

Real-time eyeglass detection using transfer learning for non-standard facial data

Ritik Jain¹, Aashi Goyal¹, Kalaichelvi Venkatesan²

¹Department of Computer Science, Birla Institute of Technology and Science Pilani, Dubai Campus, Dubai, United Arab Emirates

²Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science Pilani, Dubai Campus, Dubai, United Arab Emirates

Article Info

Article history:

Received Jul 17, 2021

Revised Mar 19, 2022

Accepted Mar 30, 2022

Keywords:

Computer vision

Convolutional neural network

Deep learning

Eyeglass detection

Ocular images

Performance analysis

Transfer learning

ABSTRACT

The aim of this paper is to build a real-time eyeglass detection framework based on deep features present in facial or ocular images, which serve as a prime factor in forensics analysis, authentication systems and many more. Generally, eyeglass detection methods were executed using cleaned and fine-tuned facial datasets; it resulted in a well-developed model, but the slightest deviation could affect the performance of the model giving poor results on real-time non-standard facial images. Therefore, a robust model is introduced which is trained on custom non-standard facial data. An Inception V3 architecture based pre-trained convolutional neural network (CNN) is used and fine-tuned using model hyper-parameters to achieve a high accuracy and good precision on non-standard facial images in real-time. This resulted in an accuracy score of about 99.2% and 99.9% for training and testing datasets respectively in less amount of time thereby showing the robustness of the model in all conditions.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ritik Jain

Department of Computer Science, Birla Institute of Technology and Science Pilani, Dubai Campus

P.O. Box No. 345055, Dubai, United Arab Emirates

Email: ritikjain@gmail.com; f20180231@dubai.bits-pilani.ac.in

1. INTRODUCTION

Human soft biometrics identification is an important and hot computer vision (CV) research field. Surveillance systems provide vast data to be visualized [1]. Facial analysis serves as backbone for most of these identification tasks [2]. The human face has recognizable demographic with lots of attributes to compare with. Although these factors like nose bridge level, forehead facial expression and all other factors seem easy for identification of faces in an image, the process is not elementary as certain technological factors like resolution and quality of the image, background and shadows, and presence of eyeglasses could affect the performance.

Human facial analysis can give inaccurate results because of certain obstruction near ocular region offered by eyeglasses in terms of reflection from lenses and frame occlusion, for example in facial analysis models using convolutional neural networks. The situation is even worse when images have sunglasses, thereby covering the entire eye area causing failure in eye detection in system. To increase the precision level for human faces by eliminating the above factors, many systems have included an eyeglasses or non-eyeglasses classification phase in their models followed by removal of eyeglasses by regenerating facial images using generative adversarial or similar network models. Therefore, a robust and accurate eyeglasses classifier is needed to improve and strengthen systems to be accurate in real world non-standard images as well.

Glasses detection is not a new problem, but very limited amount of work is dedicated for a clear detection of glasses in facial images. Various approaches were adopted by researchers involving handcrafted

feature extraction methods, edge detection in the Frame Bridge and ocular region and using shallow features to increase the prediction time of the model. The models used for this classification task can be roughly grouped into three categories: Linear classifier, neural networks, and Bayesian networks. Using a linear classifier can be pretty simple as it does not involve a lot of training data given the simplicity of the model, but the challenge arises in the feature extraction task which can be tricky as the images data may have a variety of features depending upon the background, and lighting conditions. Hidden Markov models (Bayesian network model) are quite efficient and give high accuracies. The advantage of using this type of model is that they are good at capturing certain temporal features, but prior to learning they require a clear definition of model which is difficult given the diversity of our data [3].

All the existing approaches used pre-existing standard datasets with handcrafted feature extraction methods for training which does not represent real-life scenarios. This leads to poor real-time performance. Looking at the success of deep learning in various computer vision projects, a convolutional neural network (CNN) based transfer learning model is capable to extract deep features which will have inputs for a same person's image with and without the glasses to make the comparison and easy detection of eyeglasses with respect to other facial recognition factors. CNN acts as a regularized multilayer perceptron network by convolving input data with learned features using 2D convolutional layers. As a result, CNNs can learn filters or patterns that must be explicitly defined in other classification algorithms. Similarly, in our case there is no need to explicitly define the filter or pattern for eyeglasses, rather the network automatically learns the features and creates a gradient filter to focus on specific subarea of the input image. But for successful filtering and pattern recognition using CNN, a varied dataset representing all possible real-life scenarios is required. Therefore, an attempt to generate such a dataset and fine-tuning the trained model on this dataset is made to increase the classification capability in real-time while being time efficient. The glasses would be covering the ocular region which will be the key factor for its detection. The performance of the model is tested in real-time using a custom-built framework using OpenCV.

The rest of the paper is organized as follows. Some related literature surveys will be given in section 2 describing 14 different past researchers works in the similar fields. Dataset collection and preparation is defined in section 3. Thereafter, section 4 talks about the detailed information of the implementation details and optimization of our model. Finally, the paper presents the results and conclusion in sections 5 and 6, respectively.

2. LITERATURE REVIEW

There has been a good amount of work done in the field of eyeglass detection by various researchers as the detection of glasses plays a key role in the removal of glasses from facial images which tends to cause a lot of problems for ocular and facial recognition tasks because of the presence of reflection and shadow offered by the eyeglasses. Mohammad *et al.* [4] evaluated and compared the use of various CNN structures being applied specifically on the ocular region and the frame bridge, serving as the regions of interest for feature extraction and training/learning of the model. Their CNN model was squeezed to be successfully run by mobile devices while offering parallel accuracy.

Basbrain *et al.* [5] emphasized on designing a robust shallow CNN system for detection of eyeglasses for much more faster results and high accuracy. They presented and addressed two major concepts of CNN namely, the adequate dataset length required for training and the depth of the architecture for the proposed network. After learning the parameters from a deep neural network and fine tuning it on a small dataset, some of the convolutional layers are removed to reduce the depth of the network after testing it on the validation data.

While targeting the region in CNN, some unrelated regions during learning are commonly used which leads to neglecting the effective features. Hao *et al.* [6] used an optimized CNN method instead of the traditional CNN since it does not consider weights of learning instances. They combine image recognition with bottom up segmentation considering the target selection and weight deliberation by reinforcement learning, this improved and by selecting huge amount of images robustness of the paper was shown and concluded that this model need further improvements. Another implementation by Chauhan *et al.* [7], using CNN on Modified National Institute of Standards and Technology (MNIST) and CIFAR-10 datasets for image recognition, the accuracy of model on MNIST came out to be 99.6% by using root mean square prop (RMS prop) as an optimizer, the paper concluded by saying that improvement was needed on the CIFAR-10 dataset which could be improved with larger epochs and taking more hidden layers for training the dataset.

