A

Seminar-I Report

On

GENDER IDENTIFICATION BY USING FACIAL IMAGE

Prepared by

Mr. Keval Bavadiya

P21CO008

Supervised by

Shri. R. P. Gohil, SVNIT, Surat

2022– 2023

Department of Computer Science & Engineering

Sardar Vallabhbhai National Institute of Technology,

Surat-395007 (Gujarat), India

Sardar Vallabhbhai National Institute of Technology, Surat-395007, Gujarat, India

CERTIFICATE

This is to certify that report on Seminar-I entitled ***"Gender Identification by using facial Image"*** *is prepared and presented by* ***Mr.Keval Bavadiya*** *bearing Roll No:* ***P21CO008****,2nd Year of M.Tech (Computer Engineering) and his work is satisfactory.*

SUPERVISOR	JURY	HEAD OF DEPT.
Shri. R. P. Gohil		Dr. Rupa G. Mehta

Acknowledgements

At the outset, I thank the God Almighty for the grace, strength and hope to make my endeavor a success. I express my sincere gratitude to my guide Shri. R. P. Gohil, Associate Professor, Department of Computer Science and Engineering, SVNIT, Surat for his exemplary guidance, monitoring and constant encouragement without which the successful completion of this report would not have been possible.

I am also highly grateful to the staff of Computer Department for giving me a lot of their valuable time, as well as guidance and suggestions whenever I needed them.I would like to thank to my friends, who stood by me, whenever I needed their assistance and also gave me the strengths to carry on.

I would like to express my gratitude towards my parents and family for their kind co-operation and encouragement which help me completion of my report.

Keval Bavadiya

P21CO008

Abstract

Human gender detection which is part of face recognition has received extensive attention because of its different kind of applications especially in the field of security, marketing and intelligent user interfaces. Previous research work on gender detection accomplished based on different type of static body feature like eyebrow, face, etc. support vector machine (SVM) and Convolutions neural network (CNN) have been successfully used in this domain.Because in any other algorithm first feature extraction is required which is time consuming as well as hardware.so, with few paper we explored that SVM is also not meet expectation as CNN because its also trained with different feature extraction namely HOG, LBP and Raw feature. we explored that by doing learn and classification method and with Convolution Neural Network technique, a satisfied growth in performance can be achieved on gender classification.The whole method evaluated on the data-set collected from Kaggle.

Keywords: Face detection, HOG, LBP, Convolutional Neural Network, ReLu, softmax, kaggle dataset

Table of Contents

List of Figures

List of Tables

List of Abbreviation

IoT Internet of Things

DWT Discrete Wavelet Transform

DCT Discrete Cosine Transform

SVM Support Vector Machine

CNN Convolutional Neural Network

D-CNN Deep Convolutional Neural Network

F-KNN Fuzzy K-Nearest Neighbour

LBP Local Binary pattern

HOG Histogram of oriented Gradient

KNN K-Nearest Neighbour

Chapter 1

Introduction

Gender detection plays significant role in modern technology so it's an active research area which has attracted a great deal of attention. It has applications in many domain security surveillance(Monitoring Application), targeted advertisement(marketing), human computer interface and demographics, criminal identification, video game etc. It has occupied a huge scope in the field of facial recognition.[3, 4, 6]. The main purpose of the gender identification is to differentiate male and female based on features[4]. When identify gender there are some distinct feature exist between male and female which is used in classification of gender[7].

Gender classification is binary classification task which become one of the most important task due to lots of applications[8]. In recent year many paper published and main focus on the face of the human with some different-different facial analysis[9]. Generally all pattern recognition problems when tacked with a supervised learning technique, could be dividing into following steps : (1) object detection and pre-processing (2) feature extraction (3) classification method[1].The feature extraction is classified into many categories namely, geometric, appearance based ,[4] hand crafted feature, deep-learned feature, fusion feature[3], DCT , DWT[4], etc.

Fig. 1.1 elaborate general steps of solution of the problem in computerized method. Most of the algorithm use the use Viola-Jones algorithm and SVM as classification[3].The main difference lies in the type of feature extraction based on that feature extraction method classification algorithm result may vary.The feature extraction is classified into two categories, geometric based and appearance based[4]. SVM give satisfactory results with HOG, LBP ,and raw pixel feature extraction method[3] where as KNN best with DCT and DWT[10]. However,in the field of image classification the Convolution Neural Network (CNN) has been proved best algorithm among all machine learning algorithm[4]. Main advantage of CNN is it's extract fea-

ture automatically, another difficult task with (SVM,Bayesian) any algorithm has to find such classifier that aggregate the finest among the chosen feature extractor so excellent result can be obtained[5].

