INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

IMPLEMENTATION OF DEEP LEARNING ALGORITHMS FOR CLASSIFICATION OF DEEPFAKE IMAGES

[1]Shiva Kumar Talathati, [2]Dr. Ramavtar

[1]Research Scholar, Dep of Computer Science and Engineering, Glocal University.

[2]Assistant Professor, Dep of Computer Science and Engineering, Glocal University.

[1]shiva0845@gmail.com

ABSTRACT

Facial expression recognition (FER) is a complex area of computer vision that involves detecting and evaluating how individuals express themselves through their faces. It has applications in human-computer interaction, video surveillance, and affective computing. Deep learning methods have significantly improved the field of picture facial expression recognition (FER), but the intricacy and evolution of facial expressions make it a challenging problem on video. The temporal context of video data provides potential additional clues for FER, but also presents challenges such as changes in illumination, location, and occlusion. Feature extraction is used to recognize faces and identify emotions, and the FER on video algorithm relies heavily on this stage. Convolutional Neural Networks (CNNs) are widely used in FER on video due to their ability to learn complex characteristics from data. CNNs can be used in various ways for feature extraction, including pre-trained CNNs, CNNs with two streams, CNNs in three dimensions, and CNN-Hybrids. However, due to technological constraints, such models cannot be implemented on drones for real-time facial recognition. Drones have limited storage capacity and processing speed, making large models for facial recognition unsuitable for on-board execution. Fully convolutional neural networks are suggested for use when computational resources are scarce, as they do not connect nodes in the network. Face detection, the initial step in facial recognition systems, aims to locate and extract facial portions from images or videos through face extraction. Deep learning strategies have shown high levels of success in facial recognition tests, with CNN models being the most promising approach.

ISSN: 2057-5688

KEYWORDS: CNN, Image Processing, Deep Learning Models and Feature Extraction.

I. MTCNN: Multi-task Cascaded Convolutional Networks

Face recognition is a common application for the MTCNN deep learning model, which is often employed for the purpose of identifying faces. It utilizes a cascaded design that is composed of three stages. According to Ekman, P. et al. (1971), P-Net is in charge of the formation of candidate face regions, R-Net is in charge of the refining and filtering of candidate face regions, and O-Net is in charge of the final classification of face region candidates. Even in the presence of occlusions, as well as fluctuations in position and size, the cascaded architecture makes it possible to recognize faces in a quick and accurate manner.

SSD: Single Shot MultiBox Detector

SSD is yet another well-known deep learning model that may be used for face recognition. SSD was first developed for the purpose of object identification; however, it may be extended for face detection by training on datasets that are particular to faces. In order to recognize faces at a variety of different dimensions and aspect ratios, SSD does a single feed-forward run through a deep neural network. It achieves real-time efficiency while maintaining a high level of accuracy by deriving class probabilities and bounding boxes directly from feature maps.

Face Verification using Deep Learning

The process of face verification entails evaluating whether or not two photographs of a person's face belong to the same individual. It has been shown that deep learning models may be effectively used to the job of face verification, which enables dependable and accurate matching of facial traits. Architectures such as Siamese networks and Triplet networks are often used in face verification systems.

Siamese Networks

In order to extract feature embeddings from a pair of input face photos, Siamese networks make use of a pair of CNNs that are the same and share their weights. After that, a similarity metric, is used to the embeddings in order to make the determination of whether or not the faces belong to the same individual (Zhang, K., et al. (2016)). In order to develop discriminative embeddings for face verification, Siamese networks are trained using

ISSN: 2057-5688

alternating pairs of face photos that match the same person and face images that do not match different people.

II. TRIPLET NETWORKS

The fundamental idea of Siamese networks is extended upon by triplet networks, which consist of an additional anchor photo, a positive image that portrays the same person as the anchor, and a negative image that depicts a different person from the anchor. During training, the network is taught to narrow the gap that now exists between the anchor and positive embeddings, while at the same time widening the gap that currently exists between the anchor and negative embeddings. The accuracy of face verification may be improved with triplet networks since these networks provide a larger degree of discrimination between faces that are similar to one another.

Face Identification using Deep Learning

The process of face identification includes identifying and categorizing a subject's face into recognized identities using a database of previously established people. The use of deep learning models, in particular CNNs, to the job of face detection has shown astounding levels of success. These models are taught to identify distinguishing characteristics within face photographs and to assign those characteristics to a set of predetermined identities.

