https://doi.org/10.33472/AFJBS.6.Si3.2024.459-467

AFJBS

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Enhancing Facial Gender Classification: A Deep Dive into Xception Neural Networks and Ensemble Frameworks

Anshar Ali N.[1], Dr.Thirupathi Regula[2]

[1]Research Scholar, Department of Computer Science and Engineering, Shri Venkateshwara University, Uttar Pradesh, India.
[2]Assistant Professor, Department of Computer Science and Engineering, Shri Venkateshwara University, Uttar Pradesh, India.

Email: [1]ansharali.dl@gmail.com, [2]regulathirupathi@gmail.com

Article Info

Volume 6, Issue Si3, May 2024

Received: 08 April 2024

Accepted: 03 May 2024

Published: 30 May 2024

doi: 10.33472/AFJBS.6.Si3.2024.459-467

ABSTRACT:

Gender plays an important role in many areas like health care, E-commerce. In light of the ongoing growth in the computer vision domain, various applications that depend on extracting biometric data, such as facial gender identification for purposes like access control, security, or marketing, are gaining prevalence. A standard gender classifier necessitates a substantial number of training samples to grasp a wide range of discernible features. Convolutional Neural Networks (CNNs) have established their effectiveness in tasks involving image classification, and Xception, an extension of the Inception architecture, has showcased cutting-edge performance across diverse computer vision applications. This research explores the application of the Xception neural network architecture for gender classification using facial images. The dataset used for training and evaluation consists of diverse facial images representing a broad spectrum of ages, ethnicities, and expressions. To improve the model's capability to capture significant features from facial images, preprocessing methods are employed. The devised model incorporates a pre-trained Xception model, integrated into an ensemble framework comprising Support Vector Machines (SVM), Random Forest, and AdaBoost. Our experimental outcomes indicate that training the model on a set of synthetic images yields performance comparable to existing state-of-the-art methods, which utilize authentic images of individuals. The average classification accuracy for each classifier falls within the range of 94% to 95%, mirroring the performance of previously proposed approaches.

Keywords: SVM, random Forest, Xception, residual block, adaboost, ensemble classifier, ReLU, separable convolutions, feature maps.

1. Introduction

In Biology, human gender classification generally refers to the biological attributes that determine an individual's classification as male, female, or third gender. Machine Learning (ML) methods, including but not limited to CNNs, are employed for gender classification due to their ability to learn patterns from data. ML can be applied to diverse data types, including images, voice recordings, or textual data, making it versatile for gender classification across different modalities [8]. ML-based gender classification can be integrated into human-computer interaction systems to create personalized user experiences. Interfaces and content can be adapted based on the identified gender, enhancing user satisfaction. ML-based gender classification is used in healthcare research for analyzing demographic trends, studying gender-specific health issues, and improving public health planning [9]. ML algorithms can be employed for gender classification in social media analytics. This information is valuable for businesses, advertisers, and researchers studying online behavior. ML models provide automated and data-driven insights, aiding decision-makers in making informed choices based on gender-specific patterns and trends.

Machine learning models have the potential to adopt biases existing in their training data, resulting in predictions that reflect those biases [10]. Gender classification may infringe on privacy rights, especially if applied without consent. Some datasets may lack diversity, leading to models that may not generalize well across different demographics. Gender classification based on appearance may not accurately reflect an individual's gender identity. The effectiveness of machine learning models relies significantly on the quality and representativeness of the training data. When the training data exhibits biases toward specific demographics, the model is susceptible to inheriting and perpetuating those biases. This can result in unfair predictions and the reinforcement of stereotypes. Training datasets may not fully represent the diversity of gender identities and expressions [11]. This can result in models that are less accurate when applied to individuals outside the demographics represented in the training data. ML models may inadvertently amplify existing biases present in the training data.

A CNN design that works well for classifying genders is called Xception. The design is efficient and lightweight, and it can be trained with comparatively minimal datasets. It is thus perfect for use in embedded systems or on mobile devices. The following are some benefits of classifying gender using Xception. It has been shown that Xception produces state-of-the-art outcomes on many criteria for gender categorization. Xception is an architecture that is lightweight, efficient, and trainable on comparatively tiny datasets. It is thus perfect for use in embedded systems or on mobile devices. With extensive image classification datasets like ImageNet, Xception may be pre-trained. This enables it to pick up characteristics that are helpful for a range of other tasks, such as identifying gender [12]. A basic gender categorization model based on Xception will be trained using this code. A collection of photos together with the matching gender classifications will be used to train the model. The model may be used for predicting the gender of fresh photos after it has been trained. An effective method for classifying gender is Xception. Its architecture is lightweight, effective, and capable of producing cutting-edge outcomes [13].

Xception's gender categorization feature is a strong tool with several uses. Its architecture is lightweight, effective, and capable of producing cutting-edge outcomes. Its performance on problems involving gender categorization may be further enhanced by using it for transfer learning as well. It has been shown that this produces cutting-edge outcomes on several gender categorization standards. The accuracy of other security systems, such as face recognition software, may be increased by using gender categorization [14]. Ads may be targeted to the correct audience and marketing efforts can be made more personalised by using gender categorization. In order to provide gender-specific therapies and get a deeper understanding of how gender affects illness and treatment, gender categorization may be used in the healthcare industry. In education, gender categorization may be used to identify pupils who might require more help and to provide individualised learning experiences [15]. The design is efficient and lightweight, and it can be trained with comparatively minimal datasets. It is thus perfect for use in embedded systems or on mobile devices. Large image classification datasets like ImageNet may be used to pre-train it. This enables it to pick up characteristics that are helpful for a range of other tasks, such as identifying gender. The working of the traditional Xception is shown in figure 1.