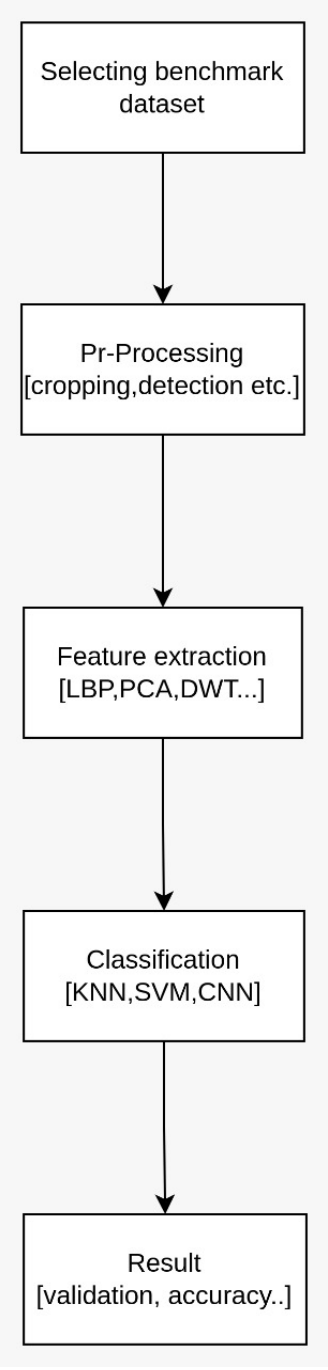

Figure 1.1: The general system of gender classification

Chapter 2
Literature Review and Theoretical Background

2.1 Machine Learning Model

2.1.1 K-Nearest Neighbors Algorithm (KNN)

A machine learning technique known as the KNN classifier is regularly utilised for regression and classification forecasting tasks. However, it is primarily employed in enterprise for categorisation and forecasting problems. The KNN learning method is non-parametric since it makes no assumption on statistical data. It is also known as the lazy learning algorithm because it does not have a dedicated training phase and uses all of the data during classification.

Thus, the training phase of this algorithm is faster, whereas the testing phase is slower and more expensive. The testing step is more expensive because it takes longer to execute and requires more memory. It needs more memory to store training data samples since it takes longer to scan all data samples [11].

Steps:

1. To determine the number of clusters, choose K.
2. Choose K centroids or random points.
3. Assign each data point to its nearest centroid, which will create the K clusters that have been predetermined.
4. Determine the variance and relocate each cluster's centroid.
5. Re-assign each data point to the new centroid of each cluster by repeating the third step.
6. Proceed to step 4 if there is any reassignment; otherwise, proceed to step 6.

7. The model is complete.

Advantage:

1. The training phase of KNN classification algorithm is faster than other classification algorithms.

2. For generalization, model training is not needed.

3. KNN is instance-based and simple learning algorithm [12].

4. KNN can be useful in non-linear data.

5. In the regression problem it can be useful where output value is calculated by the average value of K closest neighbors value.

Disadvantage:

1. The testing step of KNN classification is slow and costly in time and memory.

2. Calculating distance in each dimension is hard for high dimensional data hence KNN does not perform well.

3. Huge memory is taken for storing the entire training dataset for prediction.

4. KNN takes the Euclidean Distance between two data points/samples for finding nearest neighbors [13] thus scaling of data is necessary.

2.1.2 Support Vector Machine (SVM)

SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, it is mostly used for Classification problems in Machine Learning.

The objective of the SVM technique is to determine the optimal decision boundary or line for categorising n-dimensional area into groups so that new data sets can be added in the future and assigned to the appropriate order right away. A hyperplane is a boundary that represents the best option. The extreme points and vectors that aid in the construction of the hyperplane are chosen using the SVM. Due to the fact that support vectors depict severe circumstances, the method is known as a support vector machine.

