Int. j. inf. tecnol. (January 2023) 15(1):5–15
https://doi.org/10.1007/s41870-022-01130-5

ORIGINAL RESEARCH

An improvised CNN model for fake image detection

Yasir Hamid[1] · Sanaa Elyassami[1] · Yonis Gulzar[2] · Veeran Ranganathan Balasaraswathi[3] · Tetiana Habuza[4] · Sharyar Wani[5]

Received: 20 April 2022 / Accepted: 27 October 2022 / Published online: 16 November 2022

Abstract The last decade has witnessed a multifold growth of image data courtesy of the emergence of social networking services like Facebook, Instagram, LinkedIn etc. The major menace faced by today's world is the issue of doctored images, where-in the photographs are altered using a rich set of ways like splicing, copy-move, removal to change their meaning and hence demands serious mitigation mechanisms to be thought of. The problem when seen from the prism of Artificial intelligence is a binary classification one, where-in the characterization must be drawn between the original and the manipulated images. This research work proposes a computer vision model based on Convolution Neural Networks for fake image detection. A comparative analysis of 6 popular traditional machine learning models and 6 different CNN architectures to select the best possible model for further experimentation. The proposed model based on ResNet50 employed with powerful preprocessing techniques results in a perfect fake image detector having a total accuracy of 0.99 having an improvement of around 18% performance with other models.

✉ Yasir Hamid
yasir.hamid@adpoly.ac.ae

[1] Information Security and Engineering Technology, Abu Dhabi Polytechnic, Abu Dhabi, United Arab Emirates

[2] Department of Management Information System, College of Business Administration, King Faisal University, Riyadh, Saudi Arabia

[3] Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, Chengalpattu, India

[4] Department of Computer Science and SWE, College of Info. Tech, UAEU, Al Ain, United Arab Emirates

[5] Department of Computer Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia

Keywords Artificial Intelligence · Convolution neural networks · Classification · Doctored images · Splicing

1 Introduction

Active social media users increased by more than 400 million over the past 12 months and have passed the 4.55 billion in October 2021 [1]. Thus, 57.6% of the world's population uses social media. Sharing images through different social networking services like Facebook, Twitter, or Instagram is a common practice. Billions of images are shared online daily. Fake images have been spreading out for several purposes, ranging from unharmful reasons such as impressing friends to harmful reasons such as blackmailing.

Many media editing tools are used without requiring technical knowledge to produce fake images by simply adding, duplicating, replacing, or removing objects or people. We consider fake images as images that have been manipulated in such a way that they can trick people into believing inexact contents. such spectacular images can be speedily distributed across social media platforms and catch many viewers. A study based on groups of US college students investigated individual reactions, responses, and evaluation of images in online stories on different web channels. The study concluded that most the users trust web images and fail to question the authenticity of images [2].

The impact of doctored images might be significant. for instance, in politics, the fake images might be used as a tool against competitors during the election campaign to downgrade the opposition and incline peoples' trust toward them. In recent years, the media environment uses a large volume of information that presents serious challenges for credibility judgment. A study [3] analyzed news during the 2016 U.S. presidential election discovered that the average American adult encountered at least one fake news story per month during the presidential election and the study showed that over half of the American adults who recalled seeing them believed them. Many platforms can also make money from generating and spreading out untrue content. For blackmail reasons, criminals can use fake identities/profiles and threaten to share compromising and inappropriate images with the victims' families.

t is critical to develop preventive intervention tools to limit the propagation of doctored images and to minimize their impact that might take diverse dimensions on society. Even though fake images can look extremely real, there are always indications that may help in discovering the fake ones. However, detecting fake images is not an easy task. Detecting doctored images remains a constant problem nowadays in terms of identifying the significant factors along with achieving high classification accuracy. Many approaches have been investigated recently. Machine learning provides computational intelligence techniques to tackle the issue of analysis and prediction within large complex datasets and has been used to solve a wide range of real-world problems [4–6].

In this study, six different conventional machine learning algorithms were trained and tested. Once it was confirmed that these models don't provide impressive results, thereafter the adapted Convolutional Neural Network (CNN) based models with their basic configuration were trained and tested. At this point, a comparative analysis of the CNN models was carried out, and the best model-performing architecture was selected. The selected model was empowered with a set of pre-processing techniques to get improved results.

The following points summarize the contribution of this research work.

- A crisp review of the related works is presented to help the the guide the reader through the recent developments in fake image detection.
- A detailed comparative analysis of the conventional machine learning is provided to explain the need of the CNN's for image classification.
- A comparative analysis of popular CNN models is presented to select the best model for further experiments.
- A detailed comparative analysis of the conventional machine learning is provided to explain the need of the CNN's for image classification.
- A new optimized model based on advanced deep learning techniques has been proposed for detecting the fake images. Furthermore, different preprocessing techniques have been incorporated to avoid the change of overfitting of models.
- Latest techniques have been employed to monitor the performance of the proposed model on both the training and the test sets.

The rest of the paper is organized as follows. In Sect. 2 we present a crisp review of the related research works recently published. In section 3 a detailed discussion about the materials, methods, and the methodology of this research work is presented. Detailed discussion about the results is presented in Sect. 4, and finally, the paper concludes in Sect. 5.

2 Literature review

This section presents a detailed review of the recent works which have used one or more CNN models to differentiate between real and fake images. Efforts was put to highlight the significance and drawbacks of each work. Off the two types of AI i.e., rule based AI and Machine Learning based AI, the latest one also known as pattern based AI has been pretty popular. Pattern based AI has been found applications in diverse areas of of science and technology ranging from education [7, 8], health [9, 10], cyber security [11–13], agriculture [14, [15], [16], civil defense [17], 18] etc.