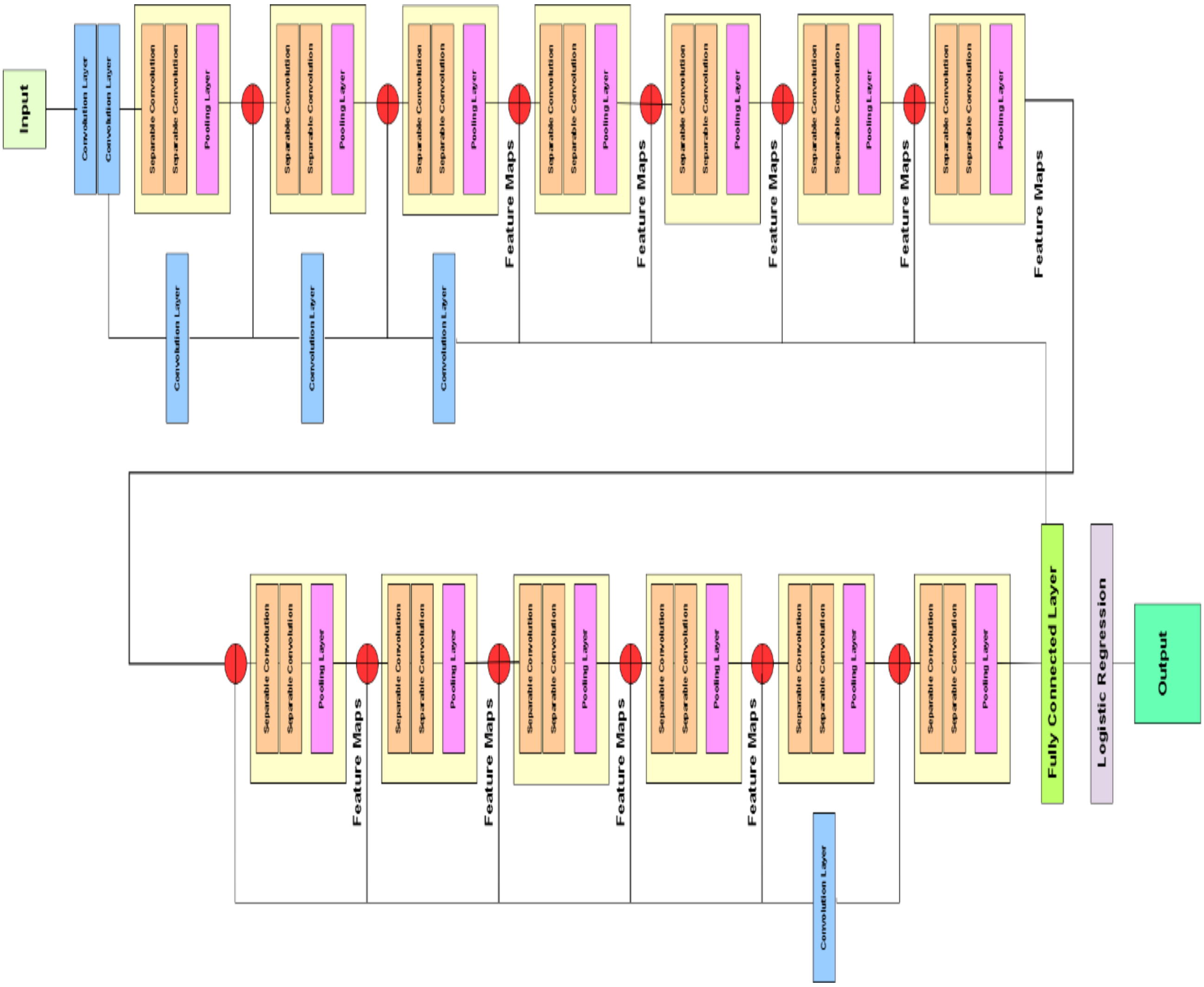


Figure 1: Traditional Architecture of Xception Model

2. Literature Survey

In [1], Valliappan Raman et al. stated that because of its resilience, deep learning is the recommended technique for facial analysis. A StyleGAN-generated dataset of synthetic facial images is proposed, exhibiting great accuracies that are on par with current techniques. To lessen the requirement for genuine facial photos and lower privacy threats, the study advises employing StyleGAN-generated examples. However, a short number of training samples might cause a model's accuracy to decline without an early termination. Increasing the sample size or using image augmentation can help the model become more accurate. A custom-built model may not perform as well as a modified pre-trained model such as VGG16, especially if the two were trained on unrelated tasks. For pre-trained models, early halting shortened the training period. To create visuals contingent on age, ethnicity, or posture, more study is required. It is advised to use crowdsourcing to increase the number of viewpoints on the generated images' labels and enhance the precision of the ground truth label.

In [2], Ammar Almomani et al. said that voice gender, age, and accent classification techniques included the use of back-propagation and bagging algorithms. The goal of the project is to create a system known as CVGAA that uses backpropagation and bagging methods, which are very accurate and precise in speech recognition systems, to predict and categorise gender, age, and accent. A voice's gender was determined using the first dataset. The second dataset is a speech corpus called Common Voice. The algorithm that bagged data had the best accuracy in classifying people by age. Out of all the methods, the Bagging technique also had the highest accuracy. People were identified using the third dataset according to their linguistic origins. The authors suggest that in order to acquire embeddings from raw waveforms, future research might investigate more recent designs such as wav2vec2.0. MLS or CommonVoice datasets might be used to adapt the suggested technique to additional languages.

In [3], Hina Tabassum et al. stated that the measurement of biological changes in a population is done through the use of anthropometry, a subfield of morphometry. It entails measuring anatomical components, which are relevant to anthropology, epidemiology, and psychology. For dimension reduction in huge datasets, classification techniques like clustering and decision trees are employed. A minimum of twelve anthropometric measurements are required for gender categorization in order to train the classifier. In this investigation, a boosting tree method yielded an accuracy rate of 98.42% with twelve important factors. Nevertheless, the study is restricted to highly skilled statistical software users and sophisticated research methodologies. Additionally, this research study's literature evaluation is lacking. Future technological developments could make it simpler to forecast offenders or similar people with minimal research.

In [4], Muhammad Mustapha et al. stated that improvements in information technology and computer science have led to a growth in computer intelligence and capabilities. Using a limited dataset, the proposed study investigates the impact of a twofold face detection technique on the accuracy of gender classification, yielding 84% accuracy for double face detection and 74% accuracy for single face detection. A balanced collection of face image data for gender categorization may be found in the FEI dataset. Greater training dataset size increases model accuracy. For certain issues, small datasets can yield great accuracy; one example of this is the FEI dataset, which has one face per image. The suggested study was assessed using the FEI dataset, yielding 84% accuracy for Experiment 2 and 74% accuracy for Experiment 1.

In [5], Faycel Abbas discussed the local binary pattern and how it relates to gender categorization based on handwriting. It is suggested to use a support vector machine classifier in conjunction with local characteristics to classify gender in online handwriting. Research indicates a relationship between writer gender and handwriting. This study examines several methods for identifying writer gender from scanned handwriting images, obtaining a 76.68% classification rate on a dataset of Arabic and English writing samples. The MLBP descriptor is assessed using an SVM classifier in this technique, which makes use of a variety of texture-based measures. Trials on a subset of the writer identification dataset from Qatar University show how successful the strategy is. A portion of the QUWI dataset—which consists of contributions from authors who have submitted four manuscripts in both Arabic and English—was utilised for the system's experimental assessment. The suggested strategy yields a classification rate that is comparable to other approaches reported in the literature and is appropriate for classifying gender.

In [6], Kota Sandeep et al revealed that the lack of readily available, precisely labelled datasets may be addressed by using a simple network design for age and gender categorization. Training distinct age classifiers for each gender can enhance gender and age categorization, which is crucial for clinical practise, behavioural science, and human-machine interactions. The necessity for private and uncommon personal information, such as gender and date of birth, is the reason for data scarcity. Using live video frames, a convolution neural network model has been created to classify people's age and gender in real time. The findings demonstrate the viability of employing neural networks for real-time human emotion detection.

In [7], Mohsen Rohani et al. report that a deep learning architecture is used in this study to provide a method for determining gender, age, and grin from face photos. Using a multi-task deep learning framework, a method was shown for detecting age, gender, and grin from face photos. This method surpassed recently developed advanced approaches in terms of accuracy and boosted total accuracy by using characteristics from many layers. The technique was tested on datasets and yielded results that were on par with state-of-the-art techniques for gender classification, age identification, and grin recognition. The procedure entails preprocessing the photos and extracting faces with a convolutional neural network. Through the use of a D-CNN and feature extraction from the GENKI-4K and IMD-WIKI datasets, the suggested framework increases the accuracy of grin recognition. Three tasks from different datasets were used in the analysis. Table 1 presents the various analysis on the existing approaches

Table 1: Comparative Analysis on the Existing Approaches

S. No	Author	Method	Merits	Demerits	Accuracy
1	Valliappan Raman	StyleGAN	Less time to process, augmentation included to improve accuracy	Misclassification of samples, not much diversity is explored, only binary classification	84.78%
2	Ammar Almomani	CVGAA	Removed all the duplicates, simple execution	Some incorrect classifications should be taken care of, and can be improved by optimization	87.1%
3	Hina Tabassum	boosting tree algorithm	Dimensionality reduction has reduced the computational time, easy to understand	Lengthy process, should work on more existing procedures	88.42%
4	Muhammad Mustapha	twofold face detection approach	FaceNet performed feature extraction, also detects multiple faces in a single frame.	More likely to identify male faces than the alter, mis-classifications	84%

5	Faycel Abbas	SVM	LBP, unique identification process.	More number of features are selected so processing will be more, binary classification	76.68%
6	Kota Sandeep	CNN	Automated process, also implemented transfer learning.	Need more memory, takes more time.	-
7	Mohsen Rohani	D-CNN	Also detects other parameters, evaluated on 3 different datasets	Scalability, robustness on real-time detection	86.63%

2.1. Materials Used: There are several datasets available for gender classification, covering different domains and modalities. Some popular datasets include:

