DEEPFAKE DETECTION TO PREVENT FALSE PROPAGANDAS USING ML ALGORITHMS: A COMPARATIVE STUDY

[1] TANMAY SRIVASTAV, [2] JUHI JAIN, [3] PRAGATI GUPTA, [4] ABHINAV PRATAP SINGH

[1,3,4] Student, Amity University, Noida
[2] Assistant Professor-III, Amity University, Noida
E-mail: [1] tanmaysri07@gmail.com, [2] erjuhijain@gmail.com, [3] gpragati18.off@gmail.com,
[4] pratapabhinavofficial@gmail.com

Abstract - Manipulation of images and videos has been made easier by deepfake technology. This manipulated content is almost indistinguishable from ordinary vision. Many algorithms have been developed and are still being developed every day to detect deepfake content. Approaches followed by these algorithms may include visual, local, deep, and temporal feature-based approaches. The primary debacle faced by the researchers in coming up with deepfake detection algorithms is the changing dynamics of the models that are used to generate this deepfake content, which keeps making deepfake videos closer to real life. In this paper, we have very meticulously analysed the working mechanisms of the three main models for deepfake detection, namely ResNeXt50 + LSTM, EfficientNet, and MesoNet. The algorithms are compared on various performance measures, such as accuracy, precision, recall, and F1 scores. Conclusively, EfficientNet, with an accuracy of 97.5 % surpassed all the other algorithms in deepfake detection.

Keywords - Deepfake detection, LSTM, ResNeXt, EfficientNet, MesoNet

I. INTRODUCTION

Deepfake content is generated by superpositioning images or audio over existing media. This is done by leveraging the techniques of Machine Learning and Neural Networks to generate visual and audio content that could be later used for deception and blackmailing. The main algorithms are based on Deep Learning(DL) and involve autoencoders, or Generative Adversarial Networks(GANs) [1]. Over the past few years, deepfakes have found infamy owing to their use in porn, financial fraud and fake news. This has brought over scrutiny from the industry as well as governments from all over the world. The penetration of deepfakes into the entertainment and gaming industry has led to widespread disruption in these industries.

Deepfake videos can be detected by identifying recurring feature similarities between different frames of the content. For example, there may be a single camera recording the video, and the facial and eye movements may be the bare minimum. This type of approach only focuses on one aspect of deepfake technology, but there may be many more things that are required in the creation of new deepfake videos as the technology to create deepfake content is evolving at the same pace as deepfake detection technologies[2].

In this paper, we are trying to answer two research questions-

RQ1: What are the accuracies, precision, recall and f1 score of individual algorithms?

RQ2: Which among these algorithms performs the best, based on the metrics used?

The first algorithm used in this paper is ResNeXt50+LSTM. First, the ResNeXt50 algorithm breaks the video into different frames and then using the LSTM algorithm, we analyze various facial features of the subject of the video and based on the facial activity, we detect whether the video is real or manipulated.

We used MesoNet (Meso-4), a potent neural network that combines classification and localization with region proposal development. MesoNet has shown outstanding performance in object identification, especially for tiny objects. It is a big development in deep learning and has real-world uses in surveillance and autonomous driving, among other things.

The third algorithm we used for comparison was EfficientNet. EfficientNet is a collection of models for convolutional neural networks that have been created to attain top-of-the-line performance using significantly fewer parameters and computation in comparison to other popular models, such as ResNet and Inception. To achieve efficiency, the EfficientNet family uses a scaling technique that ensures a balanced depth, width, and resolution of the network's architecture. This results in models that exhibit high accuracy for image classification tasks with a reduced number of parameters, making them more efficient for training and deployment.

II. LITERATURE REVIEW

Deep learning forms the basis of most of the models proposed to detect deepfake content. In Zhang et al.[3], a Generative Adversarial Network(GAN) Simulator, called AutoGAN was proposed. The GAN simulator duplicates the aggregate GAN image

artefacts and utilizes them as input to the classifier to clearly point out deepfake images.

In Zhou et al.[4], a network that is meticulous in tampering with artefacts rather than content (image) in a manipulated image was put forth. A two-stream Faster R-CNN model was proposed that could detect tampering in an image, given the tampered image. The first stream was an RGB Stream which would look for strong contrast differences, unnatural tampered boundaries and so on in the image to find tampering artefacts. The second stream was a noise stream that used a steganalysis-rich model filter to extract noise features to identify inconsistencies between authentic and manipulated images. Features from both these models are fused using a bilinear pooling layer.