CNN-based Classification Models

For the purpose of identifying people's faces, CNN-based models, such as the well-known VGGNet, ResNet, and Inception models, have seen widespread use. These models are trained on large-scale datasets including VGGFace, MS-Celeb-1M, and MegaFace. These datasets include a huge number of tagged face photos representing a variety of different people. The models are able to properly categorize fresh faces into recognized identities because they make use of the representations that they have learnt.

One-shot Learning and Few-shot Learning

Handling situations in which just a few or even a single picture of a person is available for training is one of the difficulties that might arise while attempting facial recognition. Techniques such as one-shot learning and few-shot learning have been included inside deep learning algorithms in order to circumvent this obstacle (Zhang, K., et al. (2016)). These strategies strive to generalize and identify new identities with minimal training samples by

ISSN: 2057-5688

utilizing the information learnt from a bigger collection of existing identities. This is accomplished by leveraging the knowledge gained from a larger number of known identities.

III. CHALLENGES IN DEEP LEARNING FACIAL RECOGNITION

Deep learning has made significant strides in face recognition, but there are still challenges to overcome. One of the main obstacles is data uncertainty, which is crucial for decision-making, machine learning, and statistics. Measurement uncertainty, sampling uncertainty, missing data, noise, and model uncertainty are all factors that need to be considered in academic research and analysis. Measurement uncertainty is a common issue in real-world data collection due to limitations in measurement instruments or human fallibility. It represents the potential range of values that a data point may take. Sampling uncertainty is an inherent level of uncertainty about the extent to which the sample accurately represents the larger population. Larger sample sizes often exhibit reduced levels of sampling uncertainty. Noise can arise from various sources, including limits in sensors, ambient influences, and human variability. Model uncertainty refers to the inherent uncertainty that arises when data is used to construct models, as it exists a level of uncertainty about the model's capacity to effectively represent the underlying connections present within the data. The influence on generalization is also significant. Overfitting occurs when a model with excessive noise in the data results in a common problem known as overfitting. Underfitting occurs when a model fails to capture intricate patterns in the data due to oversimplification. Bayesian Inference incorporates uncertainty into the modeling process, providing more cautious predictions and superior generalization capabilities in situations with scarcity of evidence. Confidence intervals measure the level of uncertainty around an estimated parameter, while bootstrap resampling assesses the level of uncertainty in a given dataset. Ensemble approaches, such as bagging and boosting, mitigate uncertainty by amalgamating predictions from numerous models, leading to enhanced generalization capabilities. Cross-validation methods, like k-fold cross-validation, evaluate the generalization performance of a model by dividing the data into numerous subsets. Heteroscedasticity, in regression analysis, refers to the presence of differing degrees of uncertainty in the residuals across distinct values of the independent variable, which is crucial for ensuring the appropriate generalization of a model. Privacy and ethical concerns have become increasingly prominent in the digital era due to technological advancements and the rise of data gathering and processing. The acquisition of personal data

ISSN: 2057-5688

by organizations, often without people's awareness or informed agreement, raises concerns about its use and safeguarding. Data breaches pose a significant privacy threat, as they expose sensitive personal information, financial data, and trade secrets. Mass surveillance, the systematic monitoring of people's online activity and physical movements by government and corporate entities, also poses a threat to personal privacy. The misuse of data is another concern, with entities exploiting personal data for financial gain or unethical objectives. The process of re-identification involves connecting supposedly anonymous data to particular persons even after anonymization has been completed. Algorithmic bias, the presence of prejudice in artificial intelligence (AI) systems, is a significant concern, particularly in decision-making procedures such as employment, lending, and law enforcement. AI-driven social prejudices can fortify disparities and generate discriminatory outcomes. Informed consent is crucial for the ethical development of AI and machine learning technologies, but transparency deficiencies and lack of transparency in privacy policies can hinder understanding of data use. Ethical standards are crucial for ensuring the conscientious development and implementation of AI and machine learning technologies, emphasizing principles such as justice, transparency, and the prevention of damage. The erosion of trust may occur due to recurring data breaches, unethical data practices, and biased decision-making by AI systems, leading to a loss of confidence in various entities. The rise in public knowledge about privacy and ethical concerns has resulted in a growing need for technology and data management practices that prioritize responsibility. Regulatory and legal frameworks within industries and organizations are essential for adherence to data protection laws and regulations. Agencies and government entities create these frameworks to protect individuals' rights to privacy and enforce strict limits on data handling. Governments and regulatory bodies are now considering legislation to address unethical practices in artificial intelligence R&D. The discussion around children's data consent is interesting and thought-provoking, given the risks associated with data collection and targeting, as well as children's limited understanding and capacity to offer informed consent. The effectiveness of anonymization techniques may be insufficient when confronted with sophisticated data re-identification tools, raising issues over the privacy of purportedly anonymous data.

