful to network security and personal privacy protection. By selecting and improving different deep learning models, providing the most powerful tools to fight against the increasingly severe online disinformation and fraud is urgently needed especially in this modern time with a high speed development on deep learning. Meanwhile, this research also can provide useful experience and some important insights for future development in the field of artificial intelligence.

References

[1] K. J. Hayward and M. M. Maas. Crime, Media, Culture 17.2, 209 - 233 (2021).

[2] Y. Lu, Z. Weng, Z. Zheng, Y. Zhang, and C. Gong, Expert Systems with Applications, 228, 120365 (2023).

[3] C. Cross and R. Layt, Social Science Computer Review, 40 (4), 955 - 973 (2022).

[4] X. Lei, H. Pan and X. Huang, A dilated CNN model for image classification. IEEE Access, 7, 124087 - 124095 (2019).

[5] E. Rezende, G. Ruppert, T. Carvalho, F. Ramos, and P. De Geus, 2017 16th IEEE international conference on machine learning and applications (ICMLA, 2017) pp. 1011 - 1014.

[6] A. Chadha, V. Kumar, S. Kashyap and M. Gupta, Proceedings of Second International Conference on Computing, Communications, and Cyber-Security (IC4S 2020) pp. 557 - 566.

[7] C. Zhang, H. Qi, Y. Li and S. Lyu, arXiv preprint arXiv: 2308. 01520 (2023).

[8] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, IEEE transactions on neural networks and learning systems 15 (2021).

[9] A. S. B Reddy and D. S. Juliet, 2019 International Conference on Communication and Signal Processing (ICCSP, 2019) pp. 0945 - 0949.

[10] C. G. Jignesh, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, Big Data Analytics: 8th International Conference, BDA (Sonepat, 2020), pp. 81 - 90.