Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 45 (2024), No. 3, pp. 659–675
DOI : 10.47974/JIOS-1336

Critical analysis of hybrid learning models to detect morphed images

Noble Kumari *
University School of Information, Communication and Technology
Guru Gobind Singh Indraprastha University
Delhi
India

A. K. Mohapatra †
Department of Information Technology
Indira Gandhi Delhi Technical University For Women
Delhi
India

Abstract

Machine learning (ML) algorithms produce different results with the kind of input data set or parameters passed to the algorithm. As per input data set, the analysis generally depends on the data type and amount of data set being used to train and test. Other parameters which affect the machine learning analysis are the CNN parameters or the hybrid approach being used. As there is no similar result for ML analysis on various data set, it is hard to predict which model will work most efficiently on the given data set. In this paper an analytical analysis of machine learning algorithms has been done to examine the working of various algorithms individually or in Hybrid mode. This paper studies CNN, CNN + RF, CNN +SVM approach and the theoretical parameters affecting them which helps in deciding the best suited algorithm with the given data set.

Subject Classification: *68T45, 68T05, 68M01, 68M25.*

Keywords: *Convolution neural network, Hybrid learning, Neural network, Morphing, Classification.*

1. Introduction

Multimedia images or videos have been vulnerable due to morphing attacks. These attacks have the potential to combine two or more images

* *E-mail:* **noblevashishta@gmail.com (Corresponding Author)**
† *E-mail:* **mohapatra.amar@gmail.com**

into a new image, which is hard to differentiate by a naked eye, though they are parametrically found different. These corrupted images if accepted in an ethical system may affect the whole scenario. Morphing attacks pose serious threats to the multimedia data and hence advanced algorithms have been used to detect the morphed images with efficiency.

CNN is generally used to recognize optical images which consist of Convolution Layers, Nonlinear Elements, Pooling Layers as well as of Fully Connected Layer (FC). Each layer sends processed value of input to the value of output using corresponding functions in order to accomplish the task requirement. With each layer, the most meaningful feature is retrieved and passed to the subsequent next layer. This makes semantic representation of features multidimensional with new layers.

The very first layer of the neural network retrieves the features having low level, and the subsequent higher layer extracts more features and combines them to make more abstraction in values. Each and every filter layer provides a different weight value to function the multiplication with the pixel value of the image in input to produce an image called convolution. Higher layers work for more class related features [1,2].

The basic CNN architecture contains the three general layers which are Convolution, Max_Pooling along with the Classification Layers. Convolution and Max_Pooling layers are located in the lower and the middle levels of the network. Convolution layer is identified by an even number and the Max Pooling layer is identified by an odd number. These values are then merged in a 2-dimensional plane. Convolution Layer does the feature extraction from the applied image and Max_Pooling layer performs feature abstraction using the propagation operations on the nodes. The two main processes are Feature Extraction and Classification. ConvNet is an important tool to identify faces, objects and vision [3].

Extracting features is the vital parameter which is used in the system of recognition. This mandates most distinguishable features in corresponding classes and possessing similar features in a particular class [4].

Along with the CNN models, the Hybrid models are also quite promising classification methods. The important features of Hybrid models are to automatically extract the features, to use better classifiers and to enhance the efficiency of decision parameters. CNN can be efficiently used with SVM and an RF algorithm to make the Hybrid model of CNN-SVM and CNN-RF [5].

The following sections of paper are arranged as, Section 2 of this paper provides literature survey of various forgery detection algorithms with the Hybrid approach, section 3 presents the analysis of CNN and CNN Hybrid models with CNN-SVM and CNN-RF, section 4 presents the result validation of theoretical, and implementation of above-mentioned approaches and section 5 presents the conclusion.