IV. PROPOSED RESEARCH WORK

ISSN: 2057-5688

Proposed Algorithm

1. Put together a group of pictures with the same amount of male and female faces. For now, let's call this group of facts "D."
2. Use the information to make different sets for training and validating. Let's write the name of the training set as D_train and the name of the validation set as D_val. Use a number of either 80/20 or 70/30 for teaching and evaluation, depending on what works best.
3. As part of the preparation, change the size of the pictures so that they are 224 pixels on each side, which is the usual size for the ResNet-152 model. The preprocessed training set will be called X_train, and the preprocessed validation set will be called X_val. To standardize the pixel values, take the mean out of the total and divide the result by the standard deviation.
4. Load the ResNet-152 design from a model that has already been trained, like one that was trained using the ImageNet dataset. For now, let's call this model M. It has already been trained.
5. In the ResNet-152 model, change the last fully linked layer with a new layer that has two output neurons, one male and one female. Let's call this model "M," which stands for "modified."
6. Fix all of the ResNet-152 model's layers except the top layer. From now on, the sign will show what the model's values are.
7. Figure out how well the model works on the test set. To figure out if the confirmation is right, do the following:

accuracy = sum(i=1 to D_val) * 1/D_val. [y_val(i) == M'(x_val(i);)]

Where y_val(i) is the real gender label of the ith validation case and argmax(M'(x_val(i)); is the predicted gender label of the ith validation case, which is the output neuron with the highest predicted probability. The symbol stands for y_val(i).

8. Check how well the model works on data from a test set once it has hit a point of convergence on data from the training set. D_test will stand for the test set, while X_test will stand for the test set that has already been handled. Find out if the test is right by doing the same calculations as in step 7.

ISSN: 2057-5688

9. Use the model for tasks that involve gender classification, like figuring out the gender of people in a set of pictures or a continuous video stream. Using the new data and the factors you've already learned, draw conclusions.

Core algorithm:

1. **Input Image**: Start with an input image of a certain dimension, typically 224×224×3224×224×3 for RGB images.
2. **Initial Convolutional Layer**: Pass the image through a few initial convolutional and max-pooling layers to extract basic features.
3. **Residual Blocks**: The main component of ResNet is its residual blocks. For ResNet-152, there are several of these blocks, comprising 152 layers when summed up.

a. Each residual block contains convolutional layers followed by batch normalization and a ReLU activation.

b. The output of a residual block is added to its input (the skip connection), which helps in back-propagating gradients without degradation.

4. **Stacked Residual Blocks**: In ResNet-152, the residual blocks are stacked with increasing numbers of filters and occasionally decreasing spatial dimensions using stride-2 convolutions.
5. **Global Average Pooling**: After passing through all the residual blocks, a global average.
6. **Fully Connected Layer for Classification**: The output from the global average pooling is passed through a fully connected layer with a softmax activation. Since we're focusing on gender classification, this layer would have two nodes, corresponding to male and female. The softmax activation would give the probabilities of the input image being male or female.
7. **Training**:

a. Initialize the weights using techniques like He initialization, suitable for ReLU activations.

b. Use a training set of labeled male and female images.

c. Forward propagate an image through the network to get the prediction.

d. Compute the loss by comparing the prediction with the true label.

e. Backpropagate the error through the network using optimization algorithms like Adam or SGD to adjust the weights.