Take a look at the figure illustration below:

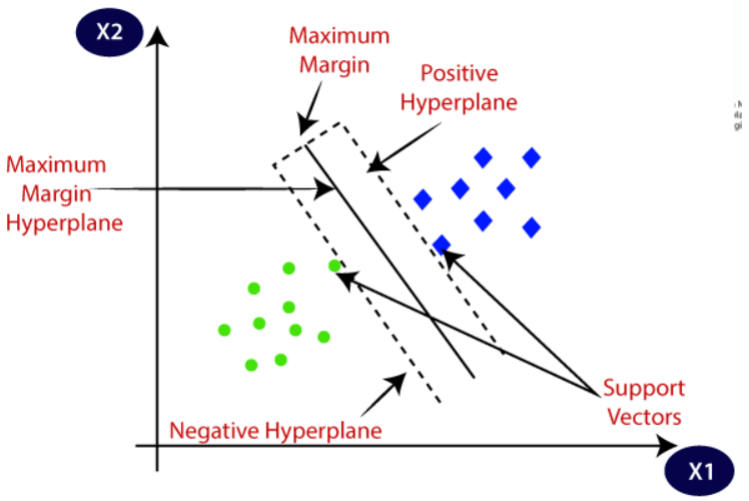

Figure 2.1: Support Vector Machine[1]

Steps:

1. Activate the essential libraries.
2. Import dataset and extract the X variables and Y separately.
3. Separate the data into train and test groups.
4. Train the SVM algorithm.
5. Make some predictions.
6. Evaluate the results of the algorithm.

Advantage:

1. It functions very well when there is a distinct range of divergence.
2. It works well when the quantity of dimensions exceeds the quantity of samples.
3. It is also memory effective, using only a fraction of training data in the classification model.

Disadvantage:

1. Because the needed training time is longer when we have a huge data set, it really doesn't achieve well.

2.1.3 Neural Network (NN)

In the area of artificial intelligence, an artificial neural network is one that makes an effort to resemble the network of neurons that composes the human brain so that computers would have the option to comprehend concepts and make decisions in a method that is similar to that of a person. Computers are programmed to function exactly like a network of interconnected brain cells to create an artificial neural network.

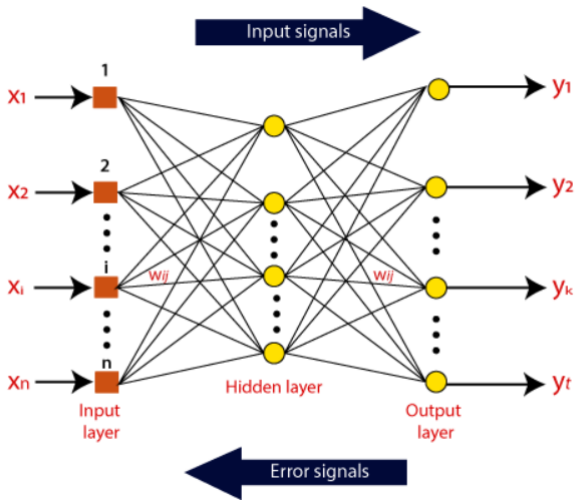

Figure 2.2: Neural Network [2]

The input signal for the artificial neural network comes from an external source as a pattern and an image as a vector shown in Fig. 2.2. Then, using the notation x, these inputs are mathematically assigned for each nth input (n). Following that, each input is multiplied by the associated weights. In the artificial neural network, these weights often indicate how well neurons are connected to one another. Inside the computer unit, a summary of each weighted input is created. Here, a certain maximum value is benchmarked to maintain the response within the bounds of the intended value, and the sum of the weighted inputs is delivered through the activation function. The set of transfer functions utilised to produce the desired output is referred to as the activation function.

Advantage:

1. Parallel processing capability.
2. Storing data on the entire network.
3. Capability to work with incomplete knowledge.

4. Having fault tolerance.

Disadvantage:

1. Confirmation of appropriate network architecture

The construction of artificial neural networks is not determined by any specific rules.

2. Dependency on hardware

2.2 Literature Survey

In [10] salma Mohamed, Nahla Nour, Serestina Viriri proposed a method for grander identification using Global Feature, for that the collect the image from ESSEX and FERET. The image has size of 256×384 pixels. ESSEX contains 153 different object and each objects contains 20 image. where as FERET contains 900 image among them 485 images for training and 415 for testing. In this proposed method first apply face detection via Viola and Jones method.Main role of this research is in feature extraction method. In feature extraction they proposed technique using DWT and DCT both focus on global feature extraction.The reason of feature extraction is to improve classification performance of the system.In this research work ,DWT and DCT are combined t form a hybrid feature extraction technique. DCT convert image into its elementFFGary frequency component.They used two-dimensional DCT, in this research. It works on 1-dimensional DCT as basic operation,

The DCT equation for 1D:

$$D(u) = C(u) \sum_{x=0}^{N-1} P(x) \cos \frac{(2x+1)\pi u}{2N} \quad \text{.....(1)}$$

The DCT equation for 2-D:

$$D(u,v) = C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} P(x,y) \cos \frac{(2x+1)\pi u}{2N} \cos \frac{(2y+1)\pi v}{2N} \quad \text{.....(2)}$$

Discrete Wavelet Transform decomposes a signal into set of basic functions called wavelets, decomposition is defined as resolution of signals. At each decomposition level low-pass filter

and high-pass filter are applied to each row/column of image to decompose image into three high frequency sub-bands.