Chugh et al.[5] proposed a Modality Dissonance Score(MDS), based on the dissimilarities between the audio and visual modalities. Manipulation of either of the modalities would lead to disharmony between the two modalities, eg- loss of any synchronized lip movement, face and movements that are not natural, etc. MDS is an average of dissimilarity scores between audio and visual channels. Using cross-entropy loss for individual modalities and contrastive loss for inter-modality similarity, modality scores for audio and visual segments are determined separately and chunk-wise.[6]-[8] proposed detection frameworks based on physiological measurements, such as heartbeat. Afchar et al.[9] proposed two inception models- (i) Meso-4 and (ii) MesoInception-4, one of which is a model that we have used in our paper for the comparative study. The mean squared error (MSE) between the actual and predicted labels makes up this model's loss function.

Authors in [10] improved upon MesoNet by changing the actual activation function permitting a 1% improvement and the consistency of results also increased. A new activation function was also created- Pish. Even though it has a time overhead, it provides higher consistency.

Authors in [11] have highlighted that deep CNNs[12]-[14], outperform shallow CNNs. Various features are extracted by different algorithms, including extraction of handcrafted features[15]-[16], spatiotemporal features[17]-[20], common texture[21],[22] and face landmarks[23]-[25]. Input for deepfake detection comprises these extracted features.

Dang et al.[26] introduced an attention mechanism. Instead of simultaneously detecting deepfaked images and predicting the deep faked mask using multi-task learning, we utilize the attention mechanism method and enhance feature maps for classification. The attention map not only improves binary classification but also visualizes the manipulated regions. [27]-[28] proposes an architecture named capsule-network(CN) that uses a number of parameters which is comparatively lesser than its alternative.

Reference No.	Techniques used	Dataset Used	Performance Measure Used	Result
[3]	GAN Simulator	CycleGAN Images	Accuracy using leave-one-out setting	97.2% accuracy on A-spec model based on AutoGAN
[4]	Extraction of noise and standard RGB data for detection of deepfake	NIST16 Dataset	F_1 Score	F_1 Score of 0.677/0.689 with a compression quality of 70 and resizing scale of 0.7
[5]	Modality Dissonance Score (MDS)	DFDC Dataset	Accuracy Score	91.54% accuracy on DFDC Dataset
[9]	Meso-4 and MesoInception-4	Deepfake and Face2Face Dataset	Aggregation Score	Aggregation score of 98% for deepfake and 95% for Face2Face
[26]	Attention mechanism (Xception + Reg.)	UADFV Dataset	AUC	98.4% AUC in UADFV Dataset
[27]	Capsule-Network(CN)	Deepfake Dataset	Accuracy	Accuracy frame level-95.93%. Accuracy at video level-99.23%.

Our work	LSTM+ResNeXt50 Meso4 EfficientNetAutoAttB4 EfficientNetB4 EfficientNetB4ST EfficientNetAutoAttB4ST	FaceForencics++	Accuracy	Accuracy of 86.43% on LSTM+ResNeXt50, 90% on MesoNet, 97.5% on EfficientNetAutoAttB4, 94.5% on EfficientNetB4, 94% on EfficientNetB4ST and 92.5% on EfficientNetAutoAttB4ST.

Table 1: Comparison between studies

III. RESEARCH METHODOLOGY

A. Dataset Used

We have extracted data from four datasets and created a unique dataset of our own containing data from these four datasets. These four datasets include ImageNet, FaceForensics++, Celeb-DF and data from Kaggle Database.

ImageNet- Millions of labelled photos make up the enormous dataset known as ImageNet, which is frequently used in the training of deep neural networks. It was developed to set a uniform benchmark for object recognition algorithms as part of the Large Scale Visual Recognition Challenge (LSVRC). Modern models including AlexNet, VGGNet, GooLeNet, and ResNet have been trained using ImageNet throughout the years and have excelled in a variety of computer vision tasks like image classification, object identification, and picture segmentation.

FaceForensics++- It is a repository consisting of 1000+ videos that were manipulated using Deepfake, Face2Face, FaceSwap and NeuralTextures Algorithm. The database has been borrowed from over 1000 publicly available videos from different social media platforms. This allowed us to use a second-level synthesis derived from the Heuristic Error Confirmation Network (HEAR) to produce face swap results that preserve high-fidelity identity and handle face occlusion. GANs and Neural Textures were also used to manipulate content in this database.

Celeb-DF- Celeb-DF contains a large Dataset that could be used for DeepFake Forensics. Celeb-DF contains over 590 authentic videos from social media sites with a plethora of features to include data points from all over the world.