2. Literature Survey

Varied approaches are being proposed to detect morphed images. Seibold et. al. [6] had proposed several approaches to detect morphed images, in which a unique technique of having attack related to morphing detection is used with deep convolutional neural networks. A complete automated morphed pipeline having images of face with modifiable attributes is projected which is helpful in making the testing and training samples. Rather than the detection of classic signs of morphing which are caused by re sampling or by the double compression of JPEG, the proposed method of anti-forensics is considered here. In order to remove detector learning from these classical morphed traces, the images are processed well in advance using various functions of cropping, rotating and scaling before it feeds to network. scaling, rotating and cropping, prior to feeding them to the network. Network which are trained earlier have outclassed the networks which have received training from scratch. This way it is suggested that the feature which are learnt from classifying objects are convenient in morph attack detection.

Lu et. al. [7] discussed the image morphing attack detecting proposal using the basic patterns and the CNN. HSV, HOG and LBP features are analysed and fused with CNN in extracting the high rating level morphed traits. The method is found effective using the morphing image having the spatial features of colour and texture.

Mrudula et. al. [8] used the combination of Deep Convolution Learning along with CNN in order to facilitate in making one unit of extracting features and their classification. A multiple layer CNN having 7 layers is used to solve the aging problem. The experiment presented the better result with the proposed architecture of CNN and classifier and provided accuracy till 92.5 %.

Raja et. al. [9] proposed new approach of using the features which are transferable for the Deep Convolution Network which is already trained. The applied approach is taken care on the base of parameters of fusion of feature level of firstly connected 2 DCNN layer which are primarily tuned with the database of morphed image.

Chen et. al. [10] proposed a deep learning framework which is based on the fusion of CNN and Naiye Bayes algorithm to analyse the video frame. CNN is projected to analyse the cracks in a video and Naiye Bayes algorithm helped in discarding positive false efficiently. The applied algorithm gives 98.3% accuracy.

Putra et. al. [11] presented their research on Base (12, 13 and 21) which are retrieved from database of Messidor classifying into two sets of classes as two classes having 'mild' and 'severe' classification and class of 3 having 'severe, 'moderate' and 'mild classification. The work is found useful regarding diabetic people. The research data is reduced and relieved using Principal Component Analysis before going to the layer of classification. Support Vector Machine Algorithm-Naïve Bayes algorithm is being used in replacing full connected CNN to optimize classification and promising results are found.

Oussama et. al. [12] applied the method of Deep Random Forest using the cascade structure of multiple random forests to enhance the input feature vector with more information.

Sourabh et. al. [13] presented the result of classifying various stages of brain tumour using CNN. The results obtained from CNN are then compared with Random Forest and K Nearest Neighbour approach. During comparison CNN is found more promising to classify data sets more efficiently. CNN provided the average accuracy of 98% with the brain tumour classification.

3. Analysis of Machine Learning Parameters

This section presents the deep analysis of CNN, SVM and the RF algorithm individually and in Hybrid mode. This gives overview of all the parameters which affects the working of an algorithm as per the input data and other functions.

3.1 *CNN*

In the convolution layer, $A \times A$ based input layer is used for convolution with a $B \times B$ based filter which gives output having size of $(A-B+1) \times (A-B+1)$. Initially the raw image is convoluted using several filters retrieving several features mapped. These features are then blurred by a downsampling layer. Lastly, the set of eigen vectors are received by the fully connected layer [14].

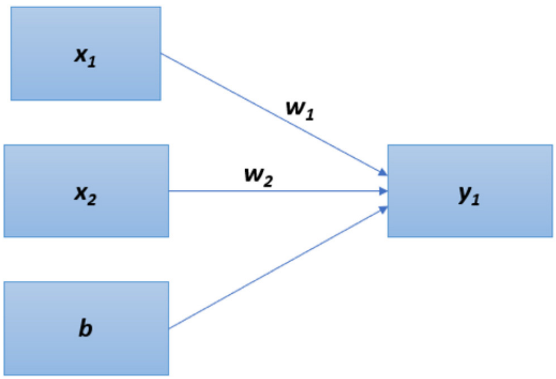

Figure 1

Perceptron having Single Layer

Activation functions being used in CNN are the core parameter to classify the data set and solve the nonlinear problem. Proper activation function maps data with better ability. Having linear properties limits the classification functionality, whereas the activation function increases the ability of expression of a neural network which is used in deep learning and has application in artificial intelligence. Figure 1 presents perceptron having single layer and without any activation function [14].