In 2017, a study by Kuruvilla et al. [19], proposed a neutral network-based system trained by analyzing the error level for 8000 images where 4000 are fake and 4000 are real. The momentum backpropagation learning rule was used with a learning rate set to 0.2 and momentum to 0.7 to adjust the neuron connection weights. The trained multilayer perceptron neural network has achieved a success rate of 83% in classifying images either to real or fake. According to [19], metadata analysis is a great technique that provides useful information about how the image was generated and handled and helps in detecting tampered images without requiring complex processing.

In 2017, a study by Kuruvilla et al. [19], proposed a neutral network-based system trained by analyzing the error level for 8000 images where 4000 are fake and 4000 are real. The momentum backpropagation learning rule was used with a learning rate set to 0.2 and momentum to 0.7 to adjust the neuron connection weights. The trained multilayer perceptron neural network has achieved a success rate of 83% in classifying images either to real or fake. According to [19], metadata analysis is a great technique that provides useful information about how the image was generated and handled and helps in detecting tampered images without requiring complex processing.

Sudiatmika et al. [20] have used error-level analysis and deep learning to detect image forgery. The forgery classification dataset CASIA V2 containing a total of 12,614 images was used for the experiment where 7491 images are original and 5123 have been tampered with. After normalizing images to a unified size of 224×224, they did perform the compression error level analysis based on the fact that an original image should have a high value in the error level analysis. They have incorporated a Convolutional Neural Network to build the model and classify images into two classes: original and fake. They have used VGG16 known for its ability for large-scale image recognition. The adopted architecture is the one including convolution layers, max-pooling layers, and fully connected layers. The selected activation function was the rectified linear unit. Sudiatmika et al. have obtained a promising result as the model achieved an accuracy of 92.2% in the training and 88.46% in the testing.

Mo, H et al. [21] have proposed a Convolutional Neural Network (CNN) based model to identify whether a face image is fake or real. They have used CelebA-HQ dataset that consists of 30,000 true images and they have used the Generative Adversarial Networks (GAN) to generate the corresponding fake face images where 12,000 pairs of fake-true images were used for the training, 15,000 pairs were used for the testing, and 3000 pairs were used for the validation. In the proposed study, they have investigated six activation functions: Hyperbolic Tangent (TanH), Rectified Linear Activation (ReLU), PReLu, LReLu, ELU, and ReLu6. The LReLu activation function has produced the best performance. They also evaluated the impact of the number of layer groups on the accuracy of the proposed model.

Arruda et al. [22] have conducted several experiments to improve car detection using an unsupervised image-to-image translator. The Berkeley deep drive dataset has been used for training and evaluating the model. Images were filtered and only 12000 day and night images were kept. Fake images were generated using CycleGAN and the model was trained using faster R-CNN to detect cars. The Recurrent-Convolutional Neural Networks (R-CNN) based model was evaluated by using the mean average precision. The authors concluded that training the model by combining the day-image dataset and fake-night-image dataset did enhance the results by more than 10% compared to the use of only one of the two datasets.

ang, Chao et al. [23] believe that the most harmful image forgery is content manipulation that uses different techniques such as copy-move, removal, and splicing. The authors raised two issues in the existing models dealing with image manipulation detection; the first one is the lack of model generalization, and the second issue is the overlooking of the integrity of the forensics task. They have proposed a 2-stage constrained R-CNN-based architecture that provides a classification of the manipulation technique used to tamper the image, simultaneously with the localization of the tampered region. The image tampering datasets that have been used for the model training and evaluation are NIST16, COVER, CASIA, and Columbia. The F1-Score and the area under the ROC curve have been used as evaluation metrics and the obtained results showed the constrained R-CNN model achieved an increase in terms of F1-Score of 73.3% on the COVER dataset, 28.4% on the INIST16 dataset, and 13.3% on the Columbia dataset.

He et al. [24] studied the incorporation of residual signals from multi-color channels with the CNN to detect fake images. The study used 10000 fake images generated by PGGAN and real images available in the CelebA dataset. After converting Input images {IR, IG, IB} into color spaces including Lab, HSV, and YCbCr, CNN was used to extract deep representations and fed them Random Forest classifier to achieve better generalization capability against different post-processing attacks. The proposed CNN model consists of 4 convolution modules including a convolutional layer, an activation layer using Rectified Linear activation function, a max-pooling layer, and 3 fully connected layers. They applied the cross-entropy loss to optimize the CNNs results. The proposed model showed promising and robust detection accuracy.

Tanaka et al. [25] investigated the use of robust hashing for detecting fake images. The robust hashing method was proposed by Li et al and was applied to four fake-image datasets: CycleGAN, Image Manipulation Dataset, UADFV, and StarGAN. JPEG compression, resizing, copy-move and splicing were applied to query images. The fake-image detection method based on robust hashing outperformed Wang's method which is one the state-of-the-art and the proposed technique was demonstrated to be robust against the combination of image manipulations.

Table 1 given below shows the a summarized information about the revived paper and some other papers related to the problem domain.

3 Materials and methods

This section defines the datasets used to perform our empirical studies, followed by the description of the proposed process to build our tampered image detection model.