Mo *et al.* [8] suggested an ensemble learning based image recognition algorithm which implements error level analysis (ELE-CNN), this method was used to solve those problems which single model cannot predict correctly. CIFAR-10 dataset was used as image dataset. Bagging ensemble was used to train the

models and a network structure was used which was a combination of DenseNet, ResNet, Inception-ResNet-V2, and DenseNet-BC architectures. The final result was the mean probability of the prediction vector.

Fernandez *et al.* [9] proposed an algorithm for automated recognition of glasses on face images derived by robust alignment along with robust local binary pattern (RLBP). Normalized and pre-processed image data is used to generate a RLBP pattern, based on the observation that the nose pin of glasses is often located at a level same as that of the center of the ocular region or eyes and finally, these features are used to perform classification task using a support vector machine. Another approach involving a unique set of Harr-like features on in-plane rotated facial images was implemented by Du *et al.* [10].

Bekhet and Alahmer [11] developed a deep learning-based recognition model for detection of glasses on a realistic selfie dataset which includes challenging and non-synthesized images of full or partial faces. After an extensive training of almost two weeks on a CNN based transfer learning model, they were able to such a model with an accuracy score of 96% on such a varied dataset.

Many research works have been done on eyeglasses recognition, removal, and localization [12], [13]. Jiang *et al.* [14] have used an eyeglasses classifier for its detection on facial images, for determining glasses on facial images only six measures were used. This method required that the position of eyes on the face in the images should be known since it will detect the presence of the glasses' bridge present in the ocular region between the eyes. Mohammad *et al.* [15] proposed two schemes learning and non-learning for eyeglasses detection. A non-learning scheme was used by the authors in contrast to the limited adaptive histogram equalization (CLAHE) filters to obtain edge information to get a hint of eyeglass detection. Learning scheme used multi-layer preceptor (MLP), support vector machine (SVM) and linear discriminant analysis (LDA) which were applied on two databases and results vary between 97.9% to 99.3% detection accuracy.

Eyeglasses can change the aspect and insight of facial images, if observing under infra-red it deteriorates the image quality, thereby making detection of glasses a prerequisite for the stratified quality. Drozdowski *et al.* [16] used deep neural networks based Caffe framework method for its detection which resulted the neural network classified images without glasses more accurately than the statistical approach and would work best on ocular images alone than the whole infrared face image. Jing and Mariani [17] used a deformable contour way for detecting the presence of eyeglasses using Bayesian network framework, they used 50 key points to mark the position and structure of glasses which was found by maximizing the posteriori and finally combining the edge and geometrical features. There were some ambiguities as grey scale of face skin was nearly identical to that of glasses which gave some wrong results. Based on the above survey, an attempt to build a model for detection of eyeglasses for non-standard facial images has been proposed. Also, an in-depth analysis has been carried out to optimize the model hyperparameters and real-time testing was performed using a custom-built framework.

3. DATASET PREPARATION

A pre-trained transfer learning model from Keras trained on the ImageNet ILSVRC dataset was used by freezing the weights of certain layers and redefining the top layers. A custom dataset of non-standard facial images was used to refactor the pre-trained model and update the weights on top layers to make the model robust in making prediction for real-life data. The dataset used for modelling contains a total of 46,372 images belonging to two classes namely eyeglasses present (label-0) and eyeglasses absent (label-1). Three sources of data from which the data was collected are mentioned below. The first part of dataset is custom built wherein, facial image data was collected from few families and friends by requesting them to click and send their pictures with and without eyeglasses in different illuminations and orientations. This part of the dataset involves a total of 4062 images from 20 subjects.

The database released by the authors of "An Indian facial database highlighting the spectacles problem contains 123,213 images of 58 subjects with and without eyeglasses in multiple sessions. These images were captured in real-time with different physical conditions like illumination, posture and orientation, alertness level, yawning and different frame types [18]. But due to the huge size of the data from just 58 subjects, using this entire dataset can cause over to overfit on this data. Hence, a random sub-sample of about 30% was selected in equal proportions from each subject to give a total of 37,388 images belonging to both output classes in a balanced ratio. The Kaggle dataset of "Glasses or No Glasses" containing 4,922 images were used in the third part of our dataset. This dataset contains high-quality pictures and thus bringing diversity to our dataset [19].

Figure 1 shows a single batch of 16 images from our labelled dataset collected from the three sources as mentioned previously. Apart from this a small dataset of about 400 images was collected from subjects whose images were not used for training for the purpose of testing the real-time performance of our model on images of subjects the model has never seen before. This dataset will be termed as real test data in rest of the paper. Additionally, two independent published datasets (Olivetti research laboratory (ORL) and

Sunglasses) were used to compare the generalization capabilities of trained model. ORL dataset includes 400 frontal facial images (in grayscale) of 40 different subjects both wearing and not wearing eyeglasses. These images were taken in various lighting and facial expressions. The second independent dataset was collected from Kaggle which includes about 3400 facial images of people with and without sunglasses [20], [21].

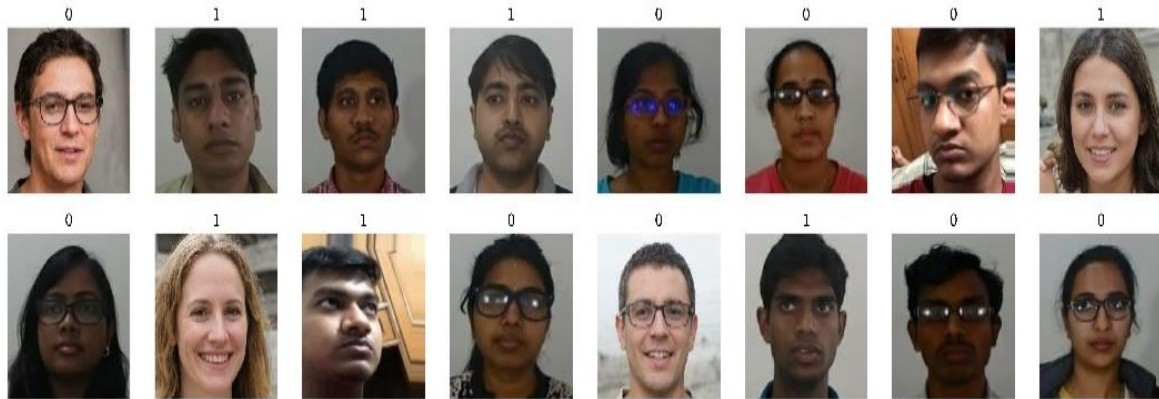


Figure 1. Single batch of 16 images from our labelled dataset

4. RESEARCH METHOD

This section is discussing about the pipeline proposed for prediction of presence of glasses in facial images in real-time along with optimizing and fine-tuning the model by varying different hyperparameters. The different is the types of optimizers, learning rates and activation function for the connected layers in the neural network model. Figure 2 represents the proposed pipeline.

