Enhancing the accuracy of fingerprint recognition

A MAJOR PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
Signal processing and digital design

Submitted by:
ABHISHEK

2K18/SPD/01

Under the supervision of

AJAI KUMAR GAUTAM (Asst. Prof)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

SEPTEMBER 2020

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I **ABHISHEK, 2K18/SPD/01** student of M.TECH (Signal processing and digital design), hereby declare that the major project report titled "**ENHANCING THE ACCURACY OF FINGERPRINT RECOGNITION SYSYTEM**" which is submitted by me to the Department of Electronics and Communication, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi **ABHISHEK**
Date: 30//09/2020

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the major project report titled "**ENHANCING THE ACCURACY OF FINGERPRINT RECOGINITION SYSTEM**" which is submitted by **ABHISHEK**, **2K18/SPD/01** of Electronics and Communication Department, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi
Date: 30/09/2020

AJAI KUMAR GAUTAM
SUPERVISOR
(Asst. Professor)
Department of Electronics and Communication
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)

ACKNOWLEDGEMENT

A successful project can never be prepared by the efforts of the person to whom the project is assigned, but it also demands the help and guardianship of people who helped in completion of the project.

I would like to thank all those people who have helped me in this research and inspired me during my study.

With profound sense of gratitude, I thank Mr. AJAI KUMAR GAUTAM., my Research Supervisor, for his encouragement, support, patience and his guidance in this research work.

I take immense delight in extending my acknowledgement to my family and friends who have helped me throughout this research work.

ABHSIHEK

ABSTRACT

In the era of the digital world everyone and everything is digitally connected with each other with the help of personnel digital assistants and internet of things (IOT).This will also give rise to the problem of identity theft whether making a banking transaction, sharing a confidential file, or making a fraud with company. So a highly secure system is thus required. Biometrics plays a dominant role for providing the security it includes face recognition, iris recognition and fingerprint recognition among all of this fingerprint provide high level of abstraction to the user security. We have discussed some approaches, comparison of the existing enhancement techniques and by using different machine learning algorithms how we can increase the accuracy of present recognition system. In this project I have used the standard dataset fvc2002 db1 and applying different machine learning algorithms on this database by first without applying the principal component analysis and then with principal component analysis (PCA) and realized that random forest classifier(without PCA) produce the best results among all of this algorithms.

CONTENTS

LIST OF TABLES

List of figures

CHAPTER-1

INTRODUCTION

In the era of growing information and communication technologies, security became the serious concern for all of us that's why demand of security system has also increased. Fingerprint recognition systems act as a bridge to this demand since every individual has their own unique fingerprint, which reduces the chances of unauthorized access. Along with that it offers the least cost, fast and reliable way of providing security. Fingerprint recognition systems has wide variety of applications including law enforcement, access control, IT systems security, border management system airport [1]. Fingerprint consists of ridges and valley and every individual has unique pattern of ridges. Generally pattern of ridges can be classified into 3 types that is arch, loop and whorls [2].

The major steps involved in the fingerprint recognition system are acquiring the data it can be either online or offline, followed by pre-processing, feature extraction, and classification and then pattern identification, and matching. The acquisition or enrolment can be done via optical finger reader or through ink impression on the paper the later method is known as offline method and former called online method. The major challenging part of this process comes when the fingerprint image is noisy or corrupted due to poor skin condition or dirt accumulated on the finger. So the fingerprint image is subjected to pre-processing and enhancement step which is then followed by Binarization and thinning which would be necessary to clearly identify the ridge structure. Generally in most of the researches done earlier Gabor filter is widely used in the enhancement process

Whether it is Texture Segmentation [3], Face recognition [4], Fingerprint enhancement [5], although Fast Fourier Transform (FFT) were also used.

The popularity of Gabor enhancement is due to its characteristics such as the frequency and inclination representation which is similar with the visual system of humans. The function for gabor filter in complex form can be described as follows:

$$g_{\lambda\theta\psi\sigma\gamma}(x,y) = exp(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2})cos(2\pi\frac{x'}{\lambda} + \psi)$$

$$x' = xcos(\theta) + ysin(\theta),$$
$$y' = ycos(\theta) - xsin(\theta).$$

Wavelength (λ): it represents the number of cycles per pixel.
Orientation (θ): Angle of the normal to the parallel strips of gabor function
Phase angle (ψ): it is the offset of the sinusoid.
Spatial aspect Ratio (γ): it represents the ellipticity which is generally less than 1.