Equation 1 defines the perceptron having Single Layer without any activation function.

$$y1 = w1x1 + w2x2 + b \quad (1)$$

When y1 = 0, classification lines are obtained. The linear indivisibility issue is not handled by the perceptron having single layer and hence the perceptron having multiple layers is used to solve the multi class problem. Equation 2 presents the preceptor having multiple layers.

$$y1 = \sum_{i=1}^{n} wixi + b \quad (2)$$

As the classifier is using linear equation only so combination becomes irrelevant in dealing the non-linear system problem. To achieve that goal the activation function comes into picture. Figure 2 presents perceptron having activation function with single layer [14].

The output of above-mentioned model can be defined as:

$$y2 = w1x1 + w2x2 + b \quad (3)$$

$$y = \sigma(y2) \quad (4)$$

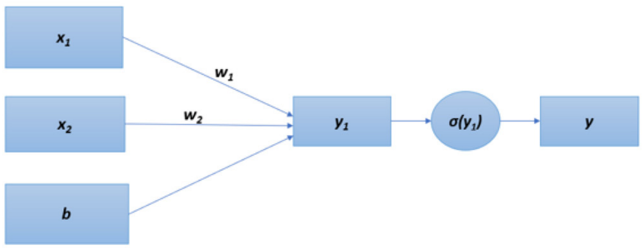

Figure 2

Perceptron with activation function on single layer.

The perceptron having any of the activation functions can work efficiently with classification of nonlinear system. This is presented by 3 and 4 equations.

Sigmoid function is a very common nonlinear activation function. Its output is bounded, and it has been used widely in deep neural networks since early ages. The sigmoid function works homologous to the synapses of biological neurons but nowadays this is rarely used. The curve of the sigmoid function has soft saturability which means the slope of graph approaches to become zero when input varies from too small to large. The gradient to the network gets modest when the slope of the function is close to 0, which makes it challenging to train the parameters. Also, the weight towards a direction update only to one as the function output is only positive. Equation 5 presents the sigmoid function [14].

$$f(x) = \frac{1}{1+e^{-x}} \tag{5}$$

Figure 3 presents Sigmoid Function.

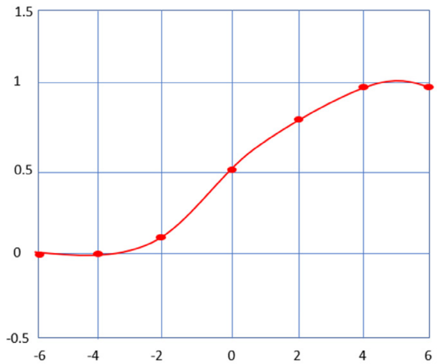

Figure 3

Sigmoid Function

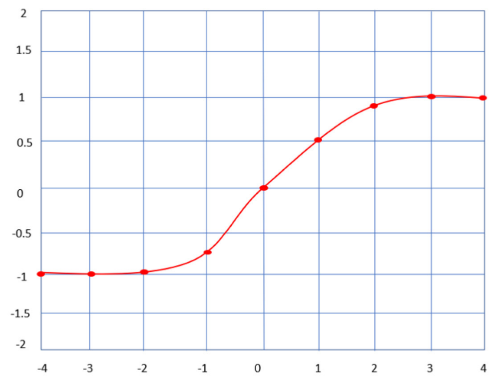

Figure 4

Tanh function

Tanh function is the advanced form of the sigmoid function being a symmetrical function which centres to zero. It also has bounded output and brings non linearity to the respective neural network. The convergence rate is higher in Tanh function then that of Sigmoid function. Equation 6 presents the Tanh function [14].