ISSN: 2057-5688

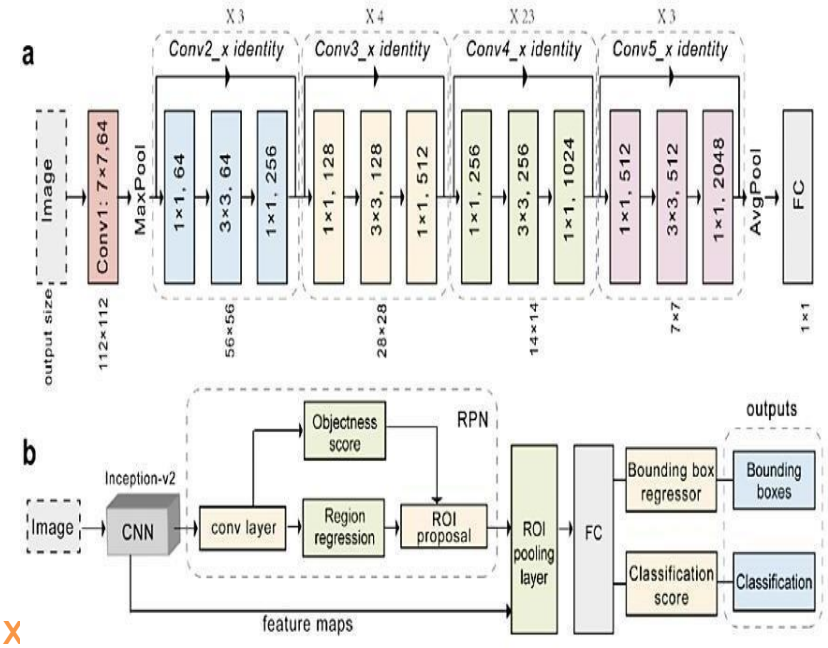
f. Repeat steps c-e for multiple epochs or until the model performance plateaus on a validation set.

8. **Evaluation**: After training, evaluate the model's performance on a test set. Check accuracy, precision, recall, and potentially other metrics to understand its classification capabilities.

Advantage of the proposed algorithm

- High accuracy: ResNet-152 is a deep convolutional neural network architecture shown to achieve high accuracy on image classification tasks. This makes it suitable for gender classification, where accuracy is critical.
- Pretrained model: The ResNet-152 architecture is a pretrained model trained on a large dataset (ImageNet) of millions of images. This allows us to leverage the pre-trained weights to improve the accuracy of our gender classification model.
- Transfer learning: By using transfer learning, This allows us to focus on training the final classification layer, which has significantly fewer parameters than the ResNet-152 model.
- Versatile: The proposed algorithm is versatile and can be applied to various gender classification tasks, including detecting gender in images, videos, or real-time streams.

Figure 1. Building a ResNet in an Organized.



ISSN: 2057-5688

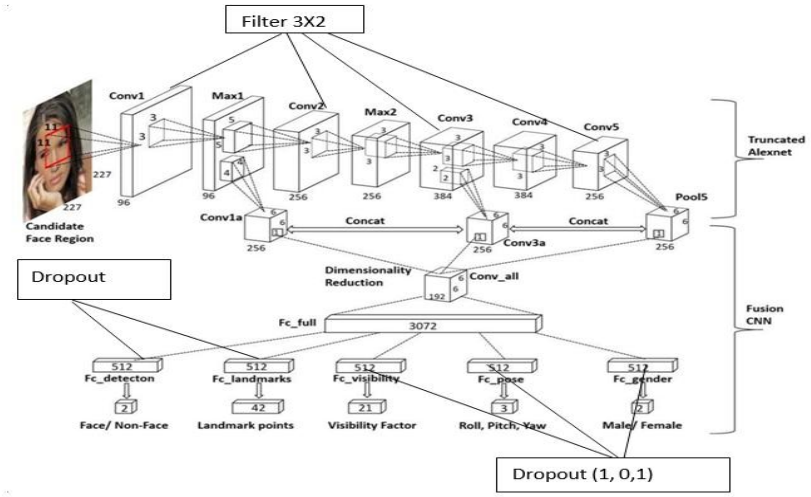


Figure 2. Proposed architecture.

The model's flow starts with finding the face region once captured. Then features are extracted from ResNet to make a feature vector for Xception model training, dataset Indian data contains noisy data. There was authentication information available for the data card that was made to train over globally available public datasets, as these were also used by the state-of-the-art methods in one or more situations.