CHAPTER-2

LITERATURE REVIEW

Fingerprint recognition process has gone through series of steps fingerprint enrolment, pre-processing, feature extraction, then fed to the classifier. Pre-processing is required to deal with noisy image so to avoid any error in the results and it can be further divided into five types normalisation, enhancement, binarization, filtering and thinning[6].

2.1 Image Pre-processing

Pre- processing of image is required to deal with noisy image so to avoid any error in the results and it can be further divided into five types normalisation, enhancement, binarization, filtering and thinning[6].

2.1.1 Normalisation

The main purpose of normalisation is to compress the pixel values of the fingerprint image so that it comes under the desired range which makes the computations efficient. Basically it makes the pixels of the image either 0 or 1. Normalisation function can be written as follow

$$F(x, y) = M_0 + \frac{\sqrt{v0 * (I(x,y) - M)^2}}{\sqrt{v}} \quad \text{if} \quad I(x, y) > M$$

$$= M_0 - \frac{\sqrt{v0 * (I(x,y) - M)^2}}{\sqrt{v}} \quad \text{otherwise}$$

M_0 : desired mean

M : mean of the image pixels

V0 : desired variance

V :variance of input image pixels

2.1.2 Image enhancement

In order to extract the feature from the image there could be the problem of scars, blurs and incipient ridges so to minimize such undesirable effects and to extract the certain quantities of good features it can be ridge bifurcation, ridge ending or can be singularities likes cores and delta image enhancement methods are required.

Original Enhanced

Fig-i

2.1.3 Binarization

In this process a grey level image is transformed into black and white image that is during the process of binarization if the pixel value of fingerprint image is more than threshold it is treated as 1(white) otherwise 0(black) so after this images has having the ridges and valleys. Ridges corresponds to black and valley as white

Original **Binarized**

Fig-ii

2.1.4 Thinning

It is a morphological operation and its primary objective is to reduce the pixels so that it becomes only one pixel by eliminating the redundant pixels present in the image

Original **Thinned**

Fig-iii

2.2 Image enhancement methods and comparison

2.2.1 Histogram equalization

Histogram equalization is a method to alter the pixels of the image it may lead to enhance the contrast of the image but not always there may be some images where histogram equalization process does fit well for contrast enhancement. The process of histogram equalization is as follows:

We are given an image in the form of matrix

1	2	7	5	6
7	2	3	4	5
0	1	5	7	3
1	2	5	6	7
6	1	0	3	4

Table 1

Assuming the image is of 3 bpp(bits per pixel) for histogram equalization we have to calculate the probability mass function(PMF) and cumulative distribution function(CDF)

Probability mass function is the probability of count of different numbers present in the image

Pixels values	Frequency of pixels	Probability mass function(PMF)
0	2	2/25
1	4	4/25
2	3	3/25
3	3	3/25
4	2	2/25
5	4	4/25
6	3	3/25
7	4	4/25

Table 2

The total frequency of the pixels value is equal to 25 and the probability mass function can be written as

$$\text{PMF} = \frac{\text{count of the frequency of each pixel}}{total\ \ frequency\ of\ the\ pixel}$$

Corresponding CDF is given by

Grey level value	CDF
0	0.11
1	0.22
2	0.55
3	0.66
4	0.77
5	0.88
6	0.99
7	1

TABLE 3

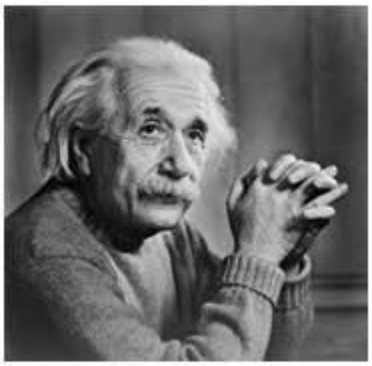
So, CDF is cumulative addition of the consecutive values of the pmf in the table. After that each value of CDF is multiplied by its corresponding grey level value to get the new grey value for the pixel

Grey level value	CDF(cumulative distribution function)	GIF{CDF * No. of Grey levels-1}
0	0.11	0
1	0.22	1
2	0.55	3
3	0.66	4
4	0.77	5
5	0.88	6
6	0.99	6
7	1	7

TABLE 4

Gif: greatest integer function

So our old grey values changed to new pixel values and our original image contrast is enhanced. It might possible for some image that after the histogram equalization the new grey level value of pixels is reduced



Original Image **Histogram equalized image**

Fig iv

2.2.2 Gabor filtering

It is a linear filter used in texture analysis and is somewhat similar to the human visual system. Whenever there is particular frequency component present in the image or localized in the region gabor filtering is generally used. There are various application in which Gabor filter is used like facial expression recognition, in document image processing they also have interesting applications in optical character recognition, it is also used in create the different activation functions for representing the sparse image object.