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{6}$$

Figure 4 presents the Tanh function.

Relu activation function is the trendiest function in neural networks. This function forces the retrieved output to be 0 if the value of input is not larger than 0. Otherwise, the value of output is same as input. This data forcing method to be 0 may produce moderately sparse characteristics to some limit. The function provides much faster computing. As the function is unsaturated therefore the gradient diffusion problem is unlikely to happen in the ReLu function which may occur in Tanh and sigmoid. ReLu function possess limitation in case of neuron with large gradient is passed. Equation 7 defines the function.

$$f(y) = \max(0, y) \tag{7}$$

Figure 5 presents the ReLu function [14].

Softplus function and Relu function are quite similar but the significant difference between both functions can be seen as the softplus function possesses less reservation for values which are lesser than 0 which decreases the chances of neuron death. The Softplus function does

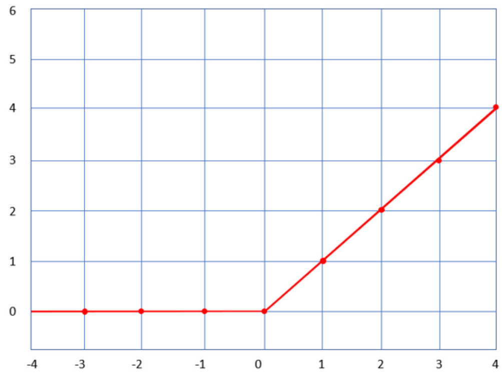

Figure 5

ReLu function

more calculation and computation than ReLu. function. Figure 6 and equation 8 present this function [14].

$$f(x) = \ln(1 + e^{x}) \tag{8}$$

In forward propagation process, convolution kernel helps in the working output of previous layer. Equation 9 presents the output of layer.

$$U_j^l = \sum_{i \in M_j} x_j^{l-1} k_{ij}^l + b_j^i \tag{9}$$

$$xij = f(u_j^l) \tag{10}$$

Where x_j^{l-1} represents feature map of output of channel *i* of previous layer, k_{ij}^l presents matrix of convolution kernel, u_j^l presents net output of

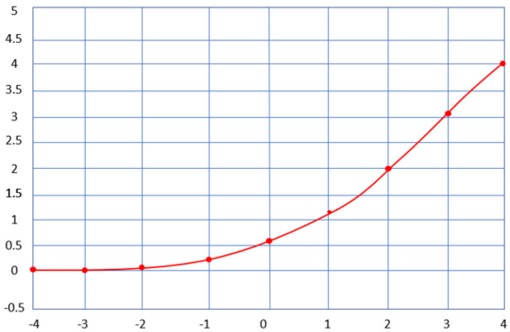

Figure 6

Softplus function

l layer which is calculated by previous layer's output feature map, output $xi_{j\,j}$ of the *l* layer and the *j* channel are received by activation function f.

Equation 9 and 10 presents that the activation function does reprocessing of convolutional operation result and does summation to the forward propagation which makes non linear relation in input and output convolutional layer. This shows that the activation function can't be any constant value and linear function. Also, activation function should be simple to speed up the process as it is used in each layer [14,15]

Characteristics of an ideal activation function [16]:

(1) To prevent the gradient, disappear while feeding it as an output to the both end data.

(2) Taking coordinates on (0,0) for centre point of symmetry in order to make gradient steady and not moving.

(3) Activation function with low cost of computation.

(4) Differential activation function to have gradient descent method for iterative training.

Generally, following properties an activation function should possess:

(1) Non-linearity: The property assures that a multiple layer network should not convert to linear network of single layer.

(2) Differentiability: This is regarding the computation of the gradient.

(3) Simple: Simple activation function is faster in speed.

(4) Saturation: It refers to the issue of making gradients close to 0 and making impossible to modify parameters.