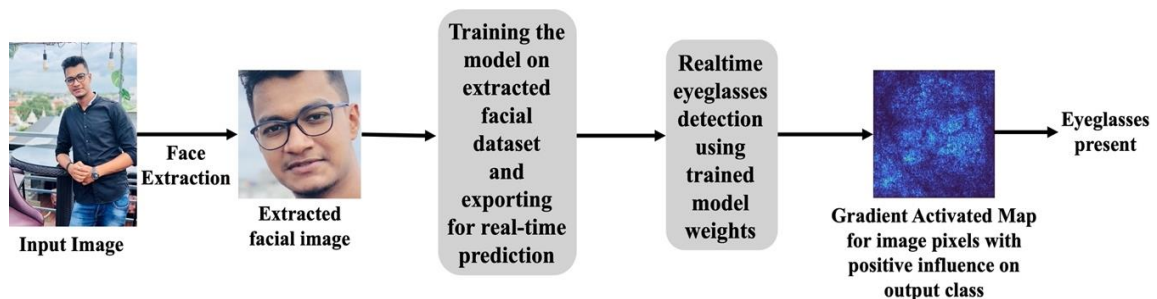


Figure 2. Pipeline of proposed work

4.1. Face detection and pre-processing of facial images

The input images need to be pre-processed before being used for training. The first step as shown in pipeline in Figure 2 is detecting and extracting the face in each input image, which is done by using a technique based on the renowned voila-jones face detection algorithm called the Haar-cascades. OpenCV supports object recognition using Haar features which can recognize faces in an image and return a rectangular region of interest (ROI) which can be cropped and saved as a new image. Using this approach, a new dataset is thus generated consisting of close shots of faces in the base dataset.

While using a pre-trained image classification model, the selection of such a model must be done carefully. There are a variety of pre-trained models available online, trained on varied datasets and using different network architectures. For this research, a pre-trained model from Keras was chosen which is based on the Inception-V3 architecture and trained on ImageNet dataset. Our pre-processing techniques thus includes resizing the images to (299,299) so as to match the input vector size of the pre-trained Inception-V3 model. Next step is to normalize each pixel value by dividing it by 255 so as to get the red, green, blue (RGB) values of each pixel in range 0 to 1 for the ease of calculations while training.

4.2. Model selection and training

Neural networks have proven to be best for solving classification problems as a result of their property to learn the most significant classification features by the virtue of neuron's collective behavior and ability of each neuron to do a fixed amount of parallel computation. Using a convolutional neural network can further boost the performance of the model as it has the advantage of having multiple layers over a simple artificial neural network. Using multiple layers helps us to automatically detect the most significant features for training without any human supervision.

Transfer learning is nowadays the most popular machine learning technique used for training classification models wherein a pre-trained model is used which is trained on significantly large datasets with high accuracy and then refactored on our concerned dataset to fit more. A typical transfer learning algorithm involves using a pre-trained model and recycling a portion of its weights while reinitializing and updating the weights of shallower layers. The elemental instance of this approach requires modifying or updating the weights on the final classification dense layer of a fully trained neural network and using it to classify a different dataset [22], [23]. Figure 3 visualizes the proposed convolutional neural network architecture used by the transfer learning model and the filters from its first and last convolutional layers.

Inception-V3 also known as the 3rd version of Google's convolutional neural network from inception family, is an architecture as shown in Figure 3(a) with several advancements offered by factorized 7×7 convolutions, label smoothing and propagating to lower down the network using an auxiliary and batch normalization layer [24]. Using a pre-trained model instance from keras based of the Inception-V3 architecture and trained on ImageNet ILSVRC dataset, a new model was defined by taking all the 311 layers except the top fully connected layers (which is used for classification) and explicitly freezing these layers' weights to prevent them from being changed or updated while training. To this model a few extra layers were added including, a flatten layer to reshape the tensor, a dropout layer with frequency of 0.5 to prevent overfitting and finally a dense layer with 2 output neurons representing out 2 output class labels namely eyeglasses present and eyeglasses absent. Our model now includes a total 314 layers with 94 convolutional filter layers with first convolutional layer having 32 filter and the last convolutional layer having 192 filters. The Figures 3(b) and 3(c) show how a typical convolutional layer filters features in an input image.