Mathematically expression of DWT,

$$DWT_{X(n)} = \begin{cases} d_{j,k} = \sum X(n)\, h_j^*(n - 2^j k) \\ a_{j,k} = \sum X(n)\, g_j^*(n - 2^j k) \end{cases} \quad \text{.....(3)}$$

In this they used K-Nearest Neighbour ,Fuzzy-KNN and Support vector machine (SVM) for classification. KNN and Fuzzy-KNN both are similar but in K-NN every data point belongs to one class, which is majority class .Whereas in Fuzzy K-NN belongs to more than one class and multiple membership function associated to these classes.SVM goal is minimizing an upper bound of the generalization error through maximizing the margin between separating hyper planes.As part of result, the KNN when using ESSEX the high accuracy achieved with hybrid DCT-DWT but when using FERET achieved higher accuracy with DCT. So Hybrid DCT-DWT achieved high accuracy when applied F-KNN and SVM.In terms of mathematics KNN achieved 88% with Hybrid DCT-DWT where as F-KNN achieved 93% and SVM achieved 95%.

In [3] George Azzopardi, Antonio Greco , and Mario Vento proposed method for gender recognition using a fusion of SVM Classifiers.In this research first the proposed, first they proposed face detection using Viola-Jones algorithm[14]. Then resize that face by cropping and convert image into 128×128 pixels. Then they transform the image into a element feature vector and convert all dimensional in same range by dividing each element by 255 so that all dimensional have range in between 0 to 1.They apply LBP descriptor to entire image by comparing intensity value of each pixel with 3×3 neighbourhood method. Generate a 256-element L2-normalized histogram of tiles using spatial tilling 3×3. At the end merge that nine histogram and form 2304 (256×9) vector for each image.This histogram is used with SVM classifier.

Another descriptor HOG they used. For that first they compute angle and gradient of every pixel by considering the response of first-ordered partial derivative of 2D Gaussian function with $\sigma = 1$. Now 32×32 pixels that overlap by 50% of each and for each tile they compute the L2-Normalized weighted histogram of 9 bins using 2×2 tiles.So HOG result has 1764-element

vector(7×7 blocks , 4 tiles ,9 bins) similar to LBP based descriptor HOG is also applied to SVM input for classification.

The result of is pair of probabilities that given image is male or female. Final decision is taken based on the they sum three male probabilities and three male probabilities,if total male probability is more that female probability then labeled as male otherwise female.

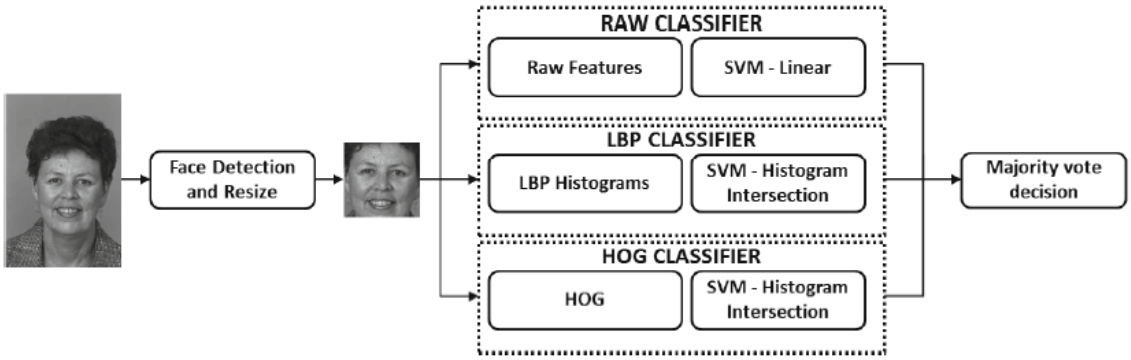

Figure 2.3: Architecture of proposed method by them[3].

The dataset they used used is subset of FERET. This contains 473 male and 473 female image among them 50% used for training and the remaining 50% for the test set.As per there result any combination of feature and classifiers achieved minimum 85% accuracy. Using leaner SVM highest accuracy achieved is 90.9%,while best accuracy achieved using histogram intersection of both LBP and HOG is 92.6%.