Dataset

Dataset 1. The dataset contains:

https://datarepository.wolframcloud.com/resources/FER-2013

- The FER-2013 dataset has 35,887 photos, all of which are grayscale and 48x48 pixels in size.
- The dataset contains seven unique facial emotion classes: anger, contempt, fear, happiness, sorrow, surprise, and neutral. These emotions may be seen on a person's face.
- All of the photographs included in the collection were obtained via the use of the Google Photos search engine and were then labeled by real people.
- The training set has a total of 28,709 photographs, while the public test set contains 3,589 pictures, and the private test set also contains 3,589 pictures.
- The photos in the dataset were resized to a uniform 48x48 pixels and subjected to histogram equalization as part of the preparation stage.

ISSN: 2057-5688

- The FER-2013 dataset is unbalanced in terms of class frequency, with the neutral expression class comprising over half of all observations.
- Difficulties One of the primary difficulties in using the FER-2013 dataset is that it has a rather low resolution, making it difficult to discriminate between nuanced facial expressions.

Dataset 2. https://www.kaggle.com/datasets/jessicali9530/celeba-dataset

More than 200,000 images of celebrities are included in the CelebA (Celebrities Attributes) collection, a big face recognition database. A team of researchers from USC and the Chinese University of Hong Kong created the dataset in 2015. There are 10,177 photos of celebrities in the CelebA collection, and 40 attributes are annotated for each. For example, "Male" or "Female," "Smiling" or "Not Smiling," "Eyeglasses" or "No Eyeglasses," and so on are all examples of binary labels that make up the characteristics. The positions of five landmarks, including the eyes, nose, and mouth, are also annotated in the dataset. Images in CelebA vary greatly from one another in terms of position, facial expression, lighting, and setting. Images with substantial facial hair or partial occlusion are also included in the dataset's frontal and profile perspectives. High-resolution photos (178x218 pixels) are included in the CelebA dataset. The dataset contains a total of 162,770 photos: a training set, a validation set, and a test set, each of which has 19,867 photographs. Numerous computer vision applications, including studies of face recognition and attribute prediction, have made heavy use of the CelebA dataset. The dataset includes examples of face detection, face alignment, face recognition, and face attribute prediction. The CelebA dataset is freely available for research purposes and may be downloaded from the official website.

Dataset 3. https://www.kaggle.com/datasets/jangedoo/utkface-new

Over 20,000 human faces are shown in the UTKFace dataset, a large-scale face recognition dataset. 2017 was the year when researchers from the University of Texas at Arlington generated the dataset.The UTKFace dataset includes images of people of different ages,

ISSN: 2057-5688

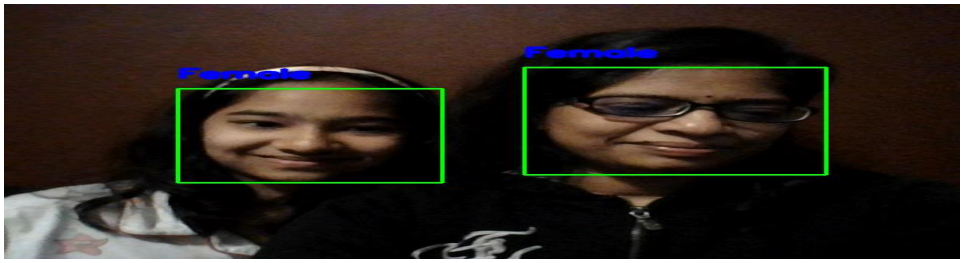
genders, and ethnicities. The images are categorised into 4 age groups: 0-17, 18-30, 31-45, and over 45. Each image also includes annotations for age, gender, and ethnicity.

The UTKFace dataset includes images with variations in pose, expression, and lighting and images with occlusions and background clutter. The images are grayscale, with a resolution of 200x200 pixels.

Dataset 4. https://www.kaggle.com/datasets/jessicali9530/lfw-dataset

LFW (Labeled Faces in the Wild) is a popular face recognition dataset including over 13,000 images of faces collected from the internet. The dataset was created in 2007 by researchers at the University of Massachusetts, Amherst. The LFW dataset has photographs of individuals of various ages, races, and nationalities. Images range in resolution and aspect ratio, and may be in either color or black & white. Images of individuals with their names annotated are also included in the collection. Training and evaluation data are separated in the LFW dataset. More than 4,000 photos are used in the training set, while over 5,000 are used in the test set. The test set was made to be tough on purpose, with several photographs including people in awkward stances or partially obscured by backgrounds. Researchers in the field of face recognition have made extensive use of the LFW dataset, primarily for the purposes of verification and identification. Various face-related tasks, including recognition, verification, and identification, have made use of the dataset. For those doing academic study, the LFW dataset is accessible for free download from the dataset's official website. In addition, benchmarks for multiple facial recognition algorithms on the dataset are available on the website.