3.1 Convolutional neural networks

Convolutional networks (CNNs) are a special form of artificial neural networks. CNN's architecture was inspired by the mammal's frontal lobe which is responsible for processing visual stimuli. These CNNs can accomplish relatively complex tasks by exploiting data such as images, sound, text, video, etc. In an image, there are very strong links between neighboring pixels; by flattening the image, this information is lost, and it becomes more difficult for the network to interpret the data during the training phase. This justifies the use of Convolutional Neural Networks instead of Multilayer perceptron which requires providing the data

Table 1 A comparison table of the related works

Work	Technique	Dataset	Preprocessing	Accuracy (%)
Kuruvilla et al. [19]	Conventional neutral network-based	In house dataset with 8000 images	No	83
Sudiatmika et al. [20]	CNN model VGG16	CASIA V2	No	92
Yang, Chao et al. [23]	R-CNN	NIST16, COVER, CASIA, and Columbia	None	73
Kakkar [41]	Conventioal Image processing	In house dataset	Transform invariant features	90
Qu [42]	SVM	In house dataset	Hand crafted feature extraction	90
Dong [43]	SVM	In house datset	Statistical moments of run-length and edge detection	76

as a vector. The CNN architecture helps process pixel data and therefore Convolutional Neural Networks are widely used in image processing and recognition. Various CNN architectures tested in this research work are listed in this experimentation section. A complete discussion about working mechanisms of all the models is beyond the scope of this paper Fig. 1. Given below presents the work follow diagram of the research work.

3.2 Dataset description

he dataset used in this research is a public dataset available on [22] and contains two sets of face images. The first set contains the face images of real people randomly selected whereas another set contains the fake face images. The fake face images are high-quality photoshopped images. These images are composites of different faces, separated by eyes, nose mouth, or whole face. It is important to note that these face images were not generated by any model such as Generative Adversarial Networks (GAN), which can generate fake images automatically, but these images are generated by a photoshop expert. Training any model on the dataset created by any GANs can provide good results and can do a great job in discriminating between real and generated face images. The classifier can easily learn patterns between the images that are generated by GANs. But when it comes to the expertise of humans these patterns become futile. It is because the exquisite counterfeits by experts are created in a completely different process. This is the aim objective to have a dataset created by experts than any GANs. The dataset encompasses a total of 2041 images where real images are 1081 and fake images are 960. There are three types of groups within the fake image set: easy, mild, and hard. To identify which part of the face is manipulated, the name of the fake image helps in that as shown in Figs. 2 and 3.

3.3 Training

In this subsection, the training of different selected models is being reported. Some exiting CNN models namely: DenseNet [26], EfficientNet [27], MobileNetV2 [28], ResNet50 [29], VGG16 [30] has been selected to be training

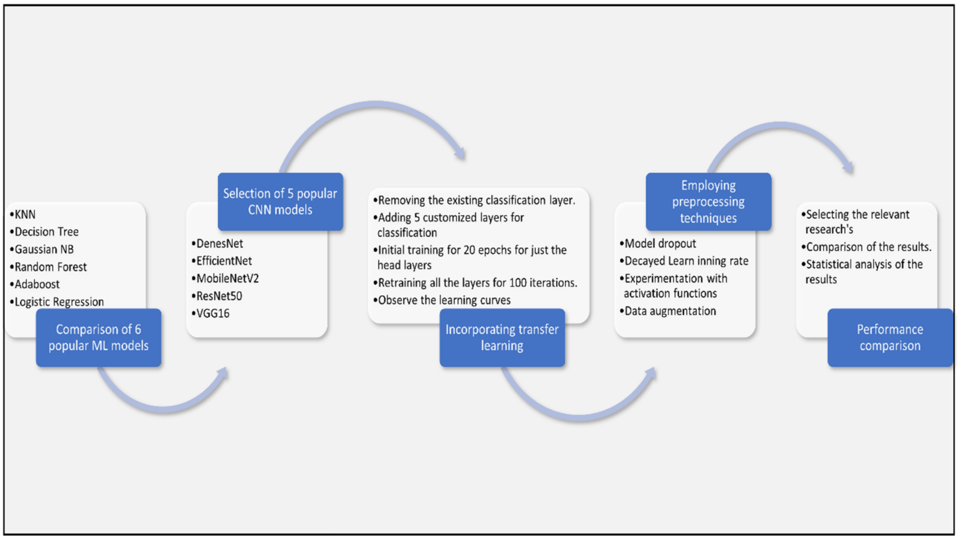

•KNN
•Decision Tree
•Gaussian NB
•Random Forest
•Adaboost
•Logistic Regression
Comparison of 6 popular ML models
Selection of 5 popular CNN models
•DenesNet
•EfficientNet
•MobileNetV2
•ResNet50
•VGG16
•Removing the existing classification layer.
•Adding 5 customized layers for classification
•Initial training for 20 epochs for just the head layers
•Retraining all the layers for 100 iterations.
•Observe the learning curves
Incorporating transfer learning
Employing preprocessing techniques
•Model dropout
•Decayed Learn inning rate
•Experimentation with activation functions
•Data augmentation
•Selecting the relevant research's
•Comparison of the results.
•Statistical analysis of the results
Performance comparison