Kaggle Database- A competition's dataset whose full training set is just over 470 GB. We have taken small chunks of it each of size ~ 10 GB. All of this data combined into one big database is effective for our preprocessing and model training.

B. Preprocessing

The 'face_recognition' library is used to take frames from each movie and process them during the video preprocessing. Each video's first 150 frames are the only ones taken into account. Each frame's face coordinates are determined using the library, and after any required scaling, the locations are recorded in batches. The new films that arise from cropping the faces in each frame are stored in the 'M', 'J', 'P', and 'G' formats, which are used to reduce the video file size and specify the video format. The preparation procedure is finished by searching for and removing any damaged video files.

C. Algorithm Used

Long Short Term Memory (LSTM)- Traditional RNNs excel at jobs with short-term dependencies but suffer from deep fake detection tasks that require analysing frames with significant temporal gaps. To address long-term dependencies in sequential data, LSTM models were created as a solution[29].
RNNs form continuous sequences of recurring modules of neural networks with a single tanh layer, as shown in figure 1.

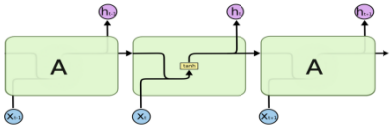


Figure 1: Repeating Modules of Traditional RNN[29]

LSTM comprises four neural network layers in a module, as shown in figure 2.

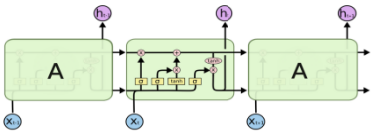


Figure 2: Repeating Modules of LSTM[29]

The single horizontal line that passes across the top of the module and represents the cell state is the most crucial component of the LSTM. With just slight linear interactions, the cell state permeates the whole chain. Gates(figure 3) control the information flow via the cell state, which is where the long-term dependencies are stored.



Figure 3: Sigmoid Layer Gate[29]

Residual Networks (ResNet)- The vanishing gradient issue in conventional CNNs was addressed by the introduction of residual blocks. By using skip connections(figure 4), they provide the convolution blocks as a middle input. The gradient flow during backpropagation is improved by this method. These Residual Blocks are stacked to create ResNet networks[30].

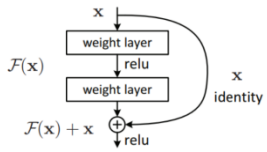


Figure 4: Skip Connections[30]

The ResNeXt model incorporates the ResNet strategy of repeating layers while also introducing an extensible, simple method of implementing the split, transform, and merge strategy.

The building block is similar to the Inception network in that it supports different transformations such as 1x1 Conv, 3x3 Conv, 5x5 Conv, and MaxPooling. However, while these transformations are stacked together in the Inception model, the ResNeXt model adds and merges them.

MesoNet (Meso4)- A deep learning technique called MesoNet is intended to categorise altered photos and videos by concentrating on minute mesoscopic characteristics. It makes use of three convolutional neural networks, trained on large datasets, to examine motion patterns, picture features, and color/texture patterns to discern between authentic and false material. MesoNet is an effective tool for preventing the spread of false information and its potential to have an impact on public opinion since it outperforms other algorithms in its capacity to identify deepfakes[9].

EfficientNet- In order to increase accuracy, traditional CNNs are frequently scaled by changing variables such as resolution, depth, or breadth. However, this method frequently necessitates human tinkering and might provide less-than-ideal outcomes. Compound Model Scaling, an ethical strategy to enhance accuracy and efficiency by optimising the scaling of CNN models, was offered as a solution to this problem. To balance performance and resource use, it takes a number of aspects into account[31].
Research shows that achieving a balance between network dimensions and available resources enhances overall model performance. A grid search is used to independently scale each dimension and determine the optimal coefficients for improvement. By adding these coefficients to the baseline network, the desired model size or computational budget is achieved. Compound scaling outperformed traditional techniques, improving MobileNet's accuracy by 1.4% and ResNet's accuracy by 0.7% on ImageNet.

Through the use of the AutoML MNAS Framework and a Neural Architecture Search (NAS), the EfficientNet model is created from a base network. The precision and efficiency of this search are to be maximised. The EfficientNet model serves as the base and is scaled to provide the EfficientNets family of models.

D. Performance Measures Used

We have used 4 parameters to measure performance-accuracy, F1 Score, recall and precision. The aggregate performance of the algorithm is measured by accuracy, which is the proportion of correct predictions in all predictions.

Precision measures the proportion of a model's genuine positive predictions to all of its positive predictions (i.e., correct positive predictions). It reveals how accurately the model identifies true positives while avoiding false positives by not misidentifying negatives.