1. CelebA : This dataset has 40 attributes. It is used for large scale requirement of such data. There are nearly 200,000 images in the dataset. It has binary class label.
2. Adience: This dataset encompasses the attributes under consideration, specifically age and gender labels. These attributes play a crucial role in the analysis and categorization of the collected data. The dataset comprises a variety of images sourced from the internet. The diversity in the image collection enhances the comprehensiveness of the dataset, providing a wide range of visual information for analysis. Within this category, the focus is on the classification of gender into two distinct classes: Male and Female. This binary classification system simplifies the categorization process and facilitates gender-related analysis in the dataset.
3. Labeled Faces in the Wild (LFW): While primarily designed for face recognition tasks, these labels extend their usability to include gender classification. This dual-purpose classification system enhances the dataset's versatility, accommodating a range of facial analysis applications. The inclusion of labeled identities enhances the dataset's utility for tasks such as identity recognition and verification.
4. IMDB-WIKI: The inclusion of age information, along with gender labels, enriches the dataset, providing valuable insights for analyses related to both demographic characteristics. Within this category, the dataset is composed of images obtained through crowdsourcing, sourced specifically from IMDb and Wikipedia. The use of crowdsourced images ensures diversity and a broad representation of individuals, contributing to the dataset's comprehensiveness.
5. UTKFace: The dataset consists of facial images specifically chosen for their diversity in terms of age, gender, and ethnicity. This intentional focus on diversity ensures a representative collection, catering to a broad range of demographic factors and making the dataset versatile for various analyses. The inclusion of age, gender, and ethnicity labels enhances the dataset's richness, allowing for a comprehensive exploration of demographic characteristics.
6. Gender Classification: Gender, as a construct shaped by societal norms, influences the distinct treatment received by males and females from birth, shaping their behaviors and preferences to align with societal expectations for their respective genders. This modest dataset is crafted to explore the possibility of predicting an individual's gender with an accuracy significantly

surpassing 50%, based on their personal preferences. Comprising 5 attributes, the data is sourced from individuals representing 21 different nationalities.

7. Gender Classification Dataset: The dataset comprises cropped images of both male and female subjects, segregated into training and validation directories. The training set encompasses approximately 23,000 images for each gender class, while the validation directory contains around 5,500 images for each class. This division facilitates the development and assessment of models on distinct subsets for effective analysis and validation.
8. Simple Gender Classification: This dataset can be used for various classification and data analysis tasks, such as predicting an individual's gender or income level based on their occupation, education level, and other demographic factors.
9. Gender Classification 200K Images: This dataset is used for Gender Classification with images. The dataset consists of almost 200K images which are almost 1.3GB in size. The dataset has 10 columns. Table 2 presents the analysis of different datasets.

Table 2: Analysis on the Dataset of Existing Dataset

S. No	Dataset	No of attributes	Class label
1.	CelebA	40	Male, Female
2.	Adience	Age, Gender	Male, Female
3.	Labeled Faces in the Wild (LFW)	Identity labels	Male, Female
4.	IMDB-WIKI	Age, Gender	Male, Female
5.	UTKFace	Age, Gender, Ethnicity	Male, Female
6.	Gender Classification	5 attributes	Male -M, Female- F
7.	Gender Classification Dataset	58700 images approx.	Male, Female
8.	Simple Gender Classification	10 attributes	Male, Female
9.	Gender Classification 200K Images	203k images	Male, Female

3. Proposed Methodology

The need for gender classification using Convolutional Neural Networks (CNNs) arises from various applications such as security, marketing, healthcare, and more. CNNs are particularly well-suited for image-based tasks, making them effective for classifying gender based on facial features or body images. CNNs are particularly well-suited for image-based tasks due to their ability to automatically learn hierarchical features from data. This makes them effective for extracting complex patterns and representations from facial or body images, which are commonly used for gender classification. Gender classification often relies on the recognition of facial features, and CNNs excel at analyzing spatial relationships in images. CNNs excel at learning hierarchical representations of features. In the context of gender classification, these features might include facial expressions, hair length, and other subtle cues that are relevant for distinguishing between male and female individuals. The proposed model integrates the Xception model with ensemble classifier because if some individual models, including the Xception model, are sensitive to noise or outliers in the data, the ensemble approach can help smooth out these variations. Outliers or misclassifications from one model may be compensated by correct predictions from other models in the ensemble.

3.1. Framework of Xception Model: Xception's use of these three types of blocks and depthwise separable convolutions contributes to its efficiency and effectiveness in image classification tasks.

3.1.1. Entry Flow Block in Xception: This block contains 4 major components i.e., seperable convolutions which can efficiently capture spatial hierarchies and reduce the number of parameters. Residual & short cut connections, these connections play a crucial role in the information flow. They enable the model to learn both the high-level abstract features and the identity mapping when necessary. Batch normalization helps to regularize the learning process, improving the model's ability to generalize to unseen data. The mathematical computation of this block is shown in equation (1)

$$Entry\_Output = \gamma(\frac{\sum_{i=1}^{n} W_i * Input_i - \mu}{\sigma}) + \beta - (1)$$

A group of convolutional layers make up the entrance flow block in Xception, which is used to analyse input images and extract information at various sizes. It is composed of the following levels. ReLU and batch normalisation are applied after two typical 3x3 layers of convolution, the first with 32 filters & the second with 64. A depthwise convolution and a pointwise convolution are the two components of each of the three depthwise separated convolutional blocks. There are 128 filters per block. two strides and a 3x3 kernel are used in the max pooling layer. The entrance flow block is intended to increase the total amount of channels while progressively decreasing the input image's spatial dimension. This aids in the network's ability to extract more intricate elements from the image. A 19x19x728 tensor is the entrance flow block's output. Xception's middle flow block, which is made up of eight similar blocks, receives this tensor next. Batch normalisation, residual connections, and depthwise separable convolutions are all present in each block. An essential component of Xception is the entrance flow block, which aids in the network's ability to extract features at various sizes from the input image. To perform well on image segmentation and classification tasks, this is necessary. Figure 2 presents the entry block of Xception module

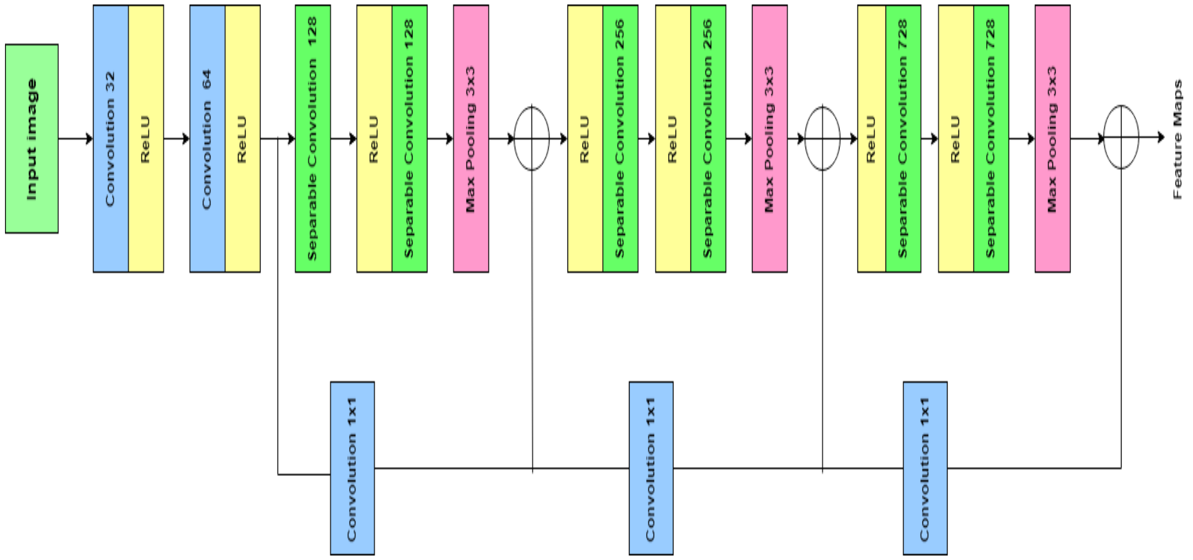


Figure 2: Architecture for the Entry Block

3.1.2. Middle Flow Block in Xception: The Middle Flow Block in Xception is responsible for building deeper representations of features by repeating a specific set of operations. It helps capture complex patterns and hierarchies in the data. The Depthwise Separable Convolution is a key component of the Middle Flow Block in the Xception architecture. This type of convolutional operation is designed to be more parameter-efficient compared to traditional convolutions.