Fig. 1 Research workflow diagram

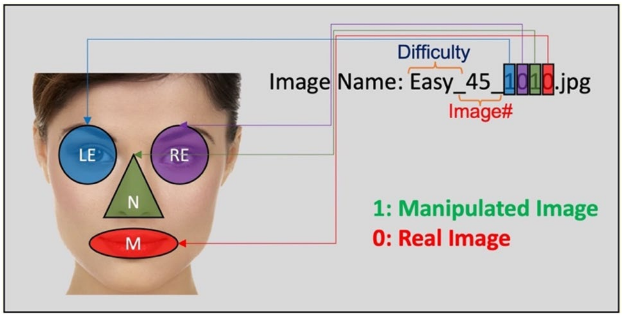


Fig. 2 Example of dataset naming

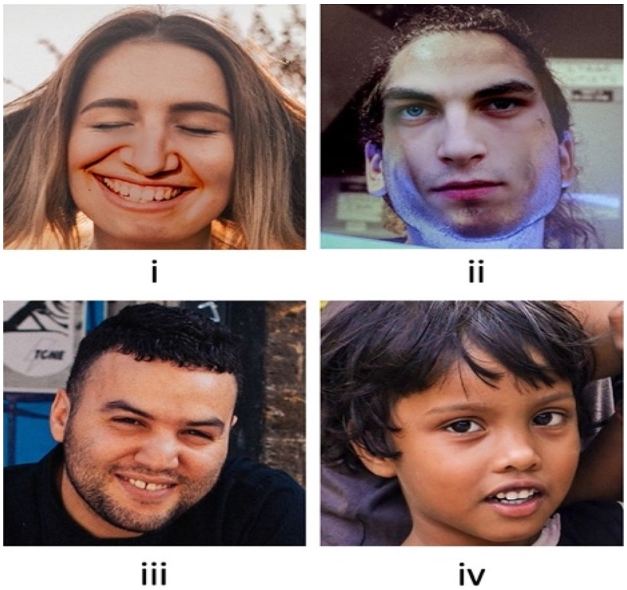


Fig. 3 Shows some of the images from the dataset as a sample. **i** and 2. **ii** Are the fake images whereas **iii** and **iv** are real images from the dataset. Sample Images

on the above said dataset. These models have been chosen based on their popularity and efficiency. These models are well known when it comes to solving image classification problems. Their accuracy rate is very high compared to other CNN models. The reason to use CNN models for the classification of the dataset was the insignificant initial results acquired by six popular conventional machine learning algorithms (KNN [31] Decision Tree [32], Gaussian Naïve Bayes [33], Random Forest [34], Adaboost Classifier [35], Logistic Regression [36]).

As these models have been proposed for different classification problems so their classification layers contain different numbers of nodes. To fit such models for our problem these models have been modified by removing the last (classification) layer. However, by doing so it does not have any effect on the weights of the learned model. A transfer learning technique has been incorporated in this study. With the help of this technique, the weights of existing layers have been frozen and only new layers have been trained on new data until the 20th iteration. after the 20th iteration, stalled payers have been unfrozen to make slight weight adjustments to the trained layers for the dataset used in this research. In addition to that, five more layers have been added to these models namely: the average pooling layer, the flatten layer, the dense layer, the dropout layer, and the SoftMax layer to these models to improve the learning rate and improve accuracy. In the average pooling layer, the pool size was set to (7,7). In flattened layer, the neurons were flattened before being fed to the dense layer, with the activation function being Relu. After that, a dropout layer with a probability of 0.5 and a classification layer with two nodes are added to the models.

3.4 Testing

At the testing stage, the best-trained model obtained from training was tested. To test the model, a subset of the dataset was used which contained 25% of the images of the entire dataset from each class. It is essential to note that the images used in testing the model were not exposed to the model before. This is to ensure the validity of the model in terms of accuracy. The results obtained during training and testing were compared to validate the proposed model.

4 Experimental setup

The objective of this study is to propose a model to distinguish between real and fake images. The performance of the proposed model was measured along with the other models found in the literature. This section describes (1) model selection (2) the data used for training and testing the proposed model and other models, (3) the performance of other models compared with the proposed model. The proposed model was implemented using Python 3.0 on Windows 10 operating system, with system configuration using *i*7 processor with 16 GB RAM. The same configuration was used for the training and testing of other models.

4.1 Model selection

It is important to identify which model will perform better on the above-mentioned dataset. for that reason, well know models such as DenseNet [13], EfficientNet [14], MobileNetV2 [15], ResNet50 [16], VGG16 [17] for image classification have been chosen and trained. Table 2 reports the precision, recall and F1-score of above-mentioned models. From the table it can be inferred that ResNet50 has outperformed all the other models in terms of Accuracy. It is important to know no preprocessing techniques has been

Table 2 Precision, Recall, and F1-score of Different Models while Training on Real and Fake images Dataset

Models	Precision	Recall	F1-Score
DenseNet	0.71	0.71	0.71
EfficientNet	0.46	0.49	0.38
MobileNetV2	0.67	0.66	0.66
ResNet	0.75	0.77	0.76
VGG16	0.66	0.66	0.66

used. Due to the high accuracy rate for among all other models ResNet50 architecture has been chosen for this study.

4.2 Training and testing data

As previously mentioned, the dataset comprises of two subsets of images; (1) real image subset contains around 1081 images and (2) fake images subset contains 960 images. Both the subsets have been divided into two sets one kept for training and another one for testing. The ratio of training and testing dataset is sat as 3:1. In other words, 75% of real and fake images are kept for training and 25% of real and fake images are kept for testing. It is important to note that the above =-mentioned ratio is before applying the data augmentation technique.

4.3 Performance metrics

To measure the performance of our model, various evaluation metrics were used such as the accuracy, the F1-Score, the sensitivity, the precision, and the specificity. The accuracy is defined as the ratio between the number of correctly classified samples and the overall number of samples and indicates the number of true positives and true negatives the model has predicated divided by the total number of predictions. The accuracy is calculated using Eq. 1.

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) \tag{1}$$

The recall is called also sensitivity and it measures the proportion of actual positives that are correctly classified as positives, and it is calculated using Eq. 2.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{2}$$

The precision called also positive predictive value is defined as the results classified as positive by the model out of all positive and it is calculated using Eq. 3.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{3}$$

The F1-Score was also used to evaluate the model precision and recall rates collectively to provide a better understanding of the misclassified records and measure the performance of the model's classification ability. F1-Score is calculated using Eq. 4.