(5) Monotonic: This guarantees that the sign of derivation should not get changed.

Hinge loss functions in training the CNN and giving stable and better accuracy of convergence. Hinge loss is generally used for training of classifiers having large margin like SVM. Equation 11 presents loss function where y represents true label and $y^{\wedge}$ represents predicted label. Hinge loss with label y > 0 and y >^ 1 for the trained example is classified rightly and has a margin beyond zero the cost is then zero and learning stops, if y < 0, the analysis is similar [17]. Thus

$$\cos t = I(y>0)\cdot\ \max(0,1-y^{\wedge})+I(y<0).\max(0,1+y^{\wedge}) \tag{11}$$

Equation 12,13 presents loss Formula of Hinge [18]

$$Ha(b)=\max\{0,a-b\} \tag{12}$$

Where, 'a' presents constant Explanation as in SVM,

$$L(y,f(x))=H1(yf(x))=\max\{0,1-yf(x)\} \tag{13}$$

The focal loss is updated from binary cross entropy. Equation 14 presents binary cross entropy [19]:

$$L(y^\wedge,y)=-y\ \log y^\wedge-(1-y)\ \log(1-y^\wedge) \tag{14}$$

$$\{=-\log y \text{ if } y=1, \text{or } \log(1-y), \text{otherwise} \tag{15}$$

Binary cross entropy is one of the types of cross entropy having prediction target 0 or 1. Sigmoid activation function is used in this to predict deep neural network [20].

Equation 16 presents standard binary cross-entropy loss function [21]:

$$J=-1\ /M(\textstyle\sum_{m=1}^{N}[ym\times\ \log(h\theta(xm))+(1-ym)\times\ \log(1-h\theta(xm))) \tag{16}$$

where M represents training number examples.

y_m represents target label for m

x_m represents input for m

h_θ performs modelling with network having weights θ

The first term, $ym\times\log(h\theta(xm))$ rectifies false negatives while training [21].

3.2 *CNN with SVM*

SVM uses supervised model of learning in machine learning which is used for regression and classification. SVM comprises linear and kernel methods. For the given set of training examples given in equation 19, the SVM function solves the problem with the function as mentioned in equation 18.

$$(x_i,y_i)\text{for } i \text{ from} 1 \text{to} n \tag{17}$$

and Kernel function K,

each y_i = {−1, +1} is taken in one of two categories. Equation 18 presents object function of SVM which is used to compute the optimization problem:

$$\max_{a}\left\{\sum_{i=1}^{n} a_i + \sum_{i,j=1}^{n} a_i a_j y_i y_j K(x_i, x_j)\right\} \tag{18}$$

$$s.t. 0 \le \alpha i \le C \tag{19}$$

$$\sum_{i=1}^{n} a_i \; yi = 0 \tag{20}$$

Here *αs* represents Lagrange coefficients

C presents a constant which penalizes the sample training errors.

SVM classifier works on the basis of making a model which assigns data into one of the categories which makes it a non-probabilistic binary linear classifier. SVM classifier represents points in space mapping the categories with a visible gap as much as possible. New data sets are then linked to that space and get categories on the basis of which side they fall on. Working of SVM depends on the regularization parameters which are *αs* and the form of kernel function *K(xi, x j)* and the result of classification depends on them [22].

Hybrid SVM-CNN model takes the best qualities of both the algorithms. CNN contains supervised learning with multiple full connected layers.

CNN works with the same mechanism as human being does and retrieves most discriminating information from the raw set of data. This output of CNN having an effective sub region from the raw image is then used by SVM to represent a multidimensional dataset into different classes which are differentiated by the hyperplane.