2.2.3 2D Fourier Transform

It states that any aperiodic signal can be transformed into periodic signal and it basically deals with the frequency domain representation of time domain signals. It has huge applications in the image processing, since image comprised of different range of pixel values which can be of high frequency or low frequency so interesting results could be obtained after applying the Fourier transform on the image pixels. The discrete Fourier transform function and its inverse transform can be given as

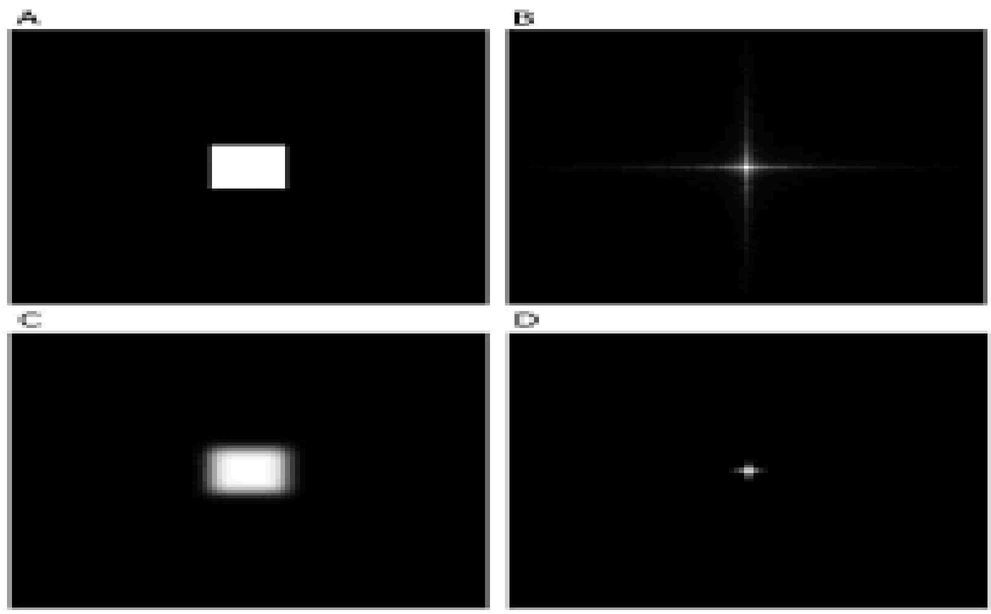
$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(\frac{ux}{M}+\frac{vy}{N})}$$

$$f(x, y) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} F(u, v) e^{+j2\pi(\frac{ux}{M}+\frac{vy}{N})}$$

Since frequency is directly related to the rate of change we can associate the change in intensity of pixels value with the frequencies of Fourier transform.

When u=0 and v=0, the Fourier transform give the slowest varying component of grey level image as we move away from the origin the low varying components corresponds low frequency components of image but as we move farther away from the origin that is towards higher frequency component the rate of change of grey level value is higher.
The original image and its Fourier transform can be described as below

Original image Fourier transform



<u>Fig-v</u>

2.2.4 Wavelet based transform

Images has smooth regions interrupted by edges or abrupt changes in contrast these abrupt changes often the most interesting part of the data both perceptually or information they provide since Fourier transform is useful tool in data analysis but it fails to represent abrupt changes efficiently the reason being is Fourier transform represent signal as the sum of sine waves which are not localized in time and space these sine waves oscillates forever therefore to accurately analyzed signals and images which has abrupt changes we need to use the new class of functions which are well localized in time and frequency these are known as wavelets. A wavelet is rapidly decaying wave like oscillations that has zero mean unlike sinusoids which are extends to infinity the wavelet extend for the finite duration wavelets come in different sizes and shapes example haar, daubechies, Morlet etc. There are two important wavelet transform concepts namely scaling and time shifting.