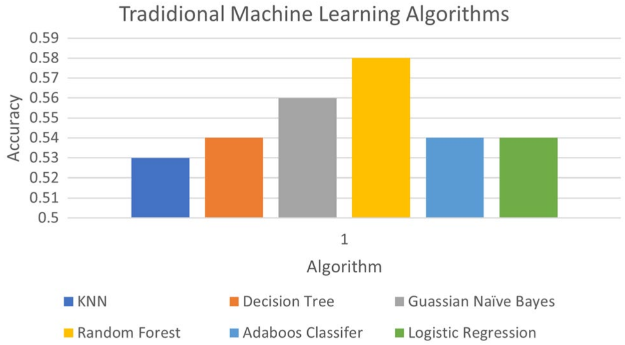


Fig. 4 Traditional machine learning Algorithms

$$\text{F1} - \text{Score} = 2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \tag{4}$$

5 Results

Initially, six different conventional machine learning algorithms listed above were trained and tested based on the dataset. Once it was confirmed that these models don't provide impressive results, thereafter the adapted CNN-based models with their basic configuration were trained and tested. At this point, a comparative analysis of the CNN models was carried out, and the best model-performing architecture was selected. The selected model was empowered with a set of pre-processing techniques to get improved results. The classification results of different traditional machine learning models are presented in Fig. 4. In this research work, we have considered accuracy as a performance metric. From the Figure, it can be noticed that the accuracy of the models on the test set is in the range of 53–58% which is not considered impressive but somehow it was expected, given the fact that the models are not suitable for the image classification. One good thing that stood out from the results is that the ensemble algorithm i.e., Random Forest has outperformed all other algorithms, which again was expected, given the fact that ensemble algorithms employ a set of weak learners to solve the difficult problem. It can be inferred from the figure the simplest algorithm KNN which doesn't have a formal learning stage has produced the least accuracy of around 53%. A reason for that is that they don't respect the spatial information present in the images. Any algorithm that just flattens the image without respecting the spatial information is bound to produce very poor results for image classifications.

After obtaining poor results from the traditional algorithms on image classification problems. We decided to test some CNN-based deep learning models for the same problem. The same dataset has been used with the same configuration. CNN is considered more suited to deal with the images and can process the image data even without compromising the spatial image data that almost all the CNN models returned better results on this dataset than any of the traditional algorithms. Figure 7 presents the results of the five CNN models namely: MobileNetV2, DenseNet121, EfficientNetB0, ResNet50, VGG16, VGG19. As explained earlier five additional layers were added to these models to have better results. From the Figure, it can be noticed that all the models showed around a 10–30% increase in accuracy compared to the traditional machine learning algorithms. Among all these models it can be seen that ResNet50 is having the highest (72%) compared to other CNN models. Whereas other models such as MobileNetV2, DenseNet121, VGG16, and VGG19 are having results of 69%, 71%, 67%, and 70% respectively. Only EfficientNetB0 are having worse results (52%) among all CNN models. Nevertheless, the results are better than traditional models, yet the results are not up to the mark.

Figure 5 shows the training and testing graphs of different models. From these graphs we can depict the training and testing performance of these models. From the graphs, it can be noticed that the models have appreciable results of the training set but have failed to perform equally well on the test set. The accuracy of the models on the training set improved with the epochs however the accuracy of the model didn't follow the trend. The accuracy of the models initially increased and later it got stagnant or decreases significantly. Likewise, the loss function decreased for the training set and didn't converge on the testing set. The reason for the significant deviation between the training and testing performance could be over-fitting, which is considered one of the common and most popular reasons as mentioned in literature. Overfitting is a case wherein the model performs very well on the training set and doesn't generalize very well on the test set. As could be inferred visually from Fig. 6, where all the models have encountered a problem of overfitting

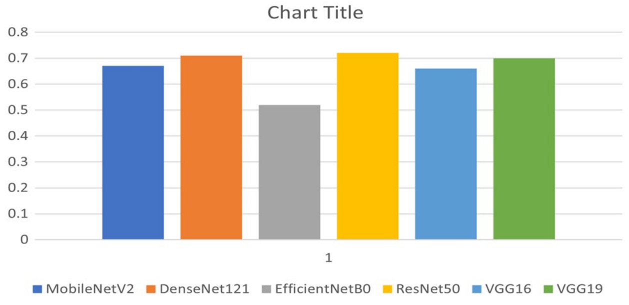


Fig. 5 Performance of CNN Models

To have better results, it is better to resolve the issue of overfitting. To do so we have adopted some pre-processing techniques such as data augmentation, adaptive learning, model checking, and dropout. However, these techniques were not implemented in all the modified CNN models but the one with the highest accuracy rate (ResNet50). These techniques are well known to overcome the case of overfitting. These techniques are briefly explained below.

Data augmentation: In data Augmentation, different types of images are artificially created by different ways of processing or a combination of multiple processing methods, such as random rotation, shifts, shear, and flip. So, ResNet50 architecture has been modified by adding five different layers as mentioned earlier, and by incorporating many pre-processing techniques. In this research work, an inbuilt function in Keras' Library [24] has been utilized to generate augmented images. For each image, 10 new images have been generated by randomly rotating images by 30%, adjusting the height and shifting width by 10%, and zooming by 20%.