Table 3: Trade-off between convolution layers

Traditional Convolution	**Depthwise Convolution**
In a standard convolutional layer, a filter/kernel is applied to the entire input volume, considering all input channels. This operation involves a significant number of parameters, especially when dealing with 3D data	Depthwise Separable Convolution decomposes the standard convolution into two steps: depthwise convolution and pointwise convolution.

3.1.2.1. Depthwise separable convolutions in Xception

One kind of convolution that may be utilised to lower the quantity of parameters and calculations needed for inference is depthwise separable convolutions. A CNN design called Xception substitutes depthwise separable convolutions for conventional convolutions. In order to create depthwise separable convolutions, a regular convolution is split into two steps: Sub-depth convolution: A convolution that is carried out separately for every channel in an input tensor is called a depthwise convolution. This indicates that every channel has its own filter. Convolution carried out using a 1x1 kernel is referred to as a pointwise convolution. This indicates that just one pixel at time is affected by the filter. While it is more efficient, the result of combining a depthwise & pointwise convolution is the same as a regular convolution. This is due to the fact that the pointwise convolution and the depth-wise convolution only need to be done once per channel and channel, respectively. Xception's convolutional blocks are all based on depthwise separable convolutions. As a result, Xception has one of the best CNN architectures out there. Figure3 presents the depth wise seperables

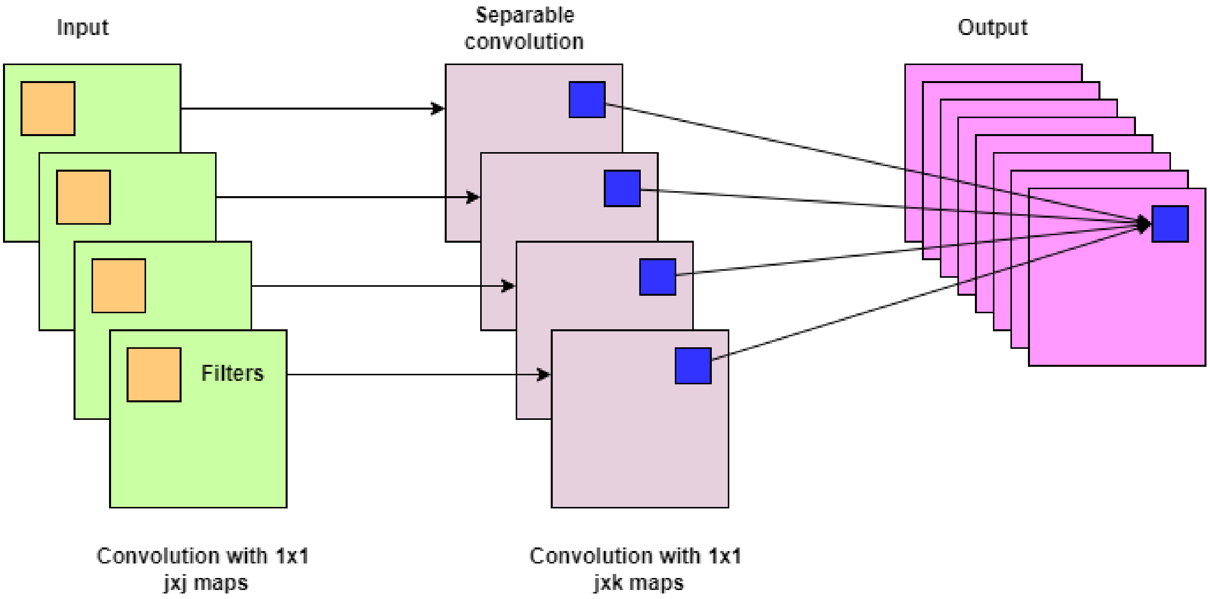


Figure 3: Working of Depth Wise Seperables

In the Middle Flow Block, residual connections are added around each depthwise separable convolutional layer. The output of the depthwise separable convolution is added to the original input, promoting the flow of information and facilitating training. Eight similar blocks make up the middle flow block in Xception, which is utilised to extract more intricate information from the input image. The following layers make up each block. a 1x1 depth-separable conv with 128 adjustments, batch normalisation, and ReLU thereafter. a 128-filter, 3x3 depthwise separable convolution, followed by batch norm and ReLU. a 1x1 point-wise conv with 128 adjustments, then batch norm and ReLU. a residual link that increases the pointwise convolution's output by the input tensor. The centre flow blocks is designed to progressively add more channels to the feature maps while maintaining the same spatial dimensions. This facilitates the network's learning of more intricate input image representations. A 19x19x728 tensor is the middle flow block's output. After that, this tensor is sent to Xception's exit flow block, which is made up of a layer of pooling global averages, a pointwise convolution, & a depthwise separable convolution. An essential component of Xception is the central flow unit, which aids in the network's ability to learn intricate models of the input image. To perform well on image segmentation and classification tasks, this is necessary. Figure 4 presents the middle flow block

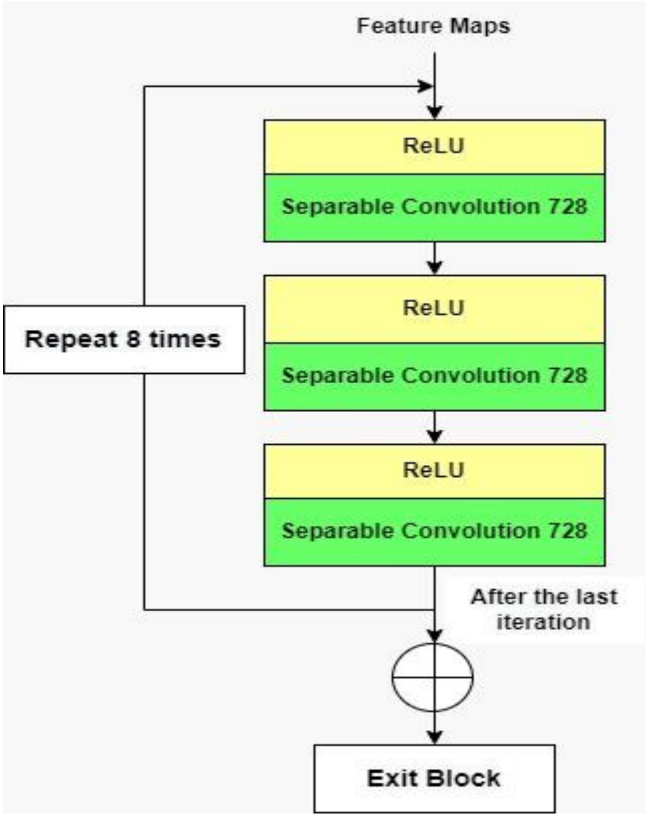

Feature Maps
ReLU
Separable Convolution 728
ReLU
Separable Convolution 728
ReLU
Separable Convolution 728
Repeat 8 times
After the last iteration
Exit Block

Figure 4: Middle Flow Architecture

3.1.3. Exit Flow Block in Xception: The exit block contains 5 components out of which, the main components are residual block and average global pooling block.

3.1.3.1 Residual Block: A Residual Block is a fundamental component of deep neural networks, introduced to address the vanishing gradient problem. It consists of a shortcut connection that allows the gradient to flow more easily during backpropagation, facilitating the training of very deep networks. Residual Blocks are commonly used in architectures like ResNet (Residual Networks) and have significantly contributed to the success of deep learning models in various tasks. the residual block in X promote Xception feature reuse, facilitates gradient flow, and contributes to the model's parameter efficiency, leading to improved performance in tasks such as image classification. Residual blocks address the vanishing gradient problem, which can occur in deep networks during backpropagation. The presence of shortcut connections allows gradients to flow directly through the network, facilitating the training of very deep networks.