In [6] Alhanoof Althnian ,Norah Aloboud,Norah Alkarashi proposed face recognition in wild.In this research they perform comprehensive study of Support vector Machine (SVM) and Convolution Neural Network(CNN) along with that they include some feature extraction method like hand- crafted ,deep-learned etc.They applied seven feature extraction method, which is majorly divide into three categories : Hand-crafted feature ,deep-learned feature and fused feature.

Hand-Crafted Feature: Hand-crafted feature categorised into pixel based features, global features, and appearance-based feature.LBP is one of the most used hand-crafted feature extraction method in Gender classification. In LBP 3×3 binary digit assign and for each pixel and compare with central(threshold value) and one digit signed if it is greater than or equal to

central value else zero assigned.The binary value for central pixel is further computed by concatenating eight digit in clock wise direction. Another method is HOG(Histogram of Oriented Gradient), is an appearance based descriptor that extract the gradient and angle. using gradient calculate magnitude of each pixel. other than same as all general method which is available in [3].Third feature extraction method is known as Principal Component Analysis(PCA)is applied to raw pixel value.they perform eigen decomposition for eigen value and eigen matrix.The principle components of data set has highest eigen value.The User may decide that need to keep all data set or subset of principle components.

Deep-learned Feature: in this input image pass through a series of hidden convolution layers,which use ReLu(Rectified Linear unit) as activation function . some layer performed non-overlapping using max-pool layer.

Fused Feature: in this feature extraction both deep learned and hand crafted feature are used according to there characteristics.

For classification they used SVM and CNN with 2000 epochs,128 batch size and binary cross entropy loss function.As result with Adience data set Hand crafted Feature and SVM classifier used and respectively there result for HOG, LBP,PCA are 65.5%,62.5%.,60.9%. but with deep learned feature result is 83.3% and using CNN classifier result improve more that SVM which is 89.2%. So conclusion of this research is CNN is better that SVM, no matter which feature extraction method is applied.

In [4] Tahmina Akter Sumi, Mohmmad Shahadat Hossian, Raihan UI Islam proposed technique for gender identification using Convolution Neural Network.In this rather than basic steps they also did pre-processing on Kaggle data set. Generally here they remove low frequency background, reflection of light so new resized image is 96×96 dimension, after that they convert the image into an array of pixel value. Each value is then convert into float and after that those value is divide by 255 so that all the value between 0 to 1.

For convolution Neural Network feature expression performed by the convolution and pooling layer. So for that convolution layer contain 32 filters with ReLU activation function.The POOL layer uses 3×3 pool size to reduce dimension from 96×96 to 32×32. ReLU applied

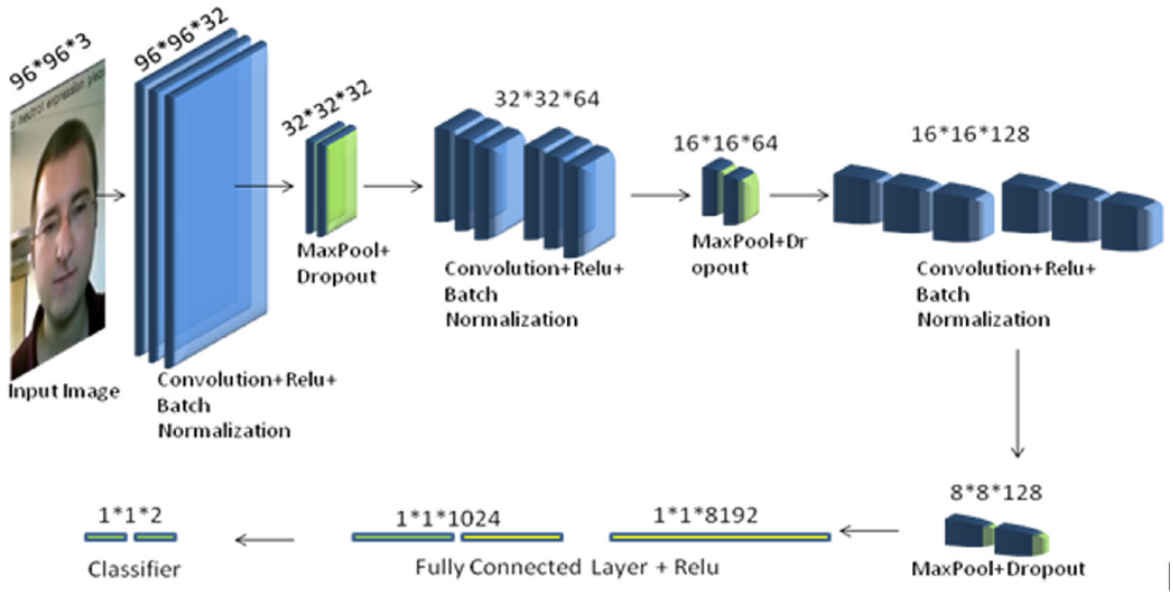

Figure 2.4: A full schematic diagram of network arch. proposed by[4].

twice applying another POOL layer.This ReLU layer and multiple convolutional layer allow to learn a wealthy set of feature. Again ReLU and convolution is applied twice before applying another POOL layer.After that sigmoid classifier is used for the classification purpose.In this 50% of the node is used for testing. System is implemented by using Matplotlib, keras, numpy libraries.