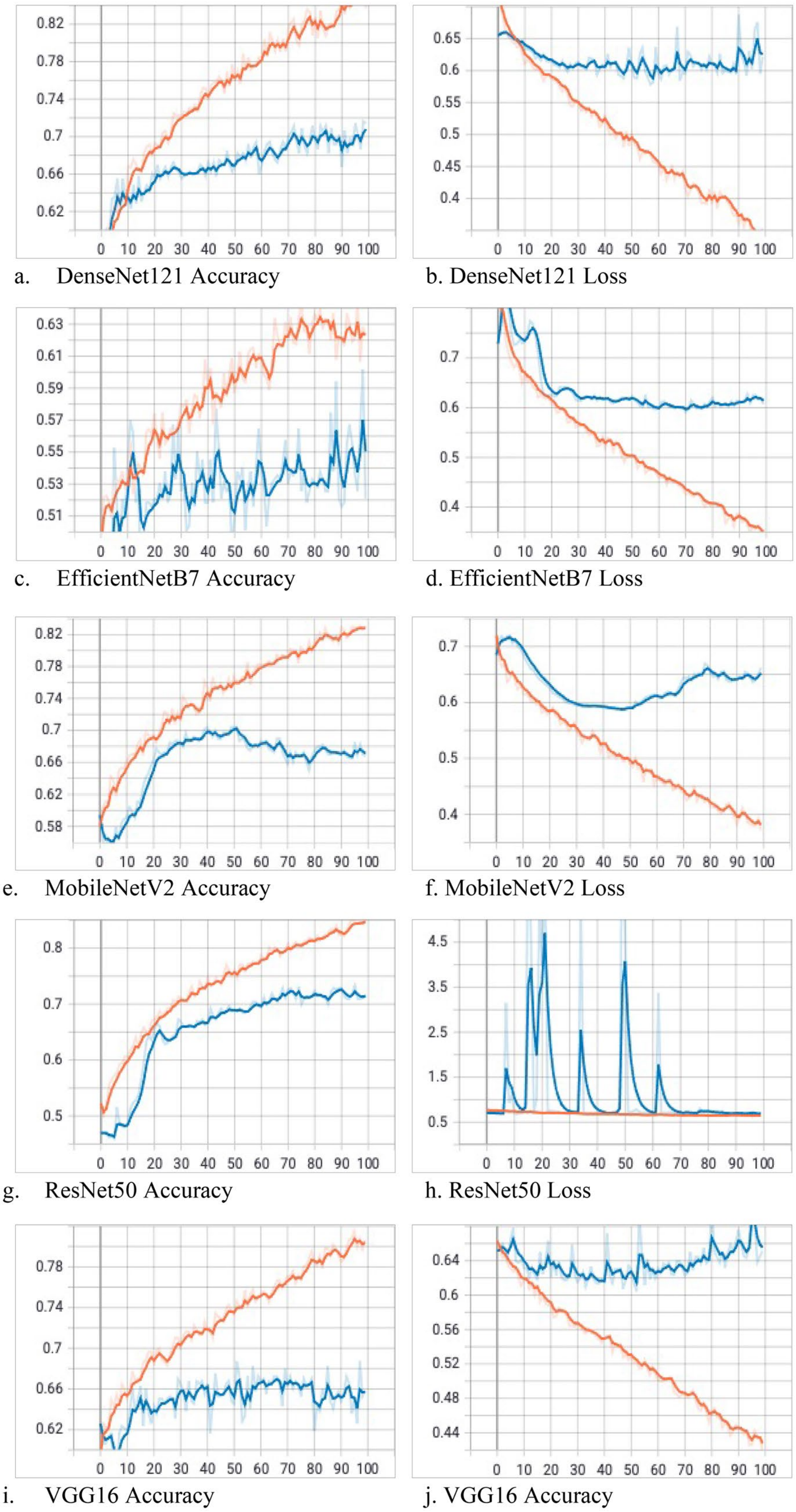
Adaptive learning rate: Learning rate schedules seek to adjust the learning rate during the training process by reducing the learning rate according to the pre-defined schedule. Common learning rate schedules include time-based decay, step decay, and exponential decay. For this work, the initial learning rate was set to 0.0001 and then the decay of the form decay = INIT_LT/EPOCHS was used.

Model checkpointing: It is the technique where checkpoints are set to save the weights of the models whenever there is a positive change in the classification accuracy on the validation dataset. It is used to control and monitor ML models during training at some frequency (for example, at the end of each epoch/batch). It allows us to specify a quantity to monitor, such as loss or accuracy on training or validation dataset, and thereafter it can save model weights or the entire model whenever the monitored quantity is optimum when compared to the last epoch/batch. In this research work, a model checkpoint of the form checkpoint = Model Checkpoint (name, monitor="val_loss", mode=" min", save_best_only = True, verbose = 1) is used, this callback monitors the validation loss of the model and overwrites the trained model only when there is a decrease in the loss as compared to the previous best model.

5.1 Dropout

Dropout is a technique used to prevent a model from overfitting. Dropout works by randomly setting the outgoing edges of hidden units (neurons that make up hidden layers) to 0 at each update of the training phase. At each training stage, individual nodes are either dropped out of the net with probability 1-p or kept with probability p, so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed.

Fig. 6 Training and validation curves



a. DenseNet121 Accuracy

b. DenseNet121 Loss

c. EfficientNetB7 Accuracy

d. EfficientNetB7 Loss

e. MobileNetV2 Accuracy

f. MobileNetV2 Loss

g. ResNet50 Accuracy

h. ResNet50 Loss

i. VGG16 Accuracy

j. VGG16 Accuracy

All the mentioned technique has been incorporated into the modified model of ResNet50. The results of the proposed model are shown in Fig. 7. From the Figure, it can be inferred that the model has performed well in terms of training accuracy and validation accuracy. With the help of the preprocessing techniques, the results have been improved. During training, the accuracy of the model has reached 100% within a few epochs and the model has sustained the same accuracy until the end of the iterations.