The middle flow block's feature maps are processed by a group of layers in Xception's exit flow block to create the final output tensor. It is composed of the following levels. a 728-filter depth-wise separable convolution, is the central flow unit. A 2048-filter pointwise convolution is succeeded by ReLU & a batch norm. an average pooling layer worldwide. The exit flow block is made to increase the number of channels while progressively decreasing the overall dimension of the feature maps. In this way, the input image's more abstract representations are taught to the network. After the feature maps are transformed into a single vector by the global average pooling layer, they may be sent to a fully connected layer for regression or classification. The exit flow block produces a vector of 2048 dimensions as its output. A FC layer may then receive this vector for regression or classification. Part of Xception that is crucial is the exit flow block, which aids the network in learning abstract models of the input image and generates an output vector that may be used for regression or classification. Figure 5 components of Exit Block

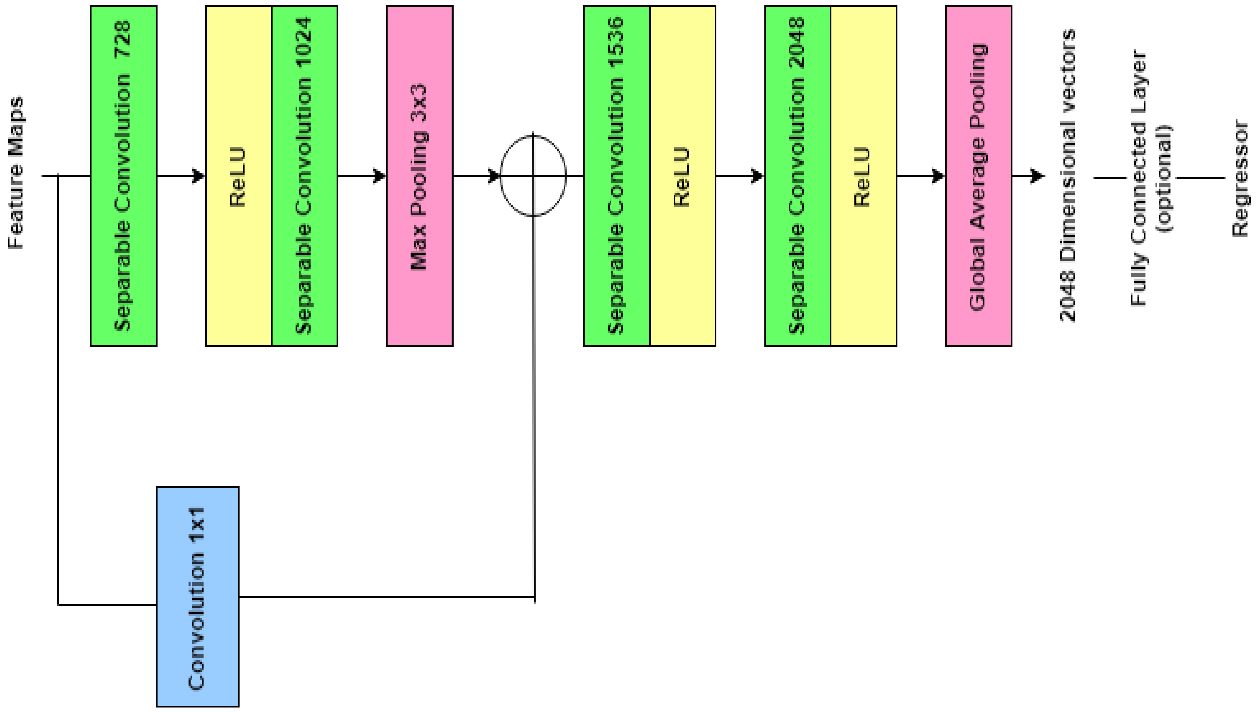

Feature Maps
Separable Convolution 728
ReLU
Separable Convolution 1024
Max Pooling 3x3
Separable Convolution 1536
ReLU
Separable Convolution 2048
ReLU
Global Average Pooling
2048 Dimensional vectors
Fully Connected Layer (optional)
Regressor
Convolution 1x1

Figure 5: Components of Exit Block

3.2. Integration with Ensemble Classifier: Ensemble can further enhance transfer learning by combining knowledge from multiple pre-trained models, each fine-tuned for different aspects of the task. When the system has a limited amount of task-specific labeled data, training a robust model from scratch might be challenging. Ensemble pre-trained models with task-specific fine-tuning or additional models can enhance the overall model's performance, leveraging the knowledge encoded in the pre-trained weights. The proposed research combines 3 different categories of algorithms ranging from basic to boosting.

3.2.1. SVM Decision for Gender Classification: The hybrid approach leverages the strengths of both the pre-trained model and the SVM for an effective gender classification system. SVMs can use different kernels (e.g., linear, polynomial, radial basis function) to handle non-linear relationships in the data. Fine-tuning the kernel and its associated parameters can significantly impact the model's ability to capture complex patterns.

3.2.1.1. Fine Tune Parameters of SVM:

A. Need of probability parameter in SVM tuning: Support Vector Machines (SVMs) belong to the category of supervised learning algorithms utilized for tasks involving classification and regression. SVMs operate by identifying the hyperplane that most effectively segregates data points into distinct classes. When engaging in the fine-tuning of SVMs, the concept of probability parameters may arise, particularly within the realm of soft-margin SVMs and probabilistic classification. The inclusion of the probability parameter in SVM tuning is closely linked to soft-margin SVMs and proves pertinent when employing SVMs for classification purposes. In the context of soft-margin SVMs, the primary goal is to determine a hyperplane that not only separates the data but also minimizes classification errors while maximizing the margin between classes. Referred to as the probability parameter, this parameter, commonly denoted as "probability" in SVM implementations such as scikit-learn in Python, provides the option to activate or deactivate probability estimates for SVM classification. The awareness of probability estimates becomes crucial when considering the adjustment of the decision threshold for class assignment.

B. Kernels: The extracted features from Xception are then used as input to the SVM with a polynomial kernel. The polynomial kernel function takes the dot product of the feature vectors and adds a constant term. It then raises the result to a specified degree. The selection of degree and their updation process is shown in figure 6

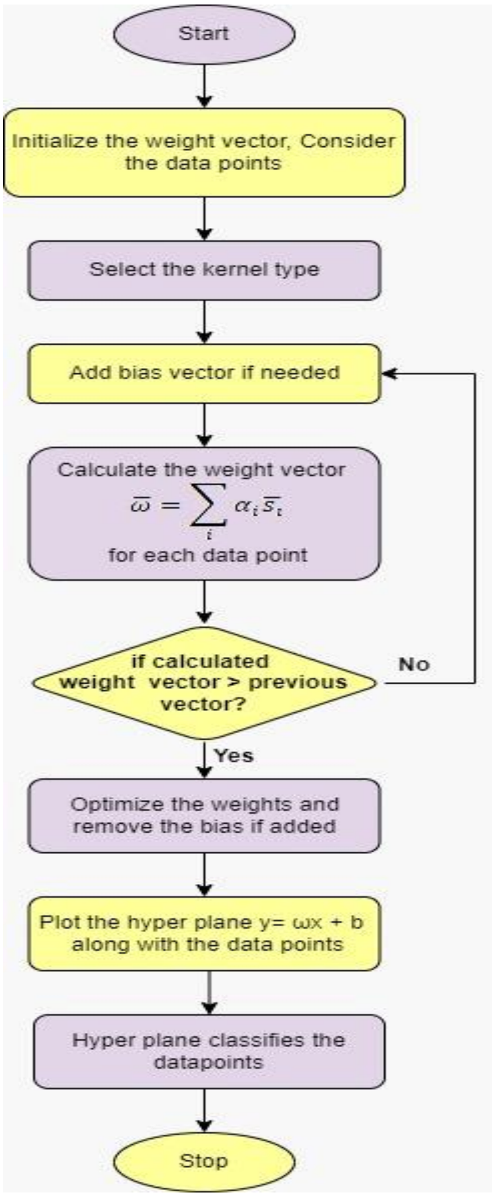

Start
Initialize the weight vector, Consider the data points
Select the kernel type
Add bias vector if needed
Calculate the weight vector
$\overline{\omega} = \sum_i \alpha_i \overline{s_i}$
for each data point
if calculated weight vector > previous vector?
No
Yes
Optimize the weights and remove the bias if added
Plot the hyper plane y= ωx + b along with the data points
Hyper plane classifies the datapoints
Stop

Figure 6: Fine Tuning of SVM based on Kernel Type

3.2.2. Random Forest for Gender Classification: The algorithm builds multiple decision trees during training and outputs the class that is the mode of the classes (classification) of the individual trees. For each decision tree, a random subset of features is selected at each split. This introduces additional randomness and ensures that different trees in the forest focus on different sets of features. The proposed model tunes 3 important parameters in this approach. Table 4 presents the fine tuning of random forest

Table 4: Fine Parameters in Random Forest

Name	Description	Values
n_estimators	It determines the number of trees that will be grown in the forest during the training phase	Multiples of 1000
random_state	Setting a specific seed ensures that the random processes, such as the random sampling of data points for building trees or the random selection of features at each split, are reproducible	Any integer value
criteria	It determines the function used to measure the quality of a split at each node of a decision tree.	1. Gini impurity is minimized when all the data points in a node belong to a single class, indicating a pure node. 2. In entropy, the goal is to maximize the reduction in entropy at each split, leading to more organized and homogenous nodes.