2.2.4.1 Scaling

Scaling refers to the process of stretching and shrinking the signal in time which can be expressed using the equation

$$\Psi(t/s) \ ; \ s>0$$

Where s is the scaling factor and it must be the positive value

The scale factor ‘s’ is inversely proportional to the frequency for example scaling the sine wave by 2 resulting it by reducing it frequency by half or by an octave

For a wavelet there is reciprocal relationship between the scale and the frequency with a constant of proportionality this constant of proportionality is called the center frequency of the wavelet this is because unlike sine wave the wavelet has band pass characteristics in the frequency domain

Mathematically it is given as

$$F_{eq} = C_f / s \, \xi t$$

Where C_f : is the center frequency

S : is the scaling factor

ξt : is the sampling interval

<u>2.2.4.2 Shifting</u>

Shifting the wavelet means advancing or delaying the onset of the wavelet along the length of the signal

Φ (t-k) this notation means the wavelet is shifted and centered by k we need to shift the wavelet allowing it to allowing the feature we are looking for in a signal.

Wavelet transform can be of two types continuous type wavelet transform and discrete type wavelet transform

2.2.4.3 Continuous wavelet transform

This transform is used to obtain the simultaneous time and frequency analysis of the signal analytic wavelets are best suited for this purpose as these wavelets do not have negative frequency component
The output of CWT are the coefficients which are the functions of time, frequency and scale

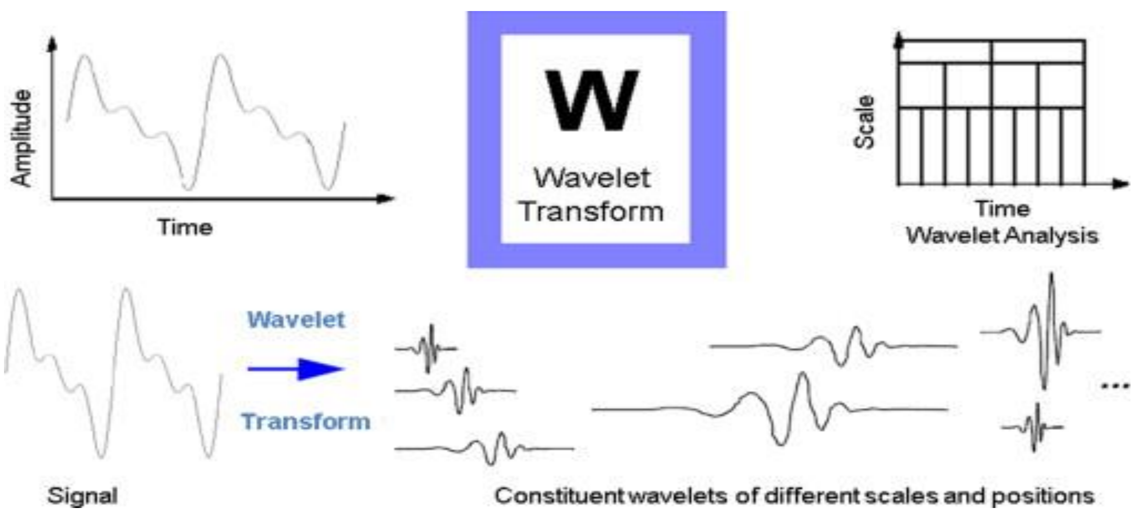


Fig-vi

With cwt we have added flexibility to analyze the signal at an intermediate scale within each octave this allows for the fine scale analysis a signal with 1000 samples analyze with 20 scales can results to 20,000 coefficients in this way we can better characterize the oscillatory behavior of signals in continuous wavelet transform

2.2.4.4 Discrete wavelet transform

Discrete wavelet transform or DWT is ideally suited for de-noising and compressing signals and images as it helps to represent many naturally occurring signals and images with fewer coefficients this enables sparse representation the base scale in DWT is set to 2^j (j=1, 2, 3, 4…..) we can obtain different scales by changing the powers of 2 and the translation occurs at rate 2^jm(m=1,2,3,4…..) this process is often referred to as dyadic scaling and shifting. This kind of sampling remove redundancy in coefficients the output of the transform yield same number of coefficients as the length of the input signal therefore it requires the less memory the discrete wavelet transform method is equivalent to signal comparable to multi rate filter banks