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688



Figre 3: Illustrative example of gender classification

V. RESULTS AND DISCUSSION

In this part, we must explain the results obtained from the four datasets used to evaluate our suggested approach and the Loss function for the current step-by-step ResNet 152.

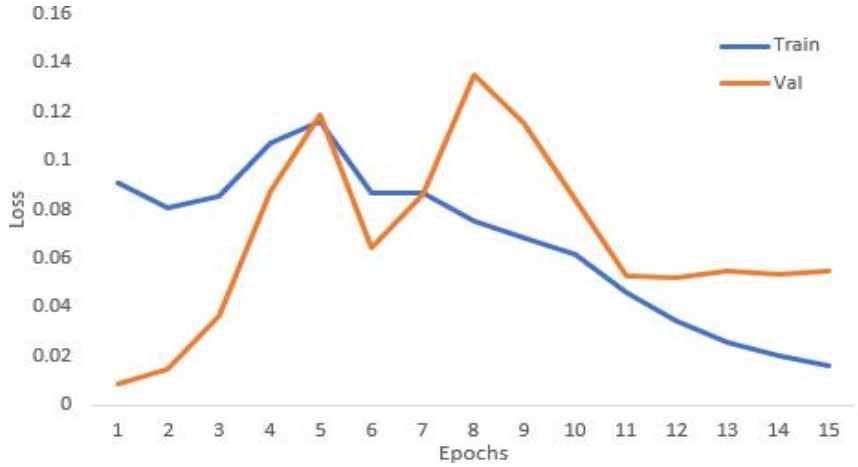


Figure 4 The loss function for the most recent step-by-step iteration of ResNet 152

Table 1: Gender classification on LFW dataset.

	LFW		
	Recall (FN)	**Precision (FP)**	**Accuracy**
Inception V3	78	78	81
VGG16	61	60	87
ResNet 50	65	60	86
EfficentNet	71	78	89
HyperFace	92	91	91
HF-ResNet	84	86	93
Proposed ResNet 152	97	98	99

ISSN: 2057-5688

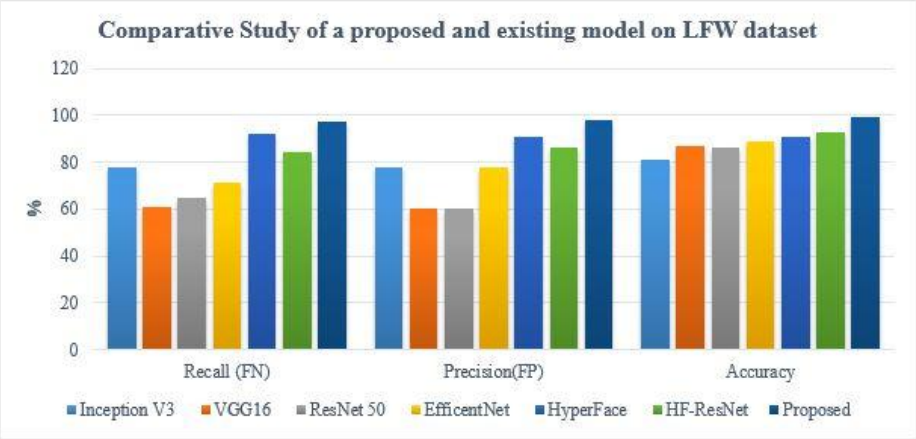
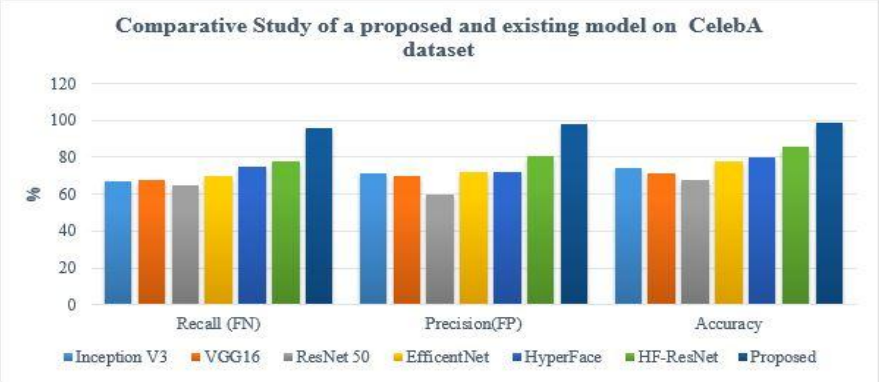