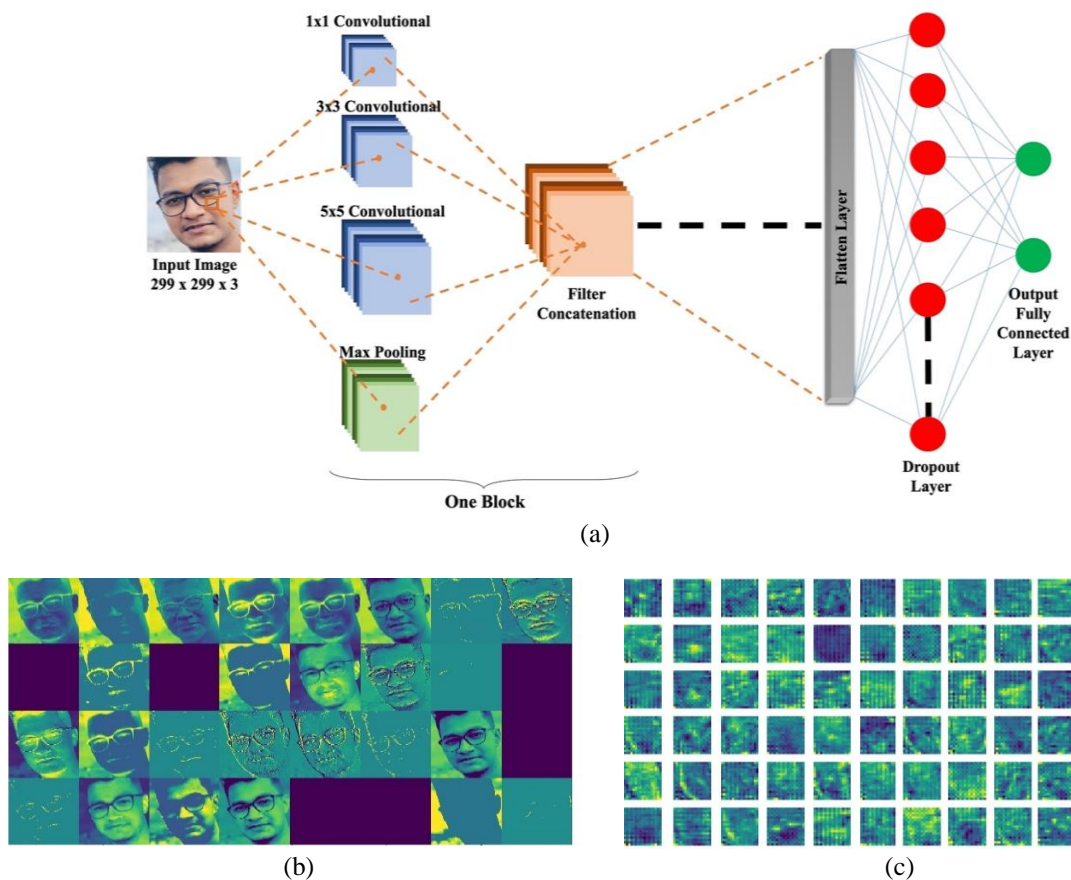


Figure 3. Proposed inception V3 based architecture and its convolutional layer filters

From Figure 3(b) it can be deduced that for an input image passing through the first convolutional layer filters, almost all the features in the image are retained although some filters are still not active at this layer which are marked by solid-colored boxes. Comparing this with the filters of our last convolutional layer in Figure 3(c), it is inferred that the activations become abstract being less interpretable visually. At this layer certain higher-level concepts are encoded like specific border, corners, and angles. Next step involves training of the model consisting of 262,146 trainable parameters over a total of 22,064,930 parameters present in the proposed model. Multiple models were trained on the batch image dataset while varying the values of different hyper parameters to find the optimized version of proposed model and hence fine tuning the model. After training all these models, the optimized model was exported and saved as “.h5” files which will be used for loading the model to make predictions in real-time using a custom framework.

4.3. Fine tuning hyper parameters

4.3.1. Training-testing splitting ratios

To study the effect of using different training-testing split ratios, three different ratios were used and trained the model on each of them to compare the performance. Table 1 shows that all the three models perform equally good, but the model with 80-20 split performing slightly better with the highest accuracy and least loss factor than the other two models. Also, batch size of 16 was found to be the best selection out of 8, 16 and 32 with significantly higher accuracy of about 15% more than the other two.

Table 1. Analysis of model for proposed different training-testing ratios

Train Ratio	Test Ratio	Train Accuracy	Test Accuracy	Train Loss
60%	40%	98.57	99.82	0.24
70%	30%	98.78	99.76	0.19
80%	20%	98.91	99.88	0.17

4.3.2. Optimizer

Optimizers are of great importance in the performance of neural networks. These are the algorithms that governs the change or updates in weights and learning rate during the training process. They also help in reducing and, hence, minimizing the losses. Hence selection of optimizes becomes crucial for the performance of our model. As a classification task is being performed in this research hence the loss function of interest is categorical cross entropy loss function which is a standard loss function for all multiclass classification tasks. To minimize the loss and maximize the performance of our model a detailed analysis is performed based on the performance of five distinct optimization algorithms namely stochastic gradient descent, RMSprop, Adam, Adagrad and Adadelta. To ensure the uniformity of our analysis a seed value of 100 was used while splitting our data into test (20%) and train (80%) datasets. The loss and accuracy plots in Figures 4 and 5, respectively represent the significantly better performance of Adagrad algorithm over other optimization algorithms as it converges to minima the fastest and has least overall loss value along with highest accuracy as mentioned in Table 2.

4.3.3. Learning rate

As Adagrad or adaptive gradient turned out to be the best performing optimization algorithm because of the highest accuracy and minimum loss offered by this algorithm in comparison to other algorithms mentioned in Table 2, Adagrad will be used for the optimization function and further fine-tuning will be performed on the learning rate of this optimization algorithm which governs the rate of change in weights while learning. Adagrad is an improvement over stochastic gradient descent with an adaptive learning rate which performs larger updates on the weights of features that occur less frequently while performing smaller updates for the features that occur more frequently. Adagrad instead of recording the sum of gradient like in momentum takes into account the sum of gradients squared, which is used to update the gradient in different directions as shown in the formula (1). The gradient history is stored in G_t as it is a diagonal matrix where each element i,i represents the sum of gradients squared for every $w_{t,i}$. ϵ is introduced to avoid division by zero. This diagonal matrix which records the history of squared sum of gradients is used to update the learning rate (α) [25].

$$w_{t+1,i} = w_{t,i} - \frac{\alpha}{\sqrt{G_{t,ii} + \epsilon}} \times g_{t,i} \quad (1)$$

Although Adagrad eliminates the process of manually tuning the learning rate but specifying the initial learning rate can change the performance of the model in terms of how fast our model converges to its

minima. Hence, comparison of model performance for five exponentially decreasing value of initial learning, ranging from 1.0 to 0.0001 while using Adagrad optimizer is represented visually in Figure 6.