Out of all actual positive cases in the dataset, the proportion of true positive predictions is called recall, though it is also referred to as sensitivity or true positive rate. This metric reveals how successfully the model is able to recognize any and all positive cases without overlooking any.

The F1 score combines precision and recalls into a single statistic, producing a fair assessment of model performance when identifying genuine positives and true negatives. The F1 score is calculated using the harmonic mean of accuracy and recall, a comprehensive combination of metrics. A low F1 score suggests an opportunity for development in both accuracy and memory, whereas a high score indicates more model precision and recall.

IV. EXPERIMENTAL SETUP

In order to preprocess videos for deepfake detection, frames from the video are extracted, and faces are identified. Training and testing sets of the data have been separated. ResNeXt50(figure 6) is used by the deepfake detection model to extract face characteristics, and LSTM to determine if videos are real or fake. The trained model is exported after the model's performance is assessed using a confusion matrix. After that, it is employed to judge the veracity of random videos.

A collection of actual and fraudulent videos is loaded as part of the MesoNet operation(figure 7). The films are divided into frames, the faces are extracted, and the cropped videos are saved as part of the data preprocessing. After splitting the preprocessed data, the labels are loaded. Each video's likelihood of being real or false is predicted using the Meso4 classifier

with pre-trained weights, and the accuracy is calculated. The dataset of actual and false films is fed into the EfficientNet pipeline(figure 8) and preprocessed by dividing the videos into frames and removing faces. EfficientNet models that have already been trained are loaded once the preprocessed data is labelled. On the basis of the data for each architecture, predictions are made, and the average prediction scores for each movie are computed. Each architecture's correctness is calculated, and models are contrasted based on accuracy ratings.