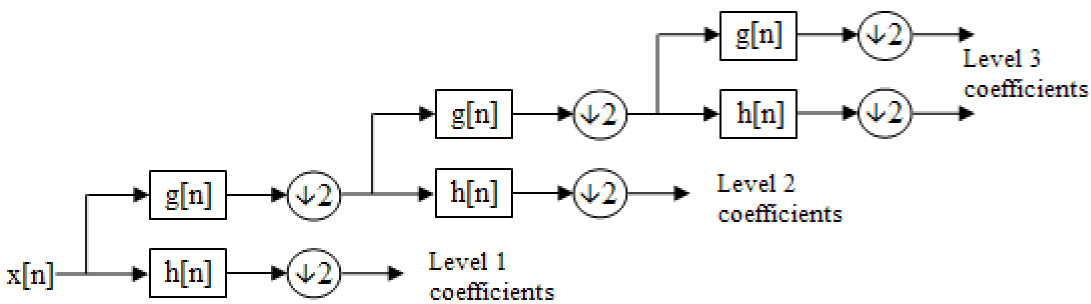


Fig vii

Given a signal is first pass through a low pass and high pass filter to yield low pass and high pass sub bands half of the samples are discarded after filtering as per the Nyquist criterion the filter typically

Has small number of coefficients and result in good computational performance these filters also has the capability to reconstruct the sub bands while cancelling any aliasing that occurs due to down sampling for the next level of decomposition the low pass sub band is iteratively filtered by the same technique to yield narrowing sub band the length of the coefficients in each sub bands is half of the number of coefficients in the preceding stage with this technique we can capture the signal of interest with a few large magnitude DWT coefficients where the noise in the signal results in smaller DWT coefficients this way DWT helps analyzing signals at progressively narrowing sub bands at different resolutions it also helps in de noise and compress signals.

2.2.5 Comparison of various enhancement methods

There have been many enhancement methods discussed [2].Each has their own merits and demerits

Methods	Demerits	Merits
Histogram Equalization	This method sometimes enhances the background noise in the image	This method straight works on the finger print image pixels
Band pass filtering	This technique fails if the image has huge noise	To some extent noise removal to retain the true structure and ridges of fingerprint
Gabor filtering	This method will not produce good results if the image contains noises	This method produce good results when the parameters of anisotropic and low pass filter are combined
Binarization and thinning	This method sometimes results in deviated lines for empty medial lines of fingerprint	fingerprint parameters and ridges connections is maintained in this method
2D Fourier Transform	This method drops the accuracy as it is continuously accepting the frequency	This is fast and categorizes the location to 16 directions
Wavelet based Transformation	Since its wavelets \are divisible therefore not suited for diagonal parameters	Fast and efficient in de noising fingerprint images
Wave atom Transform		This method hasn't any disadvantages until now as with the other methods

Table-5

2.3 Feature Extraction

Feature extraction in the fingerprint recognition system can be done via two methods pattern based approach and minutiae based approach in pattern based[7].Generally in pattern based methods fingerprint pattern(arches, loops and whorls) are used to classify the candidate fingerprint and stored template where as in minutiae based approach minutiae(fine details) are used for the identification purpose. Although there are various different local characteristics which can be singular points (core and delta), ridges, valleys etc. but most prominently ridge endings and ridge bifurcations are used because of their stability and robustness.

So in order to identify these features cross number(CN) method is used in which 3*3 mask is moved on the pixels of the image and if the central pixel in the mask is 1 and if their at least 3 neighborhood pixels are exactly 1 then it will be considered as ridge bifurcation if the central pixel in the mask is 1 and if there is only one neighborhood pixels is exactly 1 then it will be considered as ridge termination

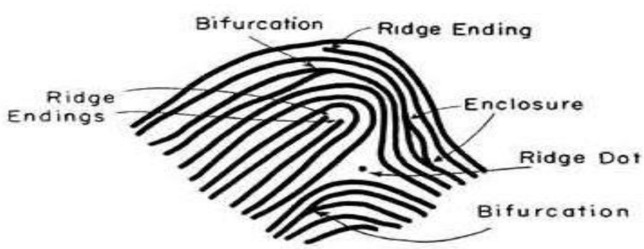


Fig viii

2.3.1 Cross number method

It is most widely used method in order to identify the features in fingerprint recognition for ridge bifurcation and ridge termination as features following methodology is used