3.2.3. Boosting Xception Model Features: AdaBoost gives less weight to misclassified samples, which can make the overall model more robust to outliers and noisy data points. This can be advantageous in scenarios where the Xception model may encounter outliers during training. AdaBoost can be adapted for multi-class classification problems using a technique called "AdaBoost.MH" (AdaBoost with Multi-class Hypotheses). It initializes the each weights using the equation (2)

$$Weight_i = \sum_{i=1}^{n} \frac{1}{nK} * \log\left(\frac{1-W_{i-1}}{W_{i-1}}\right) - (2)$$

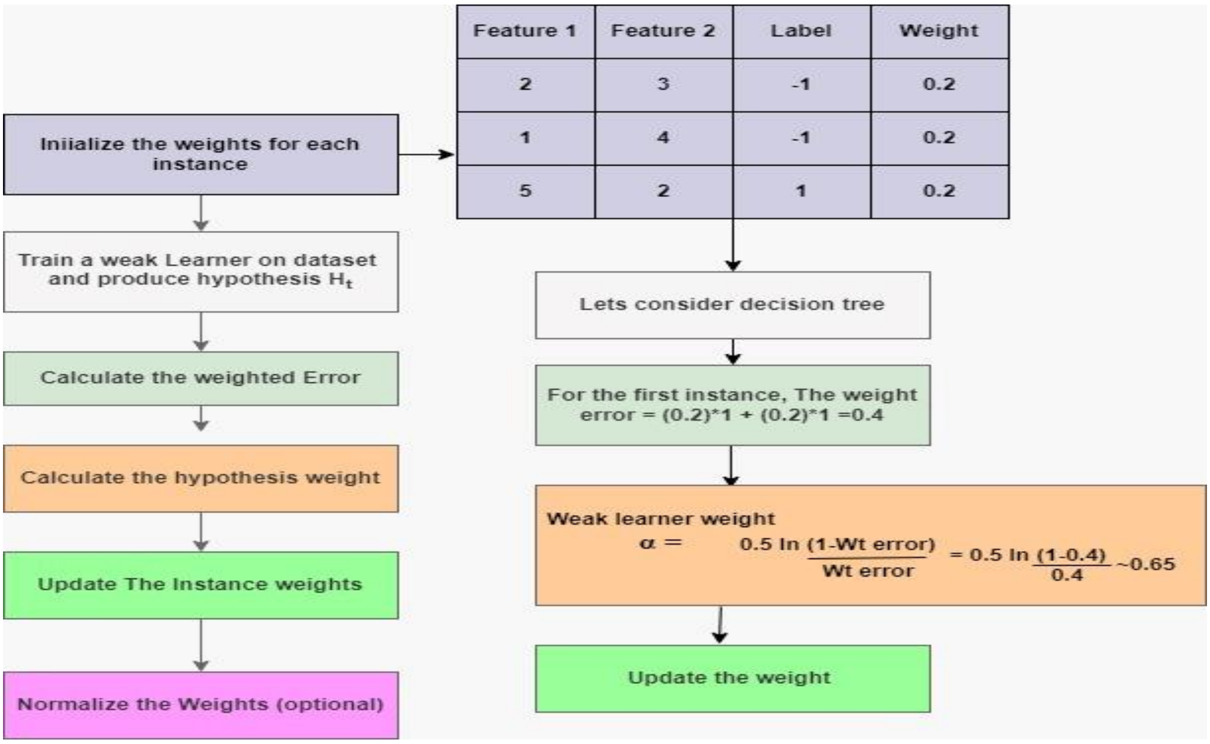
Where,

W is the weight

n is the number of samples

k is the estimators of the weak records

The key difference in the multi-class adaptation is that each weak classifier is trained to distinguish one class from the rest. The final prediction involves combining the weak classifiers for each class and selecting the class with the highest weighted vote. The AdaBoost.MH algorithm can continue for a predefined number of iterations or until a specified accuracy is reached. The intuition is similar to binary AdaBoost, where emphasis is given to misclassified samples to improve the performance of the ensemble on difficult-to-classify instances. The working of ADABOOST is shown in figure 7



Feature 1	Feature 2	Label	Weight
2	3	-1	0.2
1	4	-1	0.2
5	2	1	0.2

Iniialize the weights for each instance
Train a weak Learner on dataset and produce hypothesis H_t
Calculate the weighted Error
Calculate the hypothesis weight
Update The Instance weights
Normalize the Weights (optional)
Lets consider decision tree
For the first instance, The weight error = (0.2)*1 + (0.2)*1 =0.4
Weak learner weight
$\alpha = 0.5 \ln \frac{(1\text{-Wt error})}{\text{Wt error}} = 0.5 \ln \frac{(1\text{-}0.4)}{0.4} \sim 0.65$
Update the weight

Figure 7: Working of ADABOOST for updating the weights of Weak Samples

3.2.4. Soft Ensemble: Each model in the ensemble independently makes predictions on the input data. These predictions are typically class labels for classification tasks. Weighted averaging involves assigning a weight to each model's prediction based on its confidence or performance. Models with higher confidence or better performance may have a greater influence on the final prediction. The final working of proposed model is shown in figure 8