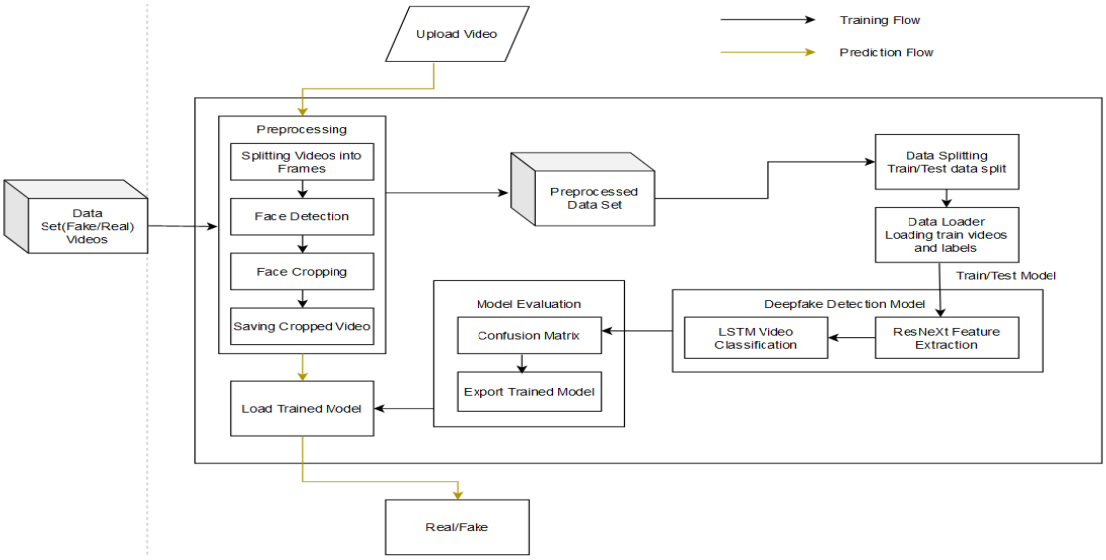


Figure 5: LSTM+ResNeXt50

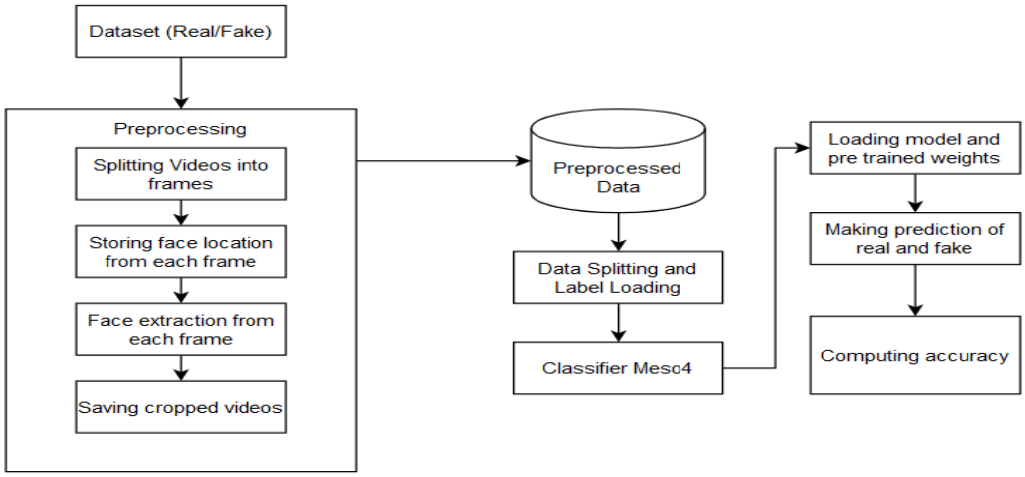


Figure 6: MesoNet(Meso4)

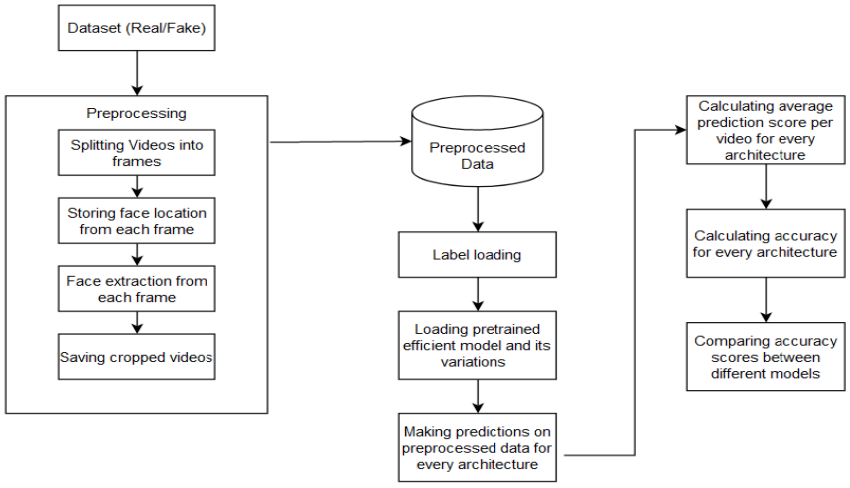


Figure 7: EfficientNet

V. RESULTS

	Accuracy	Precision	Recall	F_1-score
ResNeXt50+ LSTM	86.43	0.925	0.8157	0.867
MesoNet	90	0.924	0.968	0.945
EfficientNetAutoAttB4	97.5	0.97	0.995	0.982
EfficientNetB4	94.5	0.95	0.995	0.972
EfficientNetB4ST	94	0.948	0.989	0.968
EfficientNetAutoAttB4ST	92.5	0.938	0.983	0.959

Table 2: Performance of Different Models

As evident from the comparison graph between the models we have tested in figure 9 and table 2, EfficientNetB4 and EfficientNetB4ST have similar accuracies. If the variations of the EfficientNet model are considered, they are much more compact than the MesoNet model as the EfficientNet model is pre-trained on the ImageNet dataset whereas the MesoNet model has to be trained using the custom dataset that we have. EfficientNetB4 has an accuracy which is less than the EfficientNetAutoAttB4 model. EfficientNetB4ST is more accurate than EfficientNetAutoAttB4ST, which in itself is more accurate than MesoNet. On the other hand, ResNeXt50+LSTM has the lowest accuracy amongst all considered models.