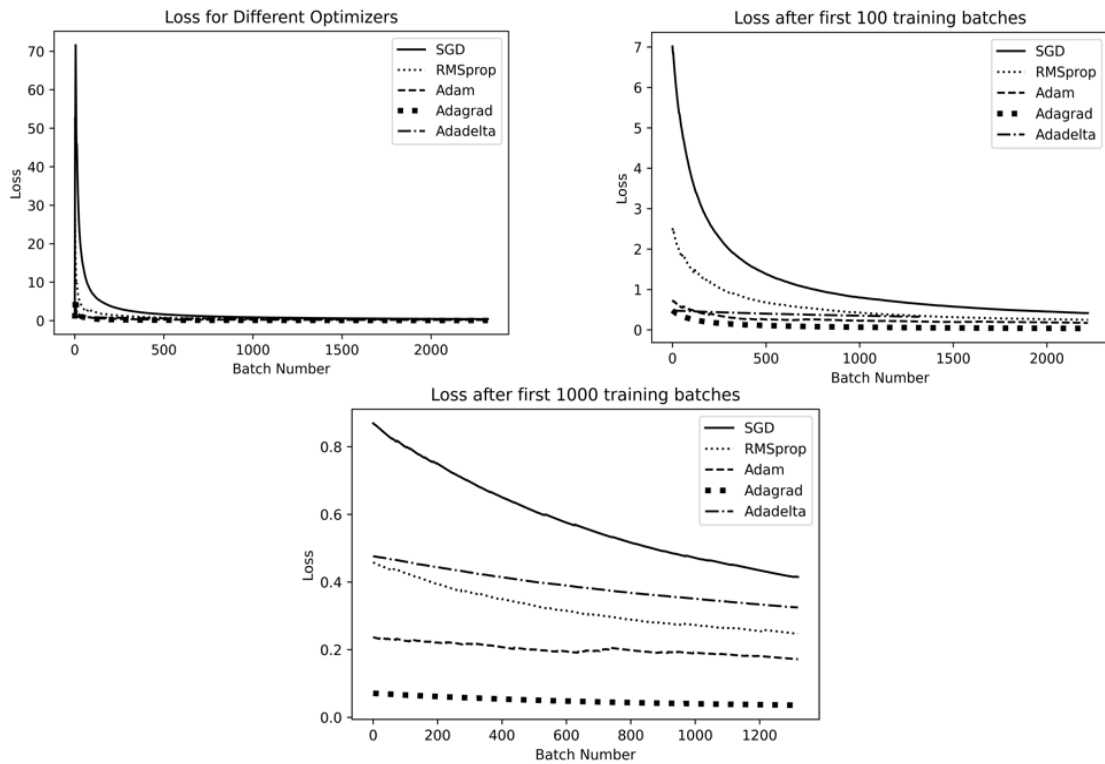


Figure 4. Training loss for different optimizer functions

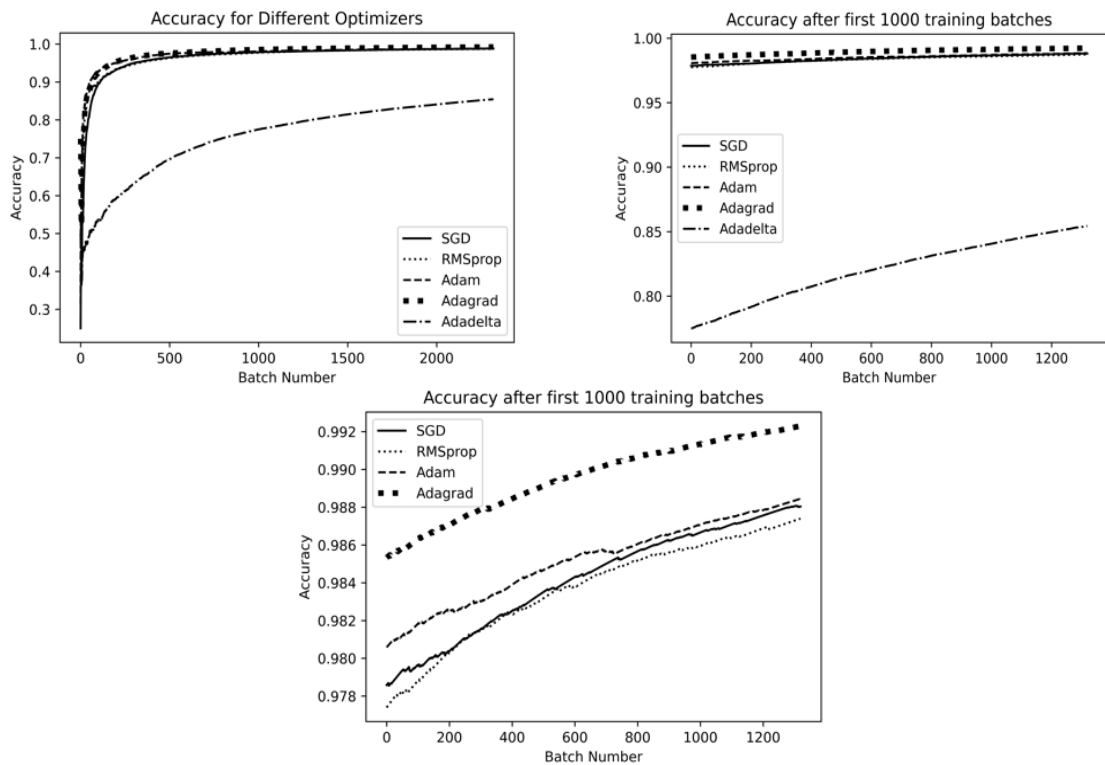


Figure 5. Training accuracy for different optimizer functions

Table 2. Analysis of model for proposed different optimizer functions

Optimizer Function	Train Accuracy	Train Loss	Test Accuracy	Test Loss
SGD	98.80	41.48	99.60	3.87
RMSprop	98.74	24.68	99.94	0.91
Adam	98.84	17.18	99.86	4.00
AdaGrad	99.23	3.56	99.95	0.25
AdaDelta	85.43	32.48	97.14	10.16