Figure 5: Gender classification on LFW dataset.

Figure 4 and Table 5 provide the LFW dataset for recall, precision, and accuracy. Methods like Inception v3, VGG16, ResNet50, Efficient Net, HyperFace, and HF are evaluated alongside the suggested technique. -ResNet.

Table 2: Gender classification on Celeba dataset.

	CELEBA		
	Recall (FN)	**Precision(FP)**	**Accuracy**
Inception V3	67	71	74
VGG16	68	70	71
ResNet 50	65	60	68
EfficentNet	70	72	78
HyperFace	75	72	80
HF-ResNet	78	81	86
Proposed ResNet 152	96	98	99



ISSN: 2057-5688

Figure 6: Gender classification on Celeba dataset.

Figure 6 and Table 3 display the and accuracy. Methods like Inception v3, VGG16, ResNet50, Efficient Net, HyperFace, and HF are evaluated alongside the suggested technique.-ResNet.

Table 3: Gender classification on UTKF dataset.

	UTK Face		
	Recall (FN)	**Precision(FP)**	**Accuracy**
Inception V3	64	70	61
VGG16	49	60	58
ResNet 50	40	60	68
EfficentNet	80	79	82
HyperFace	84	86	93
HF-ResNet	88	87	90
Proposed ResNet 152	96	97	98

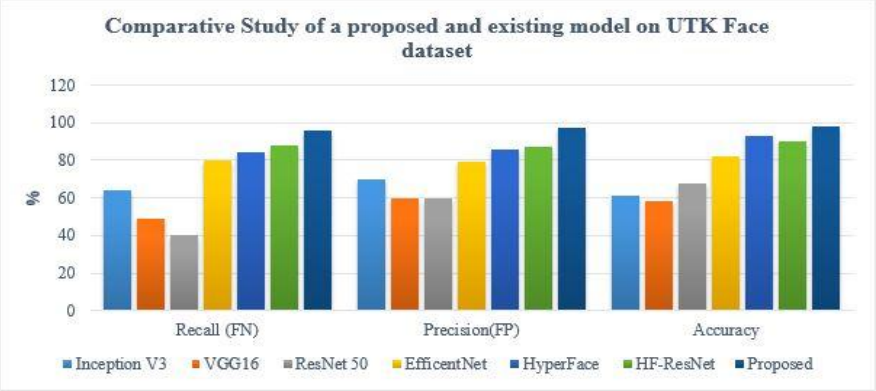


Figure 7: Gender classification on UTKF dataset.

The LFW recall, precision, and accuracy dataset is shown in Table 4 and Figure 7 Methods like Inception v3, VGG16, ResNet50, Efficient Net, Hyperface, and HF are evaluated alongside the suggested technique. -ResNet.

Table 4: Gender classification on FERET dataset.

	FER-2013		
	Recall (FN)	**Precision (FP)**	**Accuracy**
Inception V3	62	69	71
VGG16	52	59	61

ISSN: 2057-5688

ResNet 50	50	52	56
EfficentNet	68	62	76
HyperFace	80	78	84
HF-ResNet	81	84	91
Proposed ResNet 152	97	95	98