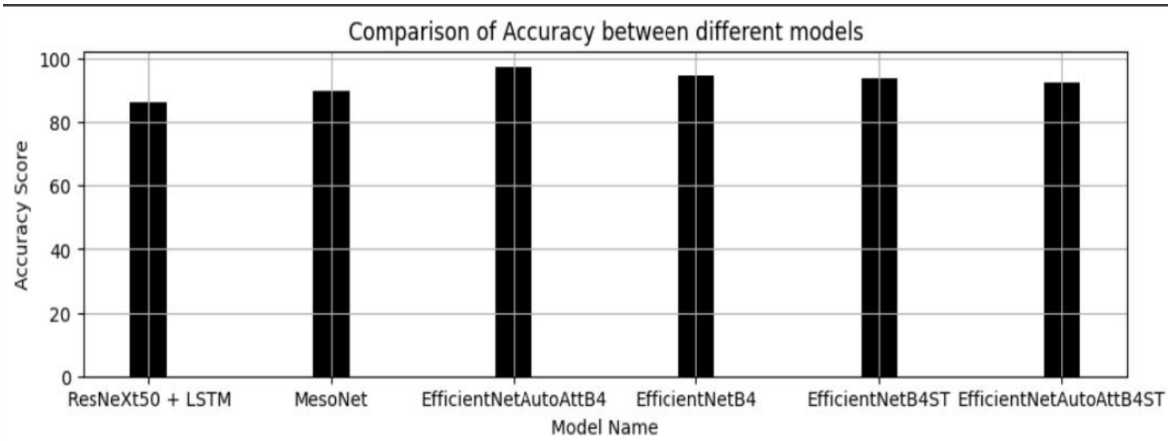


Figure 9: Comparison of Accuracy between Models

In the case of the precision score(figure 10 and Table 2), all the models are nearly equal with ResNext50+LSTM and MesoNet having a slightly lower precision score as compared to all the variations of EfficientNet.

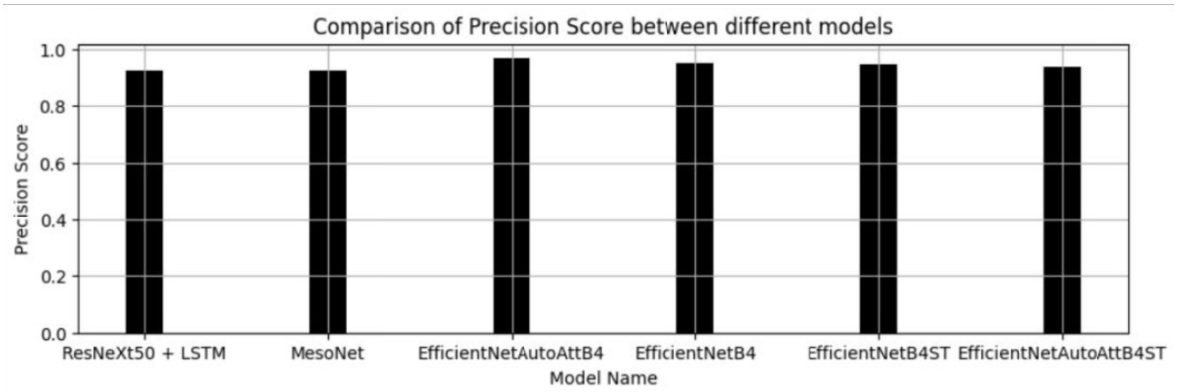


Figure 10: Comparison of Precision score between different models

EfficientNetAutoAttB4 has the highest recall score among all the variations of EfficientNet, MesoNet has a slightly lower recall whereas ResNext50+LSTM has the lowest recall, as evident from Figure 11 and Table 2.

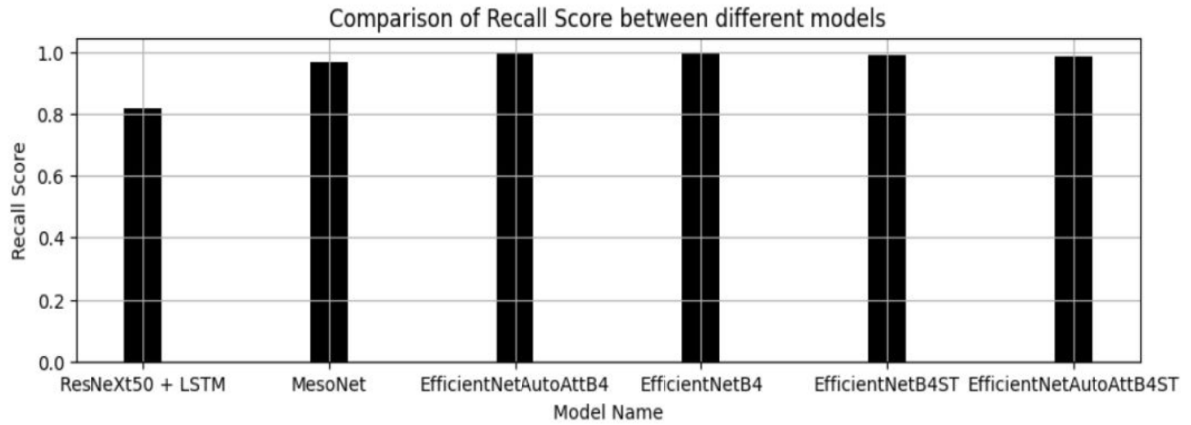


Figure 11: Comparison of recall score between different models

EfficientNetAutoAttB4 has the highest F_1 score whereas ResNext50+LSTM has the lowest F_1 score, as evident from Figure 12 and Table 2. It is known that a low F1 score has very poor precision and poor recall.