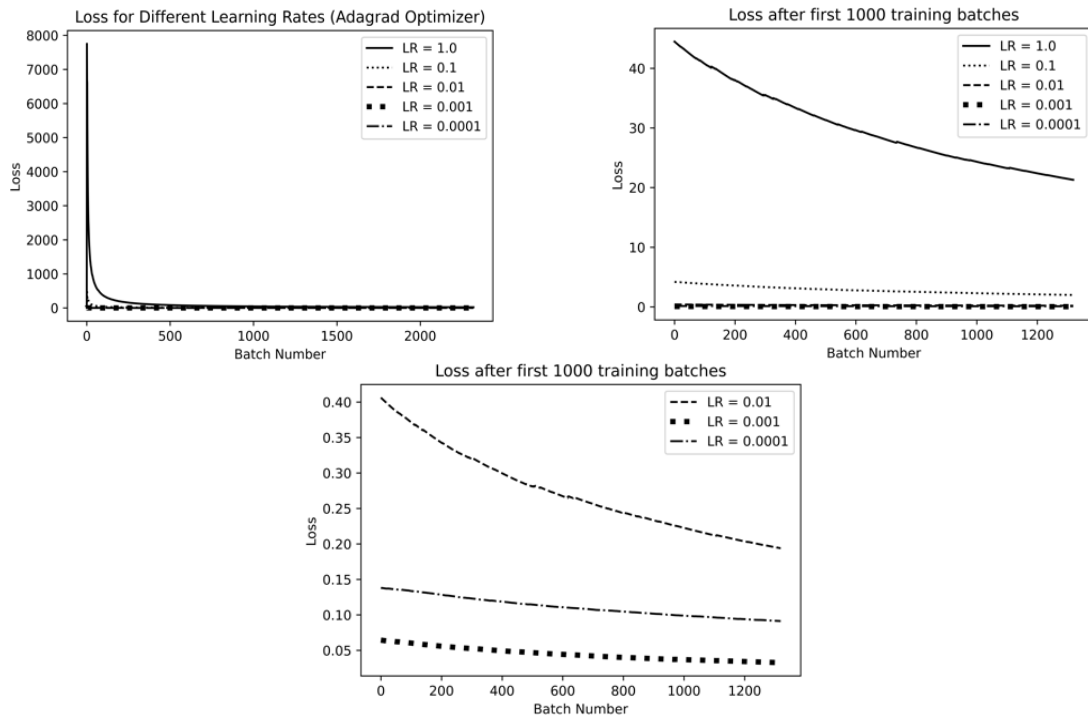


Figure 6. Training loss for different values of initial learning rate using Adagrad optimizer

In the plots in Figure 6, a pattern of loss decreasing with decreasing value of learning rate only till the 0.001 is observed. On decreasing (exponentially) the learning rate value further below 0.001, an increase in the overall loss value was observed for our trained model. Also, from the Table 3 it can be inferred accuracy not to be the suitable factor for comparing the performance based on initial learning rate because of marginal difference in accuracy on changing the learning rate. But looking at the loss values in Table 3 for models trained using different learning rates, the learning rate value of 0.001 is the best performing value as a result of its lowest loss value for both training and testing datasets.

Table 3. Analysis of model for proposed different initial learning rate values

Learning Rate	Train Accuracy	Train Loss	Test Accuracy	Test Loss
1.0	99.15	21.299	99.90	1.018
0.1	99.13	1.991	99.90	0.128
0.01	99.21	0.194	99.70	0.022
0.001	99.17	0.032	99.97	0.002
0.0001	96.81	0.091	99.47	0.024

4.3.4. Activation function (for fully connected layer)

At last, different activation functions were analyzed for the top fully connected dense layer consisting of 2 output neurons which gives a probabilistic output for an input image belonging each of the two output classes namely eyeglasses present and eyeglasses absent. Rectified linear unit (ReLU) activation function is often used with hidden layers because of its non-linear property and its ability of having negligible backpropagation errors as seen in sigmoid activation function. Therefore, the analysis of performance of Sigmoid and SoftMax activation functions in performed as shown in Figure 7.

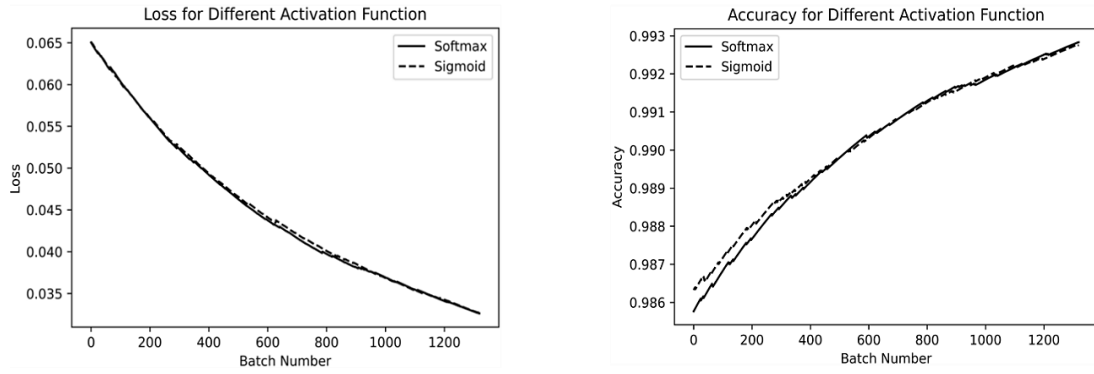


Figure 7. Training accuracy and loss for different activation function of fully connected layer

Sigmoid and SoftMax activation function are very similar to each other with only difference being Sigmoid is generally used for binary classification tasks while SoftMax is generally used for multi-class classification tasks. As a result of their similar nature, a very similar trend is seen in Figure 7 and the results of both these activation functions are found to be nearly identical. Hence the choice of proposed activation functions does not change or improve the performance of model. SoftMax activation function was used for the final optimized model.

4.4. Simulation of real-time eyeglass detection framework

For real time detection framework, the model trained on the fine-tuned hyper parameters is used and loaded through saved model in .h5 file. This loaded model is used to make predictions on frame captured in real-time by OpenCV [26]. Initially, the entire frame is captured and then using the Viola-Jones based Haar Cascading method, the faces present in the frame are extracted as mentioned in section 4.1. Before making any predictions for the extracted face image, it needs to be pre-processed using various techniques as mentioned in section 4.1. By extracting the individual frames from the live camera feed, the frame is resized according to the model input vector size which is 299×299 and scaled down to have each pixel range between 0 and 1. This processed frame then serves as input to our loaded trained model to make the probabilistic prediction for each output class. Label of the output class with the highest probability is marked as our final output. Some snips from our eyeglass detection framework predicting presence of different types of eyeglasses in real-time are attached later in the results in section 5.

5. RESULTS AND DISCUSSION

After successful collection and pre-processing of data as mentioned in section 3 and section 4.1 respectively, the model was trained on 80% of this data using Adagrad optimizer for the categorical cross entropy loss function using an initial learning rate value of 0.001 while using SoftMax activation for our final dense classification layer to get results as mentioned in the Table 4. The model was evaluated on the left out 20% of the above data termed as test data. To further test the robustness and performance of our model, another small custom-built dataset termed as real test data was used which includes subjects different from those in the main dataset. Two additional independent test datasets namely ORL and Sunglasses datasets were collected from published resources to further test the generalization capability of our model.