In terms of result,for Kaggle data set Adam , Adamax , RMSprop and Adagrad optimizer is above 90%. but when their performance vary for different activation function.It means Sigmoid perform well after that with SoftMax and Relu gave very less accuracy with respect to this two ,But it's may vary with dataset.

In [5] Amit Dhomne ,Ranjit Kumar, Vijay Bhan proposed method for gender recognition using deep learning.The main advantage of the Deep Convolutional Neural network is it automatically extract feature from an image and give output .So for deep learning based model we don't required to use feature extractor like Local Binary pattern, Histogram oriented Gradient(HOG) , Support vector Machine (SVM) , eigenvector and many more. The D-CNN perform both the work of feature extraction and feature classification so provide result with high accuracy and minimum cost.The depth of the configurations increase from left to right in given Fig 2.4. Because of more number of layer are added.

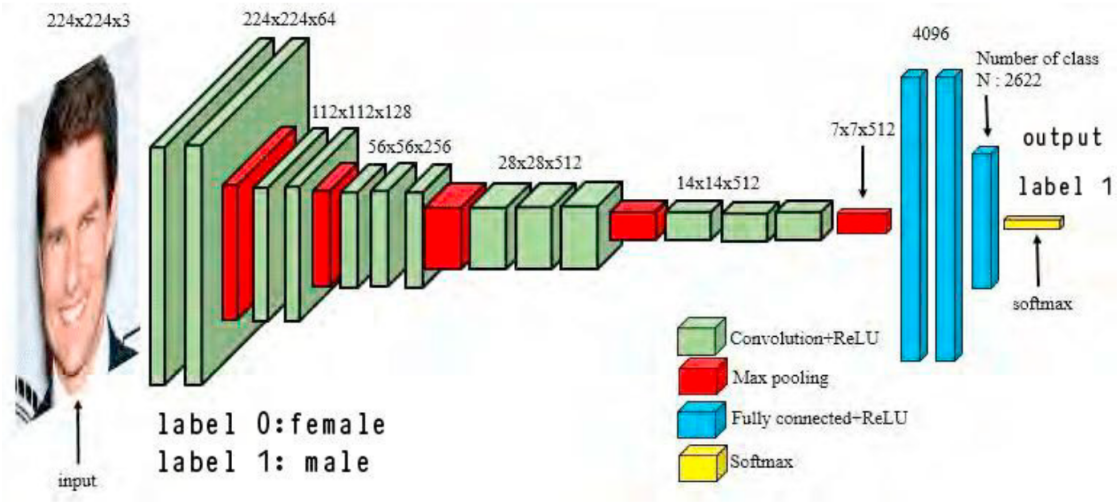

Figure 2.5: VGGNet CNN proposed Arch by[5].

A pattern on an image can be extracted by the following expression:

$$p_i^{(t)}(q,p) = F\left(\sum_{i=0}^{B}\sum_{u=0}^{R_q^{(t)}}\sum_{v=0}^{R_p^{(t)}} p_i^{(t-e)}(s_q^{(t)}q+u, s_p^{(t)}p+v)m_{ef}^{(t)}(u,v) + \theta_j^{(t)}\right) \quad \text{....(4)}$$

In equation (4), $Pi^{(t-e)}$ and $Pj^{(t)}$ are input are the input and output pattern respectively.F() is an activation function.$Sq^{(t)}$ horizontal convolution step size.

In this research they proposed, image scaled from 256×256 to 227×227 after this fed into network which consist three FC layer, first take input from third convolutional layer and exhibit neuron equals to 512 same as for second layer and third layer which maps to the final classes for gender classification. at the end result is forward to a soft-max layer function then assigns a final probabilistic label in gender detection. They use a subset of the celebrity face data-set model, which contain 200 images.Among them training-set consist of 160 imagesand test consist of images.As result it's 95% accuracy with only small size data set with huge dataset around 91%.