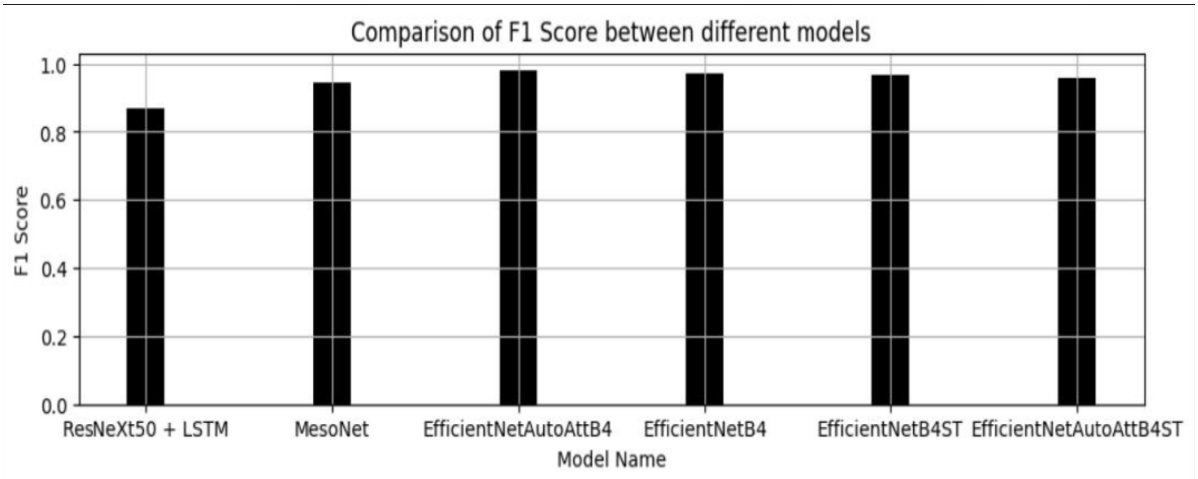


Figure 12: Comparison of F_1 Scores between different models

VI. CONCLUSION

It could be clearly stated that after assessing various models on a number of parameters as shown above, the EfficientNet model with an accuracy of 97.5% and 0.97 precision stands out, MesoNet offers appreciable performance, whereas ResNeXt50+LSTM is a bit inefficient. It is to be duly noted that Mesonet might not be as good as EfficientNet, but its performance is quite satisfactory. A substantial amount of work has already been done in the field of deepfake detection and is on the continuum. It is highly plausible that other architectures will soon be developed that offer higher efficiency than the above-mentioned models.

REFERENCES

[1] A. Tiwari, R. Dave and M. Vanamala, "Leveraging Deep Learning Approaches for Deepfake Detection: A Review", 2019.

[2] B. Dolhansky, R. Howes B. Pflaum, N.Baram, C. Ferrer, "The Deepfake Detection Challenge (DFDC) Preview Dataset",2019.

[3] X. Zhang, S. Karaman, and S.-F. Chang, ''Detecting and simulating artifacts in GAN fake images,'' in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2019, pp. 1–6.

[4] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, ''Learning rich features for image manipulation detection,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, Jun. 2018, pp. 1053–1061, doi: 10.1109/CVPR.2018.00116

[5] K. Chugh, P. Gupta, A. Dhall, and R. Subramanian, ''Not made for each other- audio-visual dissonance-based deepfake detection and localization,'' 2020, arXiv:2005.14405.

[6] H. Qi, Q. Guo, F. Juefei-Xu, X. Xie, L. Ma, W. Feng, Y. Liu, and J. Zhao, ''DeepRhythm: Exposing deepfakes with attentional visual heartbeat rhythms,'' 2020, arXiv:2006.07634.

[7] S. Fernandes, S. Raj, E. Ortiz, I. Vintila, M. Salter, G. Urosevic, and S. Jha, ''Predicting heart rate variations of deepfake videos using neural ODE,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 1721–1729.

[8] J. Hernandez-Ortega, R. Tolosana, J. Fierrez, and A. Morales, ''DeepFakesON-phys: DeepFakes detection based on heart rate estimation,'' 2020, arXiv:2010.00400.

[9] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, ‘‘MesoNet: A compact facial video forgery detection network,’’ in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2018, pp. 1–7.