Whereas when it comes to the validation accuracy the proposed model has reached 100% accuracy within the first 20 iterations and maintained the same accuracy until the 100th iteration. it is because the preprocessing techniques have been incorporated. With the help of model checkpointing, the best model has been preserved whenever there is an improvement in the weights of the model. This has led to having a model with the highest accuracy rate. Although the validation loss oscillated the magnitude of the oscillation is very little to be worried about. The results presented in the figure depict a model that is trained perfectly and has converged effectively on the datasets.

The Table 3 given below present a comparative analysis of the proposed research with the recent research efforts that are similar to this one. For each of the research work information about the dataset, machine learning models and performance metrics is presented. As can be seen from the table the proposed model outperforms all the other models in accuracy.

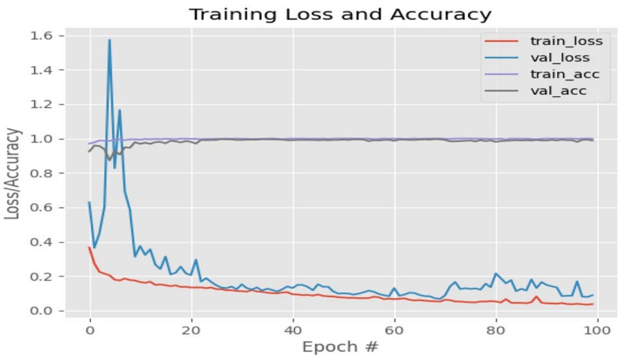


Fig. 7 Training and Validation Accuracy and Loss of Modified ResNet50

6 Conclusion

An intelligent model that decides on the originality of images based on the Convolution Neural Networks is proposed in this paper. The proposed model is evaluated using a public dataset that contains two sets of face images. The first set contains the face images of real people randomly selected whereas the second set contains the fake face images generated by a photoshop expert.

Different conventional machine learning algorithms (KNN, Decision Tree, Gaussian Naïve Bayes, Random Forest, Adaboost Classifier, and Logistic Regression) were trained and tested based on the dataset. The Random Forest algorithm has outperformed all other algorithms. The produced accuracy of the models on the test set is in the range of 53–58% which is considered not an acceptable result. This poor result is caused by the non-respect of the spatial information within images. After obtaining these poor results using traditional algorithms, we decided to implement five CNN-based deep learning models for the image classification problem. These CNN models are MobileNetV2, DenseNet121, EfficientNetB0, ResNet50, VGG16, VGG19. To enhance the result, five additional layers were added to these models. The accuracy was increased by 10–30% compared to the traditional machine learning algorithms. Among all models, ResNet50 has achieved the highest accuracy reaching 72%.

A comparative analysis of the CNN models was conducted to decide on the best model-performing architecture. The accuracy of the models initially increased and later it got stagnant or decreases significantly. Likewise, the loss function decreased for the training set and didn't converge on the testing set. To overcome this significant deviation between the training and testing performance we have adopted some pre-processing techniques and the selected model was empowered with this set of pre-processing techniques to get improved results. Data augmentation, adaptive learning, model checking, and dropout techniques were incorporated into ResNet50 that produced the highest accuracy rate. During training, the accuracy of the model has reached 100% within a few epochs and the model has sustained the same accuracy until the end of the iterations. Whereas during validation, the accuracy of the model has

Table 3 Comparison of the proposed work with other related works

Work	ML Model	Comparative analysis	Improvements	Dataset used	Performance metric
AlShariah [39]	Alexnet	Yed	None	Inhouse dataset	Accuracy 0.97
Tariq [40]	Ensemble learning	No	None	Inhouse dataset	Accuracy 0.94
Salman [41]	CNN	Yes	No	Publicly available dataset	Accuracy 0.95
Proposed research	6 popular CNN models	Yes	Yes	Public dataset available on [22]	Accuracy: 0.99

reached 100% accuracy within the first 20 iterations and maintained the same accuracy until the 100th iteration. The modified ResNet50 has performed well in terms of training accuracy as well as in terms of validation accuracy As future work we will explore many other architectures and develop a mobile application that can work detect manipulation in the images from the live feed. Also, other comprehensive datasets would be considered for the experimentation.