SVM classifier has the advantage of minimizing the generalized error on non-visible data. Optimal hyperplane is used for the classification of data sets. SVM is found good with the binary classification but considered poor for noisy data. In a Hybrid structure of CNN-SVM, the softmax layer is being replaced by SVM binary classifier where feature extraction is done by CNN and binary classification is done by SVM. The output of each layer in turn becomes an input to the subsequent layer. The CNN algorithm

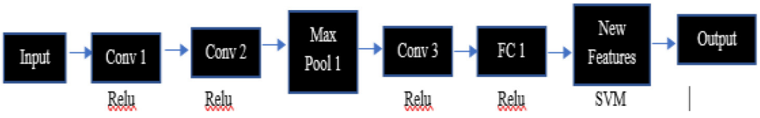

Figure 7

Hybrid CNN-SVM structure [23]

is trained after several required epochs or till the time optimum training process is not converged. Then, SVM classifier replaces the last layer of CNN and the final result is achieved [23]. Figure 7 presents the Hybrid CNN-SVM structure.

3.3 *CNN with RF*

Random forest algorithm is considered as one of the best algorithms for classification due to its interdisciplinary nature. The algorithm possesses the advantage of accessing continuous or categorical data sets, not sensitive nature in overfitting and good at dealing with outliers in training data [24].

The algorithm works on the basis of a decision tree using predictive models that uses binary rules to calculate targets.

Decision trees pose the advantage of classification of data as classification is fast, easy to interpret, no need to have unimodal distribution of data.

Decision Tree possesses' disadvantage of being over fit which gives poor accuracy results when applied to whole data. To remove this issue 'pruning' of trees is done to reduce terminal nodes but it is also a cumbersome method.

Random Forest possesses some limitations when used for regression. Regression trees have the constraint of response range and hence random forest does not work efficiently if the input data does not cover the full range [24].

Ensemble classifications are learning algorithms used to construct a multiple set of classifiers instead of single classifier. It then classifies newly found data points using vote of predictions [25].

Bagging, Boosting and Random Forest (RF) are the types of ensemble classifiers. From training data, the Bagging method generates bootstrap samples and trees are constructed from them, successively the earlier trees are used to make new trees independently.

Boosting uses the method of iterative re-training classification and increasing weights with each iteration to increase weightage subsequently. But the process has the limitation of being slow, overtrained and being sensitive to noise.

RF classifier is the tree structured classifiers being advanced in Bagging by adding randomness to it. It does not split each node whereas it follows the method of best splits using random predictors at that node. This way the new set of data keeps on replacing the older one and the tree grows with random feature selection. If the grown-up tree is not pruned accurately, RF can become inefficient. RF possesses the advantage of being fast, robust against overfitting and making user dependent number of trees. RF needs two parameters for initialization, N represents number of grown-up trees and m represents variables numbers to split each node. Out of N samples 2/3 are used for training data sets and remaining are used to test the error of prediction. Then un pruned trees grow and for each node m, the predictors have been selected to have the best split among the chosen variables. Choosing the number of variables is a crucial step to find the optimum result by providing low correlation with adequate predictive power. Breiman suggested setting m equals to square root of M gives the nearly optimum result.

RF utilizes Classification and Regression tree called CART algorithms where trees are created and split is done at each node using GINI index. GINI index calculates the class homogeneity, equation 21 represents the GINI index as:

$$\sum\sum_{j\neq i}(f(\mathcal{C}_i,T)/|T|)(f(\mathcal{C}_i,T)/|T|) \tag{21}$$

Where, T represents training set which is given

Ci represents class which selects pixel randomly

f $\mathcal{C}$ i,T/|T| suggests the probability of belonging to class Ci.

GINI index is directly proportional to the heterogeneity class. If a child has a GINI index which is lesser than that of the parent then the split becomes successful. Splitting of tree stops when GINI index becomes 0 which means homogeneous class. When all the 'N' trees are grown up, prediction on new data is done using these trees [26].