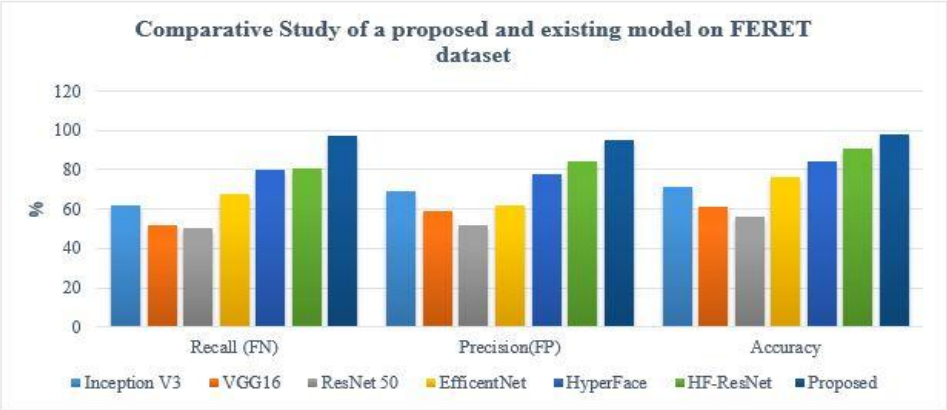


Figure 8: Gender classification on FERET dataset.

The LFW recall, precision, and accuracy dataset is shown in Table 4 and Figure 8. Methods like Inception v3, VGG16, ResNet50, Efficient Net, Hyper Face, and HF are evaluated alongside the suggested technique. -ResNet.

CONCLUSION

Using ResNet-152 for gender classification is an interesting method that has shown very accurate results when it comes to gender detection from photos. The model uses sophisticated deep learning methods to extract the relevant information, allowing it to correctly identify the input photos' gender. It may be beneficial to do further research in this area. Maybe in the future researchers may look at methods to improve the model's accuracy by using pictures shot from different perspectives or in varied lighting. This could be achieved by introducing data augmentation techniques or fine-tuning the model on a larger, more diverse dataset. Another potential direction for future research could be exploring alternative deep-learning architectures or pre-trained models. While ResNet-152 is effective, there may be other models that could achieve even better performance. A promising field of study is gender categorization via the use of ResNet-152, and more work in this area can potentially increase the accuracy and resilience of gender classification models.

ISSN: 2057-5688

REFERENCES

1. *Agarwal S, Farid H, El-Gaaly T, Lim SN. Detecting deep-fake videos from appearance and behavior. 2023 IEEE International Workshop on Information Forensics and Security, WIFS 2020. 2020. https://arxiv.org/abs/2004.14491v1. Accessed 31 Aug 2023*
2. *Bakir, V., & McStay, A. (2018). Fake News and The Economy of Emotions: Problems, causes, solutions. Digital Journalism, 6(2), 154–175. https://doi.org/10.1080/21670811.2017.1345645*
3. *N. Jain and P. Peddi, "Gender Classification Model based on the Resnet 152 Architecture," 2023 IEEE International Carnahan Conference on Security Technology (ICCST), Pune, India, 2023, pp. 1-7, doi: 10.1109/ICCST59048.2023.10474266.*
4. *Othmani, A., Taleb, A. R., Abdelkawy, H., & Hadid, A. (2020). Age estimation from faces using deep learning: A comparative analysis. Computer Vision and Image Understanding, 196, 102961.*
5. *Prasadu Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labled Gradient Boosting System for Student Activity Analysis", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73.*
6. *Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. Perception, 31(9), 1109-1121.*
7. *Remya Revi K.1 , Vidya K. R.1 and M. Wilscy," Detection of Deepfake Images Created Using Generative Adversarial Networks – A Review ", February 2021*
8. *Sun, X., & Lv, M. (2019). Facial expression recognition based on a hybrid model combining deep and shallow features. Cognitive Computation, 11(4), 587-597.*
9. *Valstar, M., Martinez, B., Binefa, X., & Pantic, M. (2010, June). Facial point detection using boosted regression and graph models. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 2729-2736). IEEE..*
10. *Wu, X., He, R., Sun, Z., & Tan, T. (2018). A light CNN for deep face representation with noisy labels. IEEE Transactions on Information Forensics and Security, 13(11), 2884-2896.*
11. *Xia, Y., Yu, H., Wang, X., Jian, M., & Wang, F. Y. (2021). Relation-aware facial expression recognition. IEEE Transactions on Cognitive and Developmental Systems, 14(3), 1143-1154.*
12. *Yan, Y., Huang, Y., Chen, S., Shen, C., & Wang, H. (2019). Joint deep learning of facial expression synthesis and recognition. IEEE Transactions on Multimedia, 22(11), 2792-2807.*
13. *Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2015). Learning social relation traits from face images. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3631-3639).*