Table 4. Performance metrics for optimized model on different datasets

Dataset	Accuracy	Loss
Train Data (80%)	99.28	0.0326
Test Data (20%)	99.95	0.0026
Real Test Data	100.00	0.0004
ORL Test Data	94.38	0.1472
Sunglasses Test Data	98.20	0.0962

The above results clearly represent the robustness of our model in all conditions. The proposed model performs exceptionally well on training, testing and real-time testing datasets. As compared to the performance of model proposed by Jiang *et al.* [14] on ORL dataset, our model is able to generalize the results well with a recall value of 100% while the proposed edge likelihood probabilistic model in [14]

performed well on the in-house dataset (maximum fisher value 28.9) but gave comparatively poor results (maximum fisher value of 9.6) on ORL test dataset. As ORL dataset includes grayscale facial images, it becomes a little difficult to extract features from such images and hence gives slightly less accuracy when compared with other testing datasets. To visualize the results of our model as shown in Figure 8, the activations of our final dense classification layer were used to plot guided gradient and saliency maps respectively which represent the unique quality and contribution of each pixel in an input image. Finally, to test the performance of our model in real-time a custom-built framework was used for detecting the presence of eyeglasses in facial images as shown in Figure 9. Two different cameras of standard-quality and high-quality resolution were used respectively.

Figure 10 shows some of the images from the two independent test datasets that were misclassified by our trained model. Possible reasons for misclassification include extremely low-quality images as in the first two images. Also, the horizontal flip (upside down) and inappropriate cropped images resulted in difficulties while making positive predictions. It was also observed that certain facial images without eyeglasses were misclassified because of the shadow present in the ocular region. As a result of aging, the ocular region develops wrinkles which also results in dark patches accounting for misclassified results.

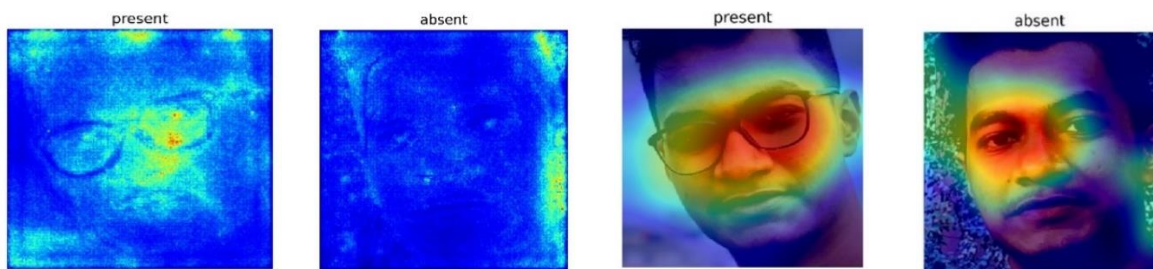


Figure 8. Guided gradient and saliency map

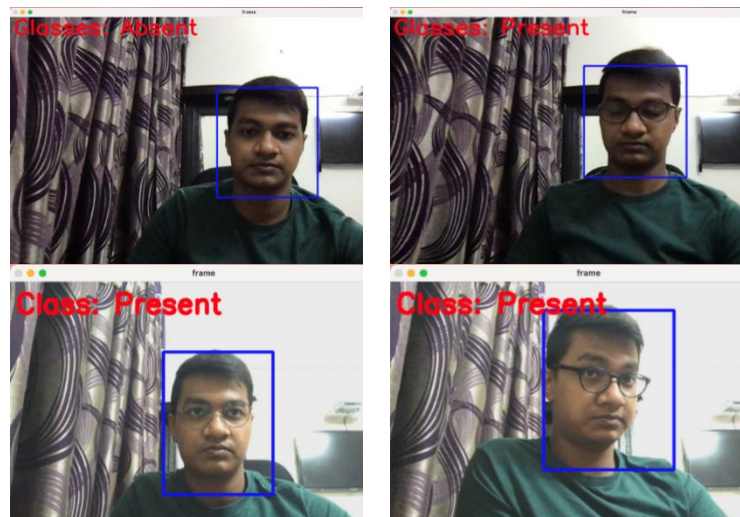


Figure 9. Real-time framework prediction sample

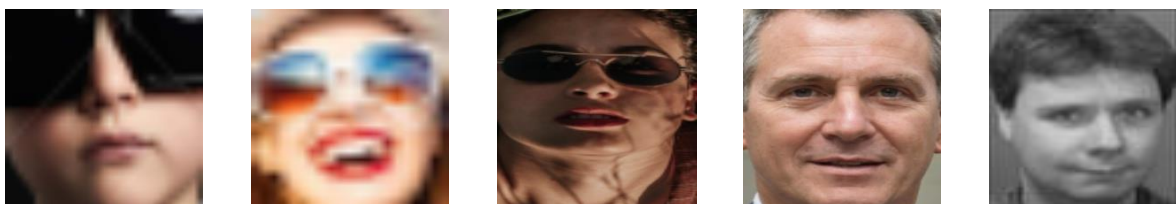


Figure 10. Misclassified test images

6. CONCLUSION

This paper presents an effective and efficient glasses detection model based on deep convolutional neural network and transfer learning. First, the input data is used and classified based on the with and without glasses images of the same person so as to make the detection of eyeglasses more accurate. Then the learned weights from pre-trained model were added to the corresponding layers according to various attributes to identify the glasses covering the ocular region can be detected. This model has proved to be highly efficient by giving a desirable accuracy of 99.9% on testing data in significantly less amount of time. Also, the model has been tested on two independent datasets while giving accuracy score of 98.2% and 94.4% respectively. Our model is robust enough to detect presence of all types of eyeglasses in real-time which gives us an edge over other similar models available. The main contribution of this paper is to design the robust eye-glass detection model using convolutional neural networks (CNN), which is expected to give a high accuracy on non-standard real life facial/ocular data.

Automatic eyeglasses detection is useful as eyeglasses offers obstruction during facial analysis and recognition tasks like security, surveillance, user authentication and other intelligent systems. The proposed detection algorithm can be extended for removal of eyeglasses from facial images and hence improve the accuracy of facial recognition algorithms. The model can be further generalized by adding more data from real life scenarios to boost the real-time performance.

ACKNOWLEDGEMENTS

The authors are grateful to the authorities of Birla Institute of Technology and Science Pilani, Dubai Campus for carrying out this research work.