The combination of CNN and RF model works on the basis of parameters of both algorithms. In a hybrid model, high level features that are taken from the CNN algorithm are classified using RF. This process considers the advantage of both the algorithms. Hybrid algorithm uses features retrieved from potential parameters of CNN used as the input for classification. And hence no input extraction is required at RF based classification. Although RF employs more complex categorization techniques like bagging and being robust to outliers and capable of extracting informative spatial features from data sets where CNN may lag due to insufficient training or input image.

The hybrid model uses the CNN network topology, but some parameters need to be changed using RF classification because the RF classifier takes the CNN output as input. Figure 8 demonstrates the full working of CNN-RF model.

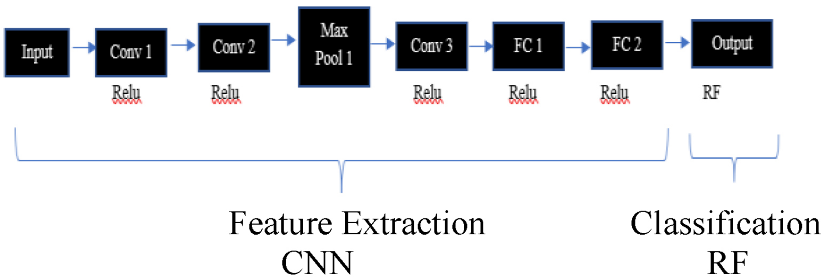

Figure 8

CNN-RF [26]

CNN-RF outperforms when fewer images or data sets are smaller over CNN and RF [26].

4. Analysis of theorotecial computation with implementation

This section analyses and validates the mathematical analysis of algorithms with the proposed result of Noble et. al. [27]. CNN-SVM and CNN-RF performed differently with the provided data set, according to the implementation results using the public data set. Therefore, for epochs 10 and 25, CNN+SVM performed better than CNN algorithm, however for epochs 40, 60, and 80, CNN algorithm performed better. Because the CNN+SVM method uses a linear function, it takes less time on average than the CNN algorithm. The most accurate result was produced by the model using CNN and Random Forest in combination on CASIA V2 data, which had a 99.18% accuracy rate and required 45 seconds to run. The model using CNN and Random Forest in combination on DVMM data had a result with 52.43% accuracy and required 25 seconds to run, demonstrating the concept of getting low accuracy when using many low range data sets for training and testing [27]. In general, the kind of data being used in testing and training along with the functions used in neural network decides the need for an algorithm to be applied for classification.

5. Conclusion

In this paper, a deep theoretical analysis of image forgery detection is done by CNN, CNN+SVM and CNN + Random Forest approach using extensive mathematical equations. It has been found that a different data set and input type requires different algorithms for efficient classification.

CNN works efficiently in most of the cases. However, CNN+SVM supports better in the case of binary classification, and CNN+RF outperforms in a small number of training and test data sets to make a decision tree. The CNN working methodology also depends on the kind of activation function used. Hence, we can conclude that the role of data dependency is a very crucial parameter, and acts as a pivotal standpoint before deciding a classification algorithm.

References

[1] Chen, Yan, and Rui Qu. "Study on infringement identification of art works based on CNN image recognition technology." *Journal of Physics: Conference Series.* Vol. 1802. No. 3. IOP Publishing (2021).

[2] Girish, Deeptha, Vineeta Singh, and Anca L. Ralescu. "Unsupervised clustering based understanding of CNN." CVPR Workshops (2019).

[3] Makkar, Tanya, et al. "Analogizing time complexity of KNN and CNN in recognizing handwritten digits." 2017 Fourth International Conference on Image Information Processing (ICIIP). IEEE (2017).

[4] Niu, Xiao-Xiao, and Ching Y. Suen. "A novel hybrid CNN–SVM classifier for recognizing handwritten digits." Pattern Recognition 45.4: 1318-1325 (2012).

[5] Knauer, Uwe, et al. "Tree species classification based on hybrid ensembles of a convolutional neural network (CNN) and random forest classifiers." Remote Sensing 11.23 : 2788 (2019).