Gender detection also done by using voice or speech in 19's. Hidden Markov model proved best in this era.parrise and carry [9] designed model which combined pitch and HMM for gender identification in this for male and female separately HMM speech recognise model is designed.This model provide accuracy of 97.3%. This experiment carried on 5 sec speech from OGI database.

In [15] shivamchaudhary and Devendea kumar sharma proposed model in which gender identification is done by using Voice signal charactristics.This research era is also same as facial image detection for that, first feature extraction from voice pitch, energy, MFCC etc. after that concatenation of that feature extraction result and apply classifier like SVM,CNN etc. at the end probabilistic decision is done for testing data.For that they used speech database is collected from TIMIT Speech cropus, which contains 140 speech files of male and 140 ;speech files from female. They divide this dataset into training and testing in 8:2. As part of result they achieved 96.45% accuracy.

Chapter 3

Methodology

We implement CNN for gender identification using Facial image. Purpose of this Convolutional Neural network model is no need to extract feature because CNN automatically extract the feature when it's passes through layer due to max poling layer result is better. The data set used here is Kaggele UTK-cropped dataset, which contains approximately 23,708 images so data set split into training and testing parts which contains 18967 and 4742 images respectively. Model contain five convolutional layer followed by relu activation and Maxpool layer. As part of result accuracy (in terms of testing) is 90.13%.