[10] P. Kawa and P. Syga, ‘‘A note on deepfake detection with low-resources,’’ 2020, arXiv:2006.05183.

[11] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, ‘‘FaceForensics++: Learning to detect manipulated facial images,’’ in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 1–11.

[12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ‘‘Densely connected convolutional networks,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[13] F. Chollet, ‘‘Xception: Deep learning with depthwise separable convolutions,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1800–1807.

[14] A. Khodabakhsh and C. Busch, ‘‘A generalizable deepfake detector based on neural conditional distribution modelling,’’ in Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, Sep. 2020, pp. 1–5.

[15] Y. Li, M.-C. Chang, and S. Lyu, ‘‘In ictu oculi: Exposing AI created fake videos by detecting eye blinking,’’ in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2018, pp. 1–7.

[16] Y. Li and S. Lyu, ‘‘Exposing deepfake videos by detecting face warping artifacts,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2019, pp. 46–52. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2019/html/Media_Forensics/Li_Exposing_DeepFake_Videos_By_Detecting_Face_Warping_Artifacts_CVPRW_2019_paper.html

[17] I. Ganiyusufoglu, L. Minh Ngô, N. Savov, S. Karaoglu, and T. Gevers, ‘‘Spatio-temporal features for generalized detection of deepfake videos,’’ 2020, arXiv:2010.11844.

[18] A. Singh, A. S. Saimbhi, N. Singh, and M. Mittal, ‘‘Deepfake video detection: A time-distributed approach,’’ SN Comput. Sci., vol. 1, p. 212, Jun. 2020, doi: 10.1007/s42979-020-00225-9.

[19] I. Kukanov, J. Karttunen, H. Sillanpää, and V. Hautamäki, ‘‘Cost sensitive optimization of deepfake detector,’’ 2020, arXiv:2012.04199.

[20] A. Haliassos, K. Vougioukas, S. Petridis, and M. Pantic, ‘‘Lips don’t lie: A generalisable and robust approach to face forgery detection,’’ 2020, arXiv:2012.07657.

[21] X. Zhu, H. Wang, H. Fei, Z. Lei, and S. Z. Li, ‘‘Face forgery detection by 3D decomposition,’’ 2020, arXiv:2011.09737.

[22] X. Wang, T. Yao, S. Ding, and L. Ma, ‘‘Face manipulation detection via auxiliary supervision,’’ in Neural Information Processing (ICONIP) (Lecture Notes in Computer Science), vol. 12532, H. Yang, K. Pasupa, A. C. Leung, J. T. Kwok, J. H. Chan, I. King, Eds. Cham, Switzerland: Springer, 2020, pp. 313–324, doi: 10.1007/978-3-030-63830-6_27.

[23] M. T. Jafar, M. Ababneh, M. Al-Zoube, and A. Elhassan, ‘‘Forensics and analysis of deepfake videos,’’ in Proc. 11th Int. Conf. Inf. Commun. Syst. (ICICS), Irbid, Jordan, Apr. 2020, pp. 053–058, doi: 10.1109/ICICS49469.2020.239493.

[24] X. Dong, J. Bao, D. Chen, W. Zhang, N. Yu, D. Chen, F. Wen, and B. Guo, ‘‘Identity-driven deepfake detection,’’ 2020, arXiv:2012.03930.

[25] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, ‘‘Learning selfconsistency for deepfake detection,’’ 2020, arXiv:2012.09311.

[26] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, ‘‘On the detection of digital face manipulation,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, Jun. 2020, pp. 5780–5789, doi: 10.1109/CVPR42600.2020.00582.

[27] H. H. Nguyen, J. Yamagishi, and I. Echizen, ‘‘Capsule-forensics: Using capsule networks to detect forged images and videos,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Brighton, U.K., May 2019, pp. 2307–2311, doi: 10.1109/ICASSP.2019.8682602.

[28] H. H. Nguyen, J. Yamagishi, and I. Echizen, ‘‘Use of a capsule network to detect fake images and videos,’’ 2019, arXiv:1910.12467.

[29] B. Lindemann, T. Müller, H. Vietz, N. Jazdi and M. Weyrich, “A Survey on Long Short-Term Memory Networks for Time Series Prediction,” 2019, doi:10.13140/RG.2.2.36761.65129/1.

[30] S. Xie, R. Girshick, P. Dollar, Z. Tu, K. He, “Aggregated Residual Transformations for Deep Neural Networks,” 2017, doi: 5987-5995. 10.1109/CVPR.2017.634.

[31] Tan, Mingxing & Le, Quoc. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2019.

★★★