References

1. Obermayer N, Kővári E, Leinonen J, Bak G, Valeri M (2022) How social media practices shape family business performance: the wine industry case study. Eur Manag J 40(3):360–371. https://doi.org/10.1016/j.emj.2021.08.003
2. Kasra M, Shen C, O'Brien JF (2018) Seeing is believing: How people fail to identify fake images on the web. Extended abstracts of the 2018 CHI conference on human factors in computing systems
3. Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–236
4. Elyassami S et al (2022) Fake news detection using ensemble learning and machine learning algorithms. Combating Fake News with Computational Intelligence Techniques. Springer, Cham, pp 149–162
5. Elyassami S, Albloushi S, Alnuaimi MA, Alhosani O, Ali HA, Almarashda K (2022) Intelligent models for mining social media data. In: Advances on smart and soft computing. Springer, Singapore, pp 199-207
6. Elyassami S, Nasir Humaid H, Ali Alhosani A, Alawadhi HT (2021) "Artificial Intelligence-Based Digital Financial Fraud Detection," in International Conference on Intelligent and Fuzzy Systems 214–221
7. Al-Karaki JN, Ababneh N, Hamid Y, Gawanmeh A (2021) Evaluating the effectiveness of distance learning in higher education during COVID-19 global crisis: UAE educators' perspectives. Contemp Educ Technol 13(3)
8. Zhai X, Chu X, Chai CS, Jong MS, Istenic A, Spector M, Liu JB, Yuan J, Li Y (2021) A review of artificial intelligence (AI) in education from 2010 to 2020. Complexity 2021
9. Yasir H et al(2020) "A simple and predictive model for COVID-19 evolution in large scale infected countries,"
10. Khan SA, Gulzar Y, Turaev S, Peng YS (1987) A modified HSIFT Descriptor for medical image classification of anatomy objects. Symmetry 13(11):2021
11. Hamid Y, Journaux L, Lee JA, Sautot L, Nabi B, Sugumaran M (2017) "Large-scale nonlinear dimensionality reduction for network intrusion detection," in 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2017) 153–158
12. Hamid Y, Shah FA, Sugumaran M (2019) Wavelet neural network model for network intrusion detection system. Int J Inf Technol 11(2):251–263
13. Balasaraswathi VR, Sugumaran M, Hamid Y (2017) Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms. J Commun Inf Netw 2(4):107–119
14. Gulzar Y, Hamid Y, Soomro AB, Alwan AA, Journaux L (2018) A convolution neural network-based seed classification system. Symmetry 12(12):2020
15. Albarrak K, Gulzar Y, Hamid Y, Mehmood A, Soomro AB (2022) A deep learning-based model for date fruit classification. Sustainability 14(10):6339
16. Hamid Y, Wani S, Soomro AB, Alwan AA, Gulzar Y (2022) "Smart seed classification system based on MobileNetV2 architecture," in 2nd International Conference on Computing and Information Technology (ICCIT) 217–222
17. Hanafi MFFM, Nasir MSFM, Wani S, Abdulghafor RAA, Gulzar Y, Hamid Y (2021) A real time deep learning based driver monitoring system. Int J Perceptive CognComput 7(1):1
18. Elyassami S, Hamid Y, Habuza T (2021) Road crashes analysis and prediction using gradient boosted and random forest trees. In: 2020 6th IEEE congress on information science and technology (CiSt), June 5. IEEE, pp 520–525
19. Villan MA, Kuruvilla A, Paul J, Elias EP (2017) Fake image detection using machine learning. IRACST-International Journal of Computer Science and Information Technology & Security (IJCSITS)
20. Sudiatmika IBK, Rahman F (2019) Image forgery detection using error level analysis and deep learning. Telkomnika 17(2):653–659
21. Mo H, Chen B, Luo W (2018) "Fake faces identification via convolutional neural network," in Proceedings of the 6th ACM workshop on information hiding and multimedia security, pp. 43–47
22. Arruda VF et al (2019) "Cross-domain car detection using unsupervised image-to-image translation: From day to night," in 2019 International Joint Conference on Neural Networks (IJCNN) 1–8
23. Powers DM (2020) "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." ArXiv Prepr.ArXiv201016061,
24. He P, Li H, Wang H (2019) "Detection of fake images via the ensemble of deep representations from multi color spaces," in IEEE International Conference on Image Processing (ICIP) 2299–2303
25. Tanaka M, Kiya H (2021) "Fake-image detection with Robust Hashing," in 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech) 40–43
26. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition 4700–4708
27. Howard AG et al. (2017) "Mobilenets: Efficient convolutional neural networks for mobile vision applications." ArXiv Prepr. ArXiv170404861
28. He K, Zhang X, Ren S, Sun J (2016) "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition 770–778
29. Simonyan K, Zisserman A (2014) "Very deep convolutional networks for large-scale image recognition. "ArXiv Prepr. ArXiv14091556
30. Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46(3):175–185
31. Quinlan J (1986) "Induction of decision trees. mach. learn,"
32. Chan TF, Golub GH, LeVeque RJ (1982) "Updating formulae and a pairwise algorithm for computing sample variances," in COMPSTAT 1982 5th Symposium held at Toulouse 1982, pp. 30–41
33. Cox DR (1958) The regression analysis of binary sequences. J R Stat Soc Ser B Methodol 20(2):215–232
34. Schapire RE (2013) "Explaining adaboost," in Empirical inference. Springer 37–52
35. Peng C-Y, Joanne KL, Lee, Gary M (2002) Ingersoll. "An introduction to logistic regression analysis and reporting. J educational Res 96(1):3–14
36. "Real and Fake Face Detection. " https://www.kaggle.com/ciplab/real-and-fake-face-detection. Accessed 15 Apr 2022
37. Arnold TB (2017) "KerasR: r interface to the keras deep learning library. " J Open Source Softw 2(14):296

38. AlShariah N, Mohammed A, Khader, Saudagar J (2019) “Detecting fake images on social media using machine learning.” Int J Adv Comput Sci Appl 10(12):170–176
39. Tariq S et al. (2018) “Detecting both machine and human created fake face images in the wild.“ Proceedings of the 2nd international workshop on multimedia privacy and security
40. Salman F, Maher, Samy S, Abu-Naser (2022) “Classification of real and fake human faces using deep learning.“ Int J Acad Eng Res (IJAER) 6.3
41. Kakar P, Sudha N (2012) Exposing postprocessed copy–paste forgeries through transform-invariant features. IEEE Trans Inf Forensics Secur 7(3):1018–1028
42. Qu Z, Luo W, Huang J (2008) “A convolutive mixing model for shifted double JPEG compression with application to passive image authentication.“ 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE
43. Dong J et al (2008) “Run-length and edge statistics based approach for image splicing detection.“ International workshop on digital watermarking, Springer, Berlin, Heidelberg

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.