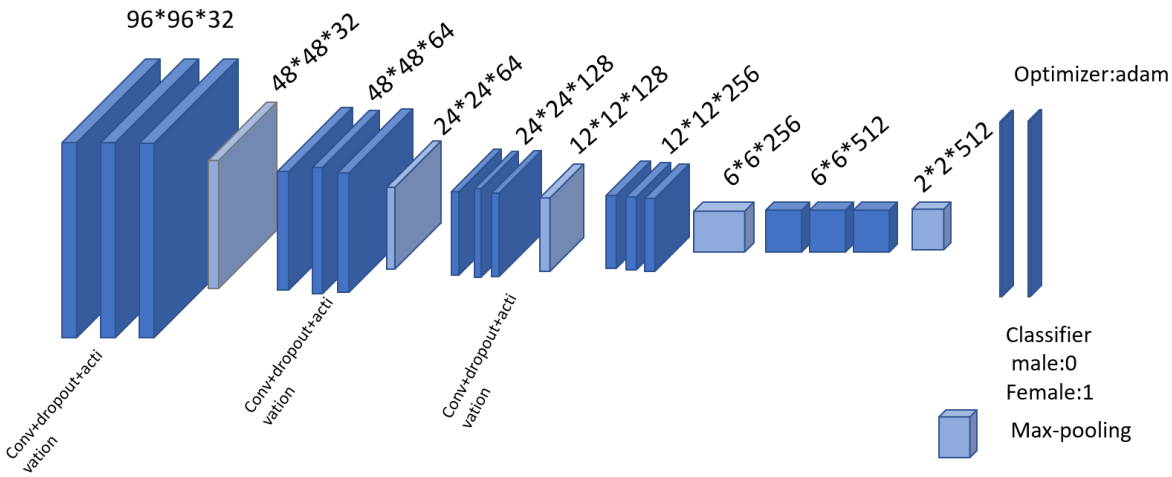

Figure 3.1: Proposed model

Figure 3.3: Result of proposed Model

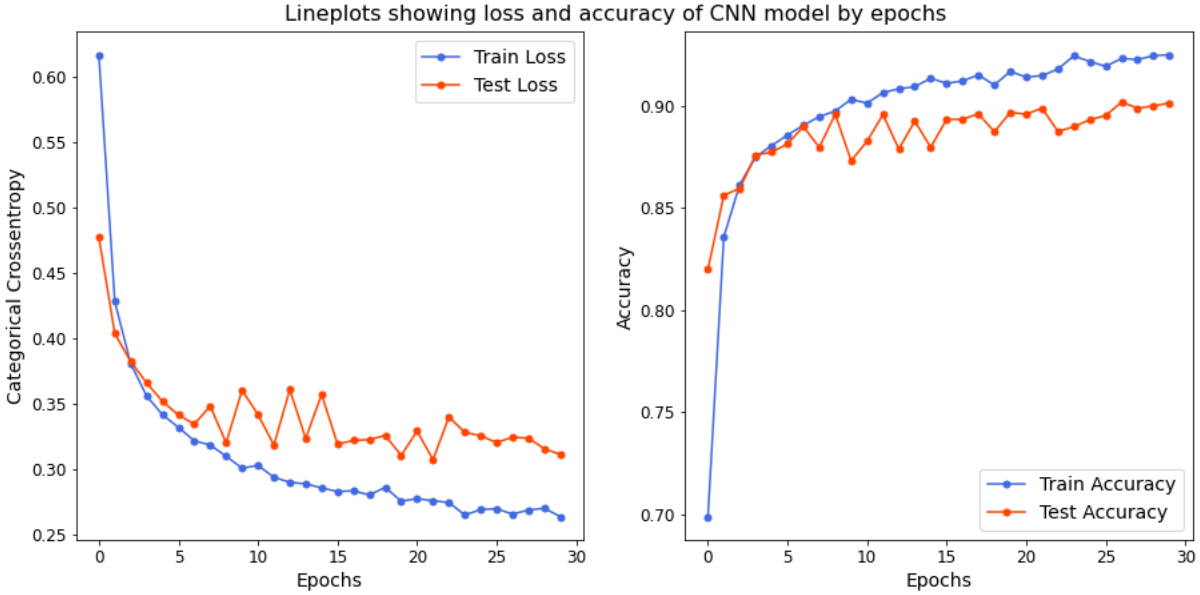

Figure 3.2: Loss and accuracy of CNN model by epochs.

```
loss,acc = model.evaluate(x_test,  y_test, verbose=2)
print("Success rate: {:5.2f}%".format(100*acc))
```

```
149/149 - 1s - loss: 0.3110 - accuracy: 0.9013 - 972ms/epoch - 7ms/step
Success rate: 90.13%
```

Figure 3.4: Model Evaluation on training and testing data.

Chapter 4

Result and comparison

In this we explored many technique is used for gender identification using facial image among them some method gave satisfactory result which are mentioned below,

Ref.	Feature Descriptor	classifier	Dataset	Result
[4]	CNN	CNN	Nottigham Scan(100 images)	80%
[4]	CNN	CNN	CELEBA (3500 images)	96.03%
[4]	CNN(Relu)	CNN	CELEBA (3500 images)	28.35%
[6]	Hand-crafted(HOG)	SVM	Adience(26000)	65.5%
[6]	Fusion(HOG-DL)	SVM	Adience	84.1%
[6]	Fusion(HOG-DL)	CNN	Adience	81.7%
[10]	DCT	KNN	FERET(900 images)	94%
[10]	DCT	SVM	FERET(900 images)	88%
Prop Method.	CNN	CNN	Kaggle(utk)(23k images)	90.13%

Table 4.1: Related work with their reported result

Bibliography

[1] "Classification algorithms - svm," https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm, [Online; accessed 12-June-2022].

[2] "Classification algorithms-logistic regression," https://www.javatpoint.com/artificial-neural-network, [Online; accessed 12-June-2022].

[3] A. G. George Azzopardi and M. Vento, "Gender recognition from face images using a fusion of svm classifiers," 2016.

[4] R. U. I. Tahmina Akter Sumi, Mohammad Shahadat Hossain and K. Ander, "Human gender detection from facial images using convolution neural network," 2021.

[5] V. B. Amit Dhomne, Ranjit Kumar, "Gender recognition through face using deep learning," 2018.

[6] A. Althnian, N. Aloboud, N. Alkharashi, F. Alduwaish, M. Alrshoud, and H. Kurdi, "Face gender recognition in the wild: An extensive performance comparison of deep-learned, hand-crafted, and fused features with deep and traditional models," *Applied Sciences*, vol. 11, no. 1, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/1/89

[7] S. V. Salma M. Osman Mohammed, "Gender identification from facial images: Survey of the state-of-the-art," 2017.

[8] X. Chu, A. Zheng, X. Zhang, and J. Sun, "Detection in crowded scenes: One proposal, multiple predictions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 214–12 223.

[9] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," *Sensors*, vol. 20, no. 2, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/2/342

[10] S. V. Salma Mohamed, Nahla Nour, "Gender identification from facial images using global features," 2018.

[11] "Classification algorithms - knn," https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.html, [Online; accessed 12-June-2022].

[12] S. Robinson, "Classification algorithms - knn," https://stackabuse.com/k-nearest-neighbors-algorithm-in-python-and-scikit-learn/, [Online; accessed 12-June-2022].

[13] A. Navlani, "k-nearest-neighbor classification using scikit-learn," https://www.datacamp.com/community/tutorials/ k-nearest-neighbor-classificationscikit-learn, [Online; accessed 12-June-2022].

[14] M. J. J. PAUL VIOLA, "Robust real-time face detection," 2003.

[15] D. K. S. ShivamChaudhary, "Gender identification based on voice signal characteristics," 2018.