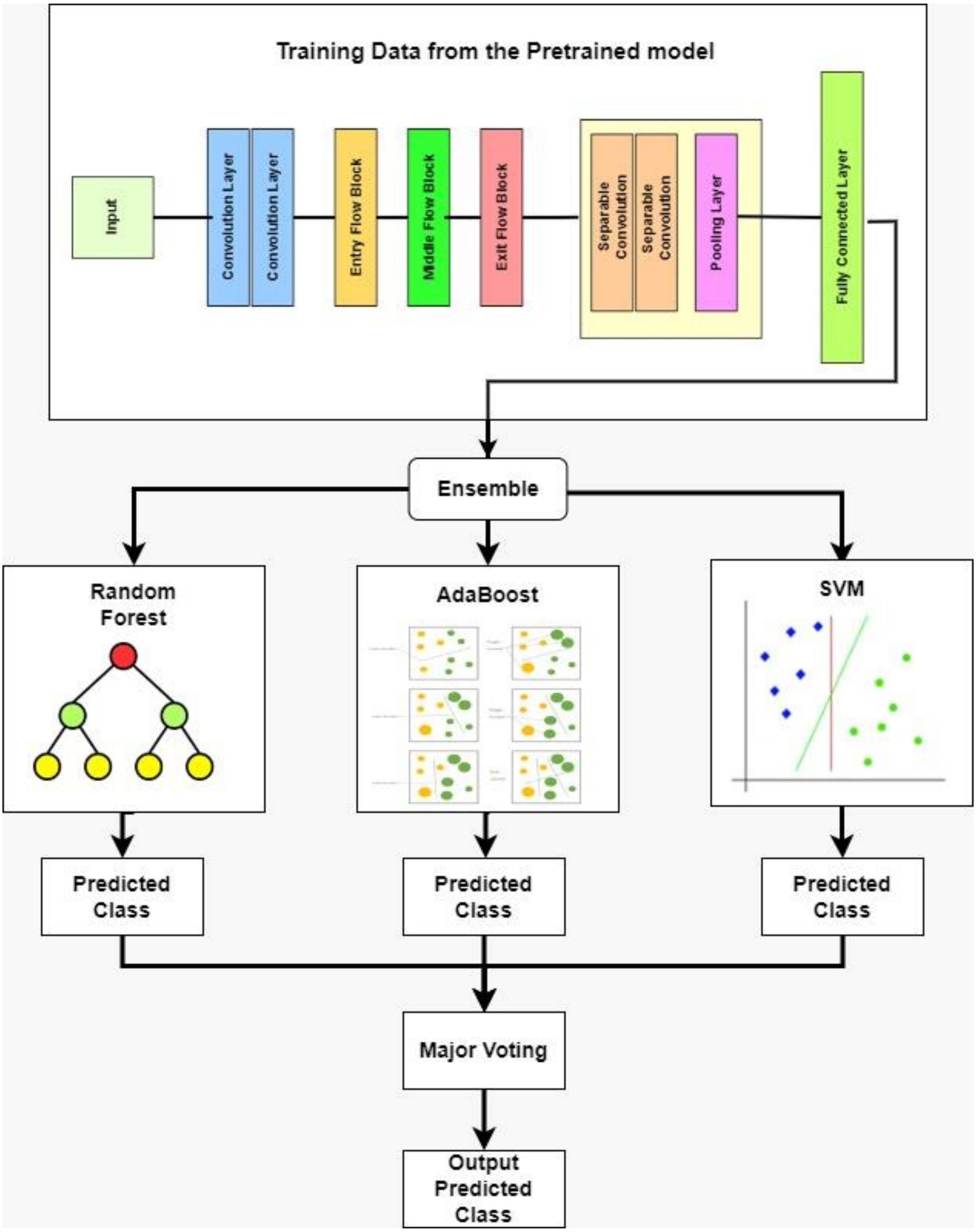
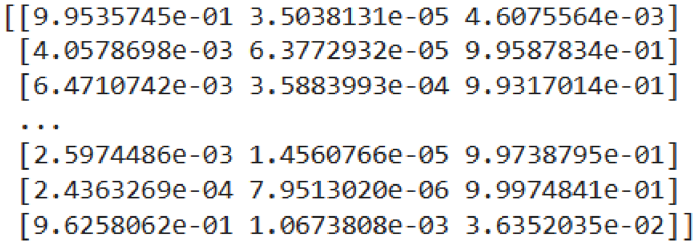


Figure 8: Integration of Xception with Soft Ensemble for Gender Classification

4. Results & Discussion



```
[[9.9535745e-01 3.5038131e-05 4.6075564e-03]
 [4.0578698e-03 6.3772932e-05 9.9587834e-01]
 [6.4710742e-03 3.5883993e-04 9.9317014e-01]
 ...
 [2.5974486e-03 1.4560766e-05 9.9738795e-01]
 [2.4363269e-04 7.9513020e-06 9.9974841e-01]
 [9.6258062e-01 1.0673808e-03 3.6352035e-02]]
```

Figure 9: Selected Features in Xception

In figure 9, it represents the extracted features from each sample using the Xception module. The architecture of Xception allows for hierarchical feature learning. Different layers in the network capture features at different levels of abstraction. Lower layers capture low-level features like edges and textures, while deeper layers capture more complex and abstract features. This hierarchical representation is beneficial for various computer vision tasks.

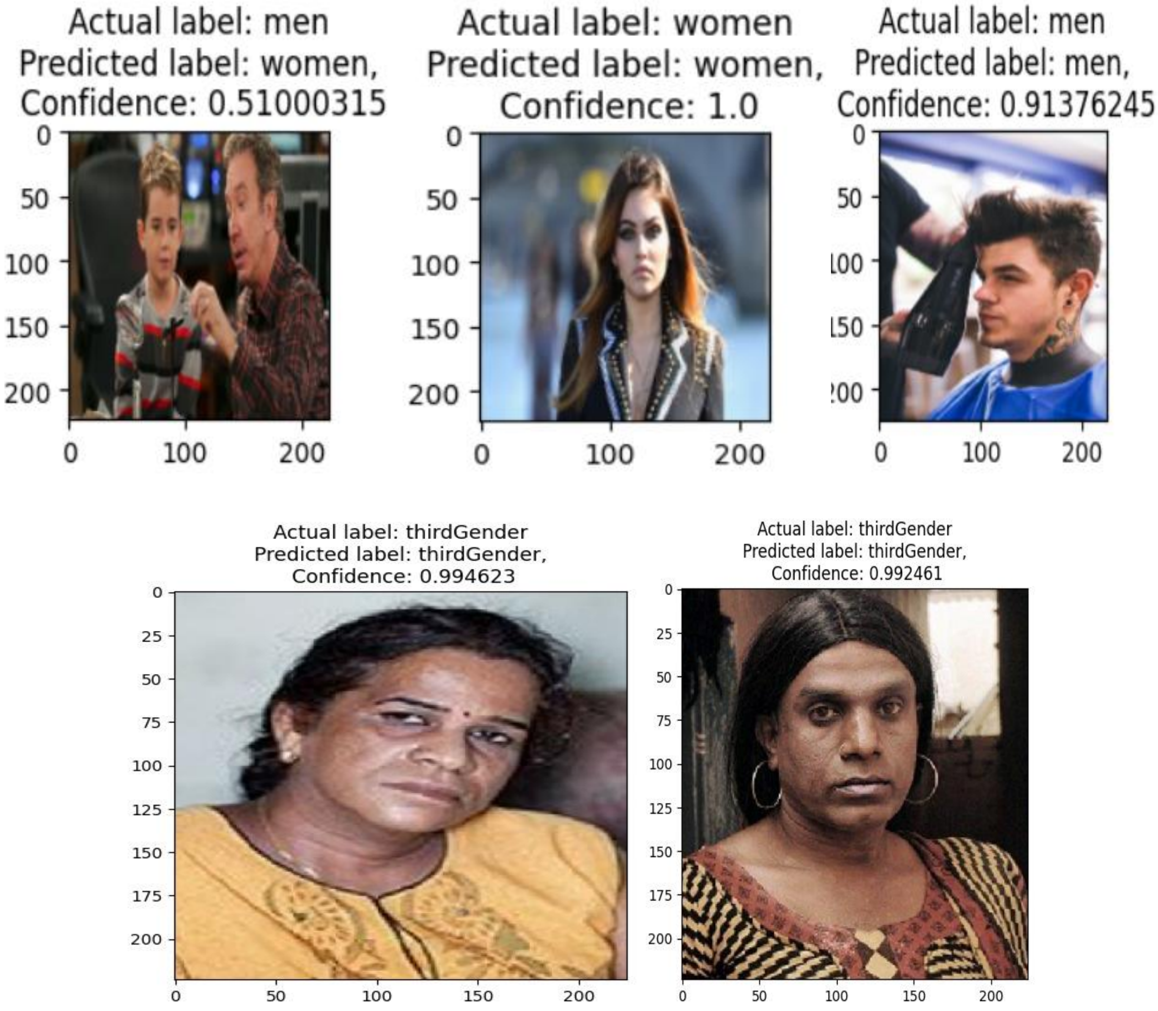


Figure 10: Gender Classification using Soft Xception Module

In Figure 10, for classification it computes the confidence level. In the realm of machine learning, especially within classification tasks, confidence serves as a metric indicating the level of certainty a model possesses regarding its predictions. This metric is closely linked to the probabilities assigned to the predicted classes. In simpler terms, confidence corresponds to the highest probability assigned to any class. If a model predicts probabilities for both Class 1 and Class 0, the confidence is determined by the greater of the two probabilities. This provides insight into the model's certainty in its prediction, with a higher confidence value indicating a stronger level of certainty. For scenarios involving multi-class classification with more than two classes, the formula extends to:

Confidence = max(P(Class_1), P(Class_2), ..., P(Class_N))

Here, N represents the number of classes, and the maximum probability is selected from among all classes. As an illustration, consider a model predicting probabilities for three classes (A, B, C), with the following probability outputs:

P(A) = 0.2
P(B) = 0.6
P(C) = 0.3

In this case, the confidence is computed as max (0.2, 0.6, 0.3) = 0.6, indicating that the model is most confident in predicting Class B. It's essential to recognize that the interpretation of confidence values can vary depending on the context and the specific model employed. For instance, neural networks often employ a softmax function to output probabilities, ensuring their summation to 1 across all classes. In such instances, the class with the highest probability is typically chosen as the predicted class, and its probability can be regarded as the confidence score.