[6] Seibold, Clemens, et al. "Detection of face morphing attacks by deep learning." International Workshop on Digital Watermarking. Springer, Cham (2017).

[7] Lu, Yi, et al. "Face morphing detection with convolutional neural networks based on multi-features." Proceedings of the 2020 International Conference on Aviation Safety and Information Technology (2020).

[8] Nimbarte, Mrudula, and Kishor Bhoyar. "Age Invariant Face Recognition using Convolutional Neural Network." *International Journal of Electrical & Computer Engineering,* 2088-8708, 8.4 (2018).

[9] Raja, Kiran, Sushma Venkatesh, and R. B. Christoph Busch. "Transferable deep-cnn features for detecting digital and print-scanned morphed face images." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2017).

[10] Chen, Fu-Chen, and Mohammad R. Jahanshahi. "NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion." *IEEE Transactions on Industrial Electronics* 65.5 : 4392-4400 (2017).

[11] Putra, Ricky Eka, Handayani Tjandrasa, and Nanik Suciati. "Severity classification of non-proliferative diabetic retinopathy using convolutional support vector machines." *International Journal of Intelligent Engineering and Systems* 13.4 : 156-170 (2020).

[12] Guehairia, Oussama, et al. "Deep random forest for facial age estimation based on face images." 2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP). IEEE (2020).

[13] Hanwat, Sourabh, and Chandra Jayaraman. "Convolutional neural network for brain tumor analysis using MRI images." *Int. J. Eng. Technol* 11 : 67-77 (2019).

[14] Wang, Yingying, et al. "The influence of the activation function in a convolution neural network model of facial expression recognition." Applied Sciences 10.5 (2020): 1897.

[15] Feng, Jianli, and Shengnan Lu. "Performance analysis of various activation functions in artificial neural networks." *Journal of Physics: Conference Series.* Vol. 1237. No. 2. IOP Publishing (2019).

[16] Hao, Wang, et al. "The role of activation function in cnn." 2020 2nd International Conference on Information Technology and Computer Application (ITCA). IEEE (2020).

[17] Jin, Junqi, Kun Fu, and Changshui Zhang. "Traffic sign recognition with hinge loss trained convolutional neural networks." IEEE transactions on intelligent transportation systems 15.5 : 1991-2000 (2014).

[18] Wang, Qi, et al. "A comprehensive survey of loss functions in machine learning." Annals of Data Science 9.2 : 187-212 (2022).

[19] Luo, Xiong, et al. "Ophthalmic disease detection via deep learning with a novel mixture loss function." *IEEE Journal of Biomedical and Health Informatics* 25.9 : 3332-3339 (2021).

[20] Ruby, Usha, and Vamsidhar Yendapalli. "Binary cross entropy with deep learning techniques for image classification." *Int. J. Adv. Trends Comput. Sci. Eng* 9.10 (2020).

[21] Ho, Yaoshiang, and Samuel Wookey. "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling." *IEEE Access* 8 : 4806-4813 (2019).

[22] Liu, Peng, et al. "SVM or deep learning? A comparative study on remote sensing image classification." *Soft Computing* 21.23 : 7053-7065 (2017).

[23] Ahlawat, Savita, and Amit Choudhary. "Hybrid CNN-SVM classifier for handwritten digit recognition." Procedia Computer Science 167 : 2554-2560 (2020).

[24] Horning, Ned. "Random Forests: An algorithm for image classification and generation of continuous fields data sets." Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan. Vol. 911 (2010).

[25] Akar, Özlem, and Oguz Güngör. "Classification of multispectral images using Random Forest algorithm." *Journal of Geodesy and Geoinformation* 1.2 : 105-112 (2012).

[26] Kwak, Geun-Ho, et al. "Potential of hybrid CNN-RF model for early crop mapping with limited input data." Remote Sensing 13.9 : 1629 (2021).

[27] http://paper.ijcsns.org/07_book/202206/20220646.pdf

Received October, 2022