REFERENCES

- [1] D. A. Reid, S. Samangoei, C. Chen, M. S. Nixon, and A. Ross, "Soft biometrics for surveillance: an overview," in *Handbook of statistics-Machine Learning: Theory and Applications*, 2013, pp. 327–352.
- [2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, Jan. 2004, doi: 10.1109/TCSVT.2003.818349.
- [3] B. Garcia and S. A. Viesca, "Real-time American sign language recognition with convolutional neural networks," *Convolutional Neural Networks for Visual Recognition*, vol. 2, pp. 225–232.
- [4] A. S. Mohammad, A. Rattani, and R. Derakhshani, "Comparison of squeezed convolutional neural network models for eyeglasses detection in mobile environment," *Journal of the ACM*, vol. 33, no. 5, pp. 136–144, 2018.
- [5] A. M. Basbrain, I. Al-Taie, N. Azeez, J. Q. Gan, and A. Clark, "Shallow convolutional neural network for eyeglasses detection in facial images," in *2017 9th Computer Science and Electronic Engineering (CEECE)*, Sep. 2017, pp. 157–161, doi: 10.1109/CEECE.2017.8101617.
- [6] W. Hao, R. Bie, J. Guo, X. Meng, and S. Wang, "Optimized CNN based image recognition through target region selection," *Optik*, vol. 156, pp. 772–777, Mar. 2018, doi: 10.1016/j.ijleo.2017.11.153.
- [7] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional neural network (CNN) for image detection and recognition," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, Dec. 2018, pp. 278–282, doi: 10.1109/ICSCCC.2018.8703316.
- [8] W. Mo, X. Luo, Y. Zhong, and W. Jiang, "Image recognition using convolutional neural network combined with ensemble learning algorithm," *Journal of Physics: Conference Series*, vol. 1237, no. 2, Jun. 2019, doi: 10.1088/1742-6596/1237/2/022026.
- [9] A. Fernández, R. García, R. Usamentiaga, and R. Casado, "Glasses detection on real images based on robust alignment," *Machine Vision and Applications*, vol. 26, no. 4, pp. 519–531, May 2015, doi: 10.1007/s00138-015-0674-1.
- [10] S. Du, J. Liu, Y. Liu, X. Zhang, and J. Xue, "Precise glasses detection algorithm for face with in-plane rotation," *Multimedia Systems*, vol. 23, no. 3, pp. 293–302, Jun. 2017, doi: 10.1007/s00530-015-0483-4.
- [11] S. Bekhet and H. Alahmer, "A robust deep learning approach for glasses detection in non-standard facial images," *IET Biometrics*, vol. 10, no. 1, pp. 74–86, Jan. 2021, doi: 10.1049/bme2.12004.
- [12] S. Mehta, S. Gupta, B. Bhushan, and C. K. Nagpal, "Face recognition using neuro-fuzzy inference system," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 7, no. 1, pp. 331–344, Feb. 2014, doi: 10.14257/ijsp.2014.7.1.31.
- [13] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, and Z. Zhu, "Robust face recognition via occlusion dictionary learning," *Pattern Recognition*, vol. 47, no. 4, pp. 1559–1572, Apr. 2014, doi: 10.1016/j.patcog.2013.10.017.
- [14] X. Jiang, M. Binkert, B. Achermann, and H. Bunke, "Towards detection of glasses in facial images," in *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, 2000, vol. 2, pp. 1071–1073, doi: 10.1109/ICPR.1998.711877.
- [15] A. S. Mohammad, A. Rattani, and R. Derakhshani, "Eyeglasses detection based on learning and non-learning based classification schemes," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, Apr. 2017, pp. 1–5, doi: 10.1109/THS.2017.7943484.
- [16] P. Drozdowski, F. Struck, C. Rathgeb, and C. Busch, "Detection of glasses in near-infrared ocular images," in *2018 International Conference on Biometrics (ICB)*, Feb. 2018, pp. 202–208, doi: 10.1109/ICB2018.2018.00039.
- [17] Z. Jing and R. Mariani, "Glasses detection and extraction by deformable contour," in *Proceedings - International Conference on Pattern Recognition*, 2000, vol. 15, no. 2, pp. 933–936, doi: 10.1109/icpr.2000.906227.
- [18] M. Z. Lazarus, S. Gupta, and N. Panda, "An Indian facial database highlighting the spectacle problems," *IEEE International Conference on Innovative Technologies in Engineering 2018 (ICITE OU)*, 2018, doi: 10.21227/ay8v-kj06.
- [19] J. Heaton, "Glasses or no glasses." Kaggle, 2020, Accessed: Apr. 20, 2021. [Online]. Available: <https://www.kaggle.com/jeffheaton/glasses-or-no-glasses>.




- [20] J. Zhang, Y. Yan, and M. Lades, "Face recognition: eigenface, elastic matching, and neural nets," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1423–1435, 1997, doi: 10.1109/5.628712.
- [21] Amol, "Sunglasses/no sunglasses." Kaggle, 2021, Accessed: Apr. 21, 2021. [Online]. Available: <https://www.kaggle.com/amol07/sunglasses-no-sunglasses>.
- [22] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, Dec. 2016, doi: 10.1186/s40537-016-0043-6.
- [23] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," in *Proceedings of the IEEE*, Nov. 2019, vol. 109, no. 1, pp. 43–76.
- [24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.
- [25] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv:1609.04747v2*, Sep. 2016.
- [26] N. Mahankali and V. Ayyasamy, "OpenCV for computer vision applications," in *Proceedings of National Conference on Big Data and Cloud Computing (NCBDC'15)*, 2015, pp. 52–56.

BIOGRAPHIES OF AUTHORS






Ritik Jain    is a final year B.E. Computer Science student at BITS Pilani, Dubai Campus. He is the General Secretary of Microsoft Tech Club at BITS Pilani, Dubai Campus and a Microsoft Learn Student Ambassador since 2020. He is a machine learning enthusiast and has worked on various projects in the same field. His research interests are in the areas of Computer Vision, Sentiment Analysis, and Intelligent Automated Systems. He can be contacted at email: ritikjain@gmail.com.



Aashi Goyal    is a final year B.E. Computer Science student at Birla Institute of Technology and Science (BITS) Pilani, Dubai Campus. She has worked on various projects in the same field. She is currently the E-Portal lead at the Artificial Intelligence club at BITS Pilani, Dubai Campus. She has her work published at an international conference. Her area of research interest includes computer vision, prediction and sentiment analysis. She can be contacted at email: aashigoyal2109@gmail.com.



Kalaichelvi Venkatesan    is an Associate Professor in the Department of Electrical and Electronics Engineering at BITS Pilani, Dubai Campus. She is working with Bits Pilani, Dubai Campus since 2008 and she is having 29 years of teaching experience. She has her research work published in refereed international journals and many international conferences. She has also reviewed many papers in International Journals and Conferences. Her research area of expertise includes Process Control, Control Systems, Neural Networks, Fuzzy Logic and Computer Vision. She is currently guiding students in the area of Intelligent Control Techniques applied to Robotics and Mechatronics. She can be contacted at email: kalaichelvi@dubai.bits-pilani.ac.in.