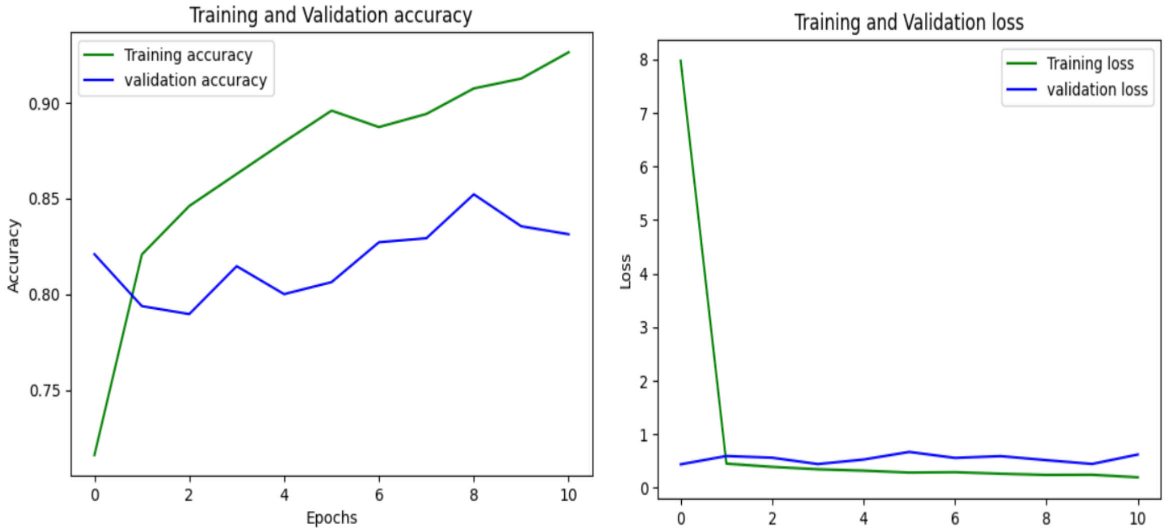


Figure 11: Accuracy and loss with proposed model

Figure 11 represents the accuracy and loss graph per epoch. It compares training and testing data for comparing the performance at every step. Training data is gradually increasing and testing data is decreasing. Both are stable in their performances.

5. Conclusion

In this research, we initiate the image processing pipeline by identifying and extracting the facial regions within a given image, streamlining the subsequent training process. These processed images serve as the training dataset for a gender classifier, aiming to differentiate between females and males. This study has demonstrated the effectiveness of the Xception neural network architecture in the task of gender classification using facial images. The utilization of deep learning techniques, especially harnessing the capabilities of Xception, has played a pivotal role in achieving precise and robust results in gender classification. The high accuracies attained in our study are nearly on par with those of existing state-of-the-art methods. These results open avenues for further exploration, suggesting the potential use of Xception and an ensemble model comprising random forest, Support Vector Machine, and AdaBoost-generated samples as training data. This approach aims to minimize reliance on actual facial images, thereby mitigating potential privacy risks. The experiments not only present a proficient gender classifier but also articulate discernible and easily understandable discriminant rules. For future endeavors, it would be worthwhile to delve into additional fine-tuning strategies, scrutinize the interpretability of the model's decisions, and address any inherent biases in the training data. This study contributes significantly to advancing our comprehension of deep learning applications in gender classification, establishing a foundation for ongoing research in this vital and interdisciplinary field. Furthermore, incorporating multiple perspectives on the gender of the subject in the image could potentially enhance the accuracy of the ground truth label assigned to each image.

6. References

1. Raman, V., ELKarazle, K., & Then, P. (2023). Artificially Generated Facial Images for Gender Classification Using Deep Learning. Computer Systems Science and Engineering, 44(2), 1341–1355. https://doi.org/10.32604/csse.2023.026674
2. Almomani, A., Alweshah, M., Alomoush, W., Alauthman, M., Jabai, A., Abbass, A., Hamad, G., Abdalla, M., & B. Gupta, B. (2023). Age and Gender Classification Using Backpropagation and Bagging algorithms. Computers, Materials & Continua, 74(2), 3045–3062. https://doi.org/10.32604/cmc.2023.030567
3. Tabassum, H., Iqbal, M. M., Mahmood, Z., Parveen, M., & Ullah, I. (2023). Gender classification from anthropometric measurement by boosting decision tree: A novel machine learning approach. Journal of the National Medical Association, 115(3), 273–282. https://doi.org/10.1016/j.jnma.2022.12.005
4. Mustapha, M. F., Mohamad, N. M., & Ab Hamid, S. H. (2023). TWOFOLD FACE DETECTION APPROACH IN GENDER CLASSIFICATION USING DEEP LEARNING. International Journal of Software Engineering and Computer Systems, 9(1), 59–67. https://doi.org/10.15282/ijsecs.9.1.2023.6.0110
5. Abbas, F., Gattal, A., & Menassel, R. (2023). Local binary pattern and its derivatives to handwriting-based gender classification. Bulletin of Electrical Engineering and Informatics, 12(6), 3571–3583. https://doi.org/10.11591/eei.v12i6.5488
6. Kota Sandeep B-Tech Student (2023). AGE AND GENDER CLASSIFICATION USING CNN.
7. Rohani, M., Farsi, H., & Mohamadzadeh, S. (2023). Deep Multi-task Convolutional Neural Networks for Efficient Classification of Face Attributes. International Journal of Engineering, 36(11), 2102–2111. https://doi.org/10.5829/ije.2023.36.11b.14
8. Wang, Z., Meng, Z., Saho, K. et al. Deep learning-based elderly gender classification using Doppler radar. Pers Ubiquit Comput 26, 1067–1079 (2022). https://doi.org/10.1007/s00779-020-01490-4
9. Adhinata, F. D., & Junaidi, A. (2022). Gender Classification on Video Using FaceNet Algorithm and Supervised Machine Learning. In International Journal of Computing and Digital Systems (Vol. 11, Issue 1, pp. 199–208). Deanship of Scientific Research. https://doi.org/10.12785/ijcds/110116
10. Costantini G, Parada-Cabaleiro E, Casali D, Cesarini V. The Emotion Probe: On the Universality of Cross-Linguistic and Cross-Gender Speech Emotion Recognition via Machine Learning. Sensors. 2022; 22(7):2461. https://doi.org/10.3390/s22072461

11. Chaari, N., Gharsallaoui, M. A., Akdağ, H. C., & Rekik, I. (2022). Multigraph classification using learnable integration network with application to gender fingerprinting. In Neural Networks (Vol. 151, pp. 250–263). Elsevier BV. https://doi.org/10.1016/j.neunet.2022.03.035
12. Yücesoy, E. Speaker age and gender classification using GMM supervector and NAP channel compensation method. J Ambient Intell Human Comput 13, 3633–3642 (2022). https://doi.org/10.1007/s12652-020-02045-4
13. Shin, J., Maniruzzaman, Md., Uchida, Y., Hasan, Md. A. M., Megumi, A., Suzuki, A., & Yasumura, A. (2022). Important Features Selection and Classification of Adult and Child from Handwriting Using Machine Learning Methods. In Applied Sciences (Vol. 12, Issue 10, p. 5256). MDPI AG. https://doi.org/10.3390/app12105256
14. Gupta, S.K., Nain, N. Review: Single attribute and multi attribute facial gender and age estimation. Multimed Tools Appl 82, 1289–1311 (2023). https://doi.org/10.1007/s11042-022-12678-6
15. Nayak, J. S., & Indiramma, M. (2022). An approach to enhance age invariant face recognition performance based on gender classification. In Journal of King Saud University - Computer and Information Sciences (Vol. 34, Issue 8, pp. 5183–5191). Elsevier BV. https://doi.org/10.1016/j.jksuci.2021.01.005