0	0	1
0	1	0
0	0	0

Table-6

3 X 3 Mask

If the center pixel is 1 and exactly only one pixel in its neighborhood is 1 then it is called ridge termination

1	0	1
0	1	0
0	1	0

Table-7

If the center pixel is 1 and at least 3 pixels in its neighborhood are 1 then it is called ridge bifurcation

2.4 Classification and identification

The main purpose of the classification and identification of fingerprint image is to classify the finger print of different owners by train our algorithm so that when training is completed it can able to predict on test dataset. The algorithm we discussed are random forest K-NN, SVM (support vector machine), Naïve bayes and deep learning model like CNN (convolutional neural networks)

We shall discuss all these algorithms in next subsequent chapter.

CHAPTER-3

Algorithms proposed

3.1 Machine learning algorithms

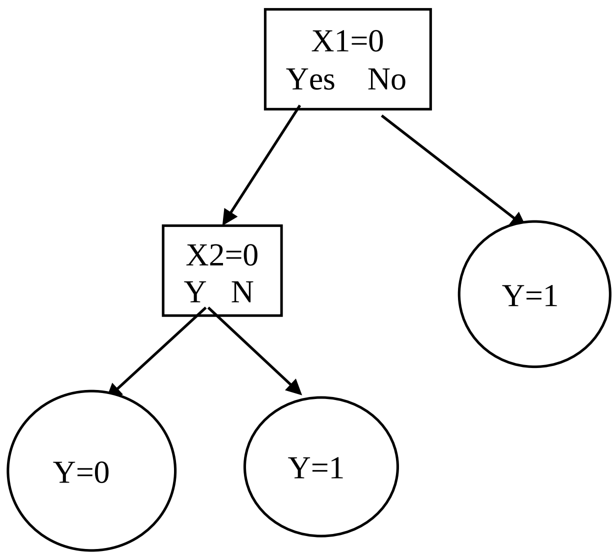
3.1.1 Random forest

For understanding the random forest we should first understand the decision tree.

A decision tree is like a flow chart where node(except terminal node) represents a test on an attribute and each branch represents the test outcome and the terminal node which is known as leaf is used to identify the class of the dataset for example a simple decision tree for The given 'OR' gate truth table can be described

X1	X2	Y
0	0	0
0	1	1
1	0	1
1	1	1

Table-8



<u>Fig ix</u>

(‘OR’ gate decision tree)

So now the question arises how to build the decision tree. In order to build the decision tree first of all we have to find a feature to split on our data into two parts or more than two parts now what will be the feature do we need to split our decision tree so that we get the best possible decision tree for that we need to check our metrics that is accuracy, information gain, Gini index and gain ratio the score of these parameters will decide that what could be the best possible feature to split upon.

Drawbacks of decision trees: There are some limitations associated with the decision tree so a very common problem is that we will start from a node and we will keep splitting it until we try to put a control on it to a point till the point when we end up in reaching absolutely zero training errors therefore it leads to overfitting of the data. One thing which will help us out to not reaching the zero training error is that to run out of the features in that case we might end up with the leaf which is having impurity and you will have training errors.

The ways to avoid overfitting can be stop early and pruning. In stop early we try to simplify the decision tree and stop whereas in pruning we build the whole tree and then we do the pruning by deciding which split is not that useful lets discuss these two ways in detail

Stop early: There are two most intuitive way of doing this firstly by defining the maximum path in which we define the maximum number of levels 'k' beyond which we are not going further but there are problems associated with this approach for example if the tree is going in a particular direction so imposing a max depth on it will not going beyond further that is force balance on both sides.

If in the stop early method we may stop at point when there is no significant in the accuracy but this is also has the drawback suppose if for 2 iterations accuracy is same but on next iteration accuracy suddenly increased so we cannot rely on this approach also.

Pruning: for the training data as the complexity of the data increase error will get to decrease but for the testing data after a certain point it starts increasing so defining a cost function as the complexity of tree is directly related with the number of splits, more the number of splits more complex will be the tree

$$\text{Cost} = \text{errors} + \lambda \text{ (no. of leafs)}$$

So we need to reduce the cost as much as possible. Pruning is done from bottom to top that is reduce the tree complexity by reducing the number of leafs but one thing should be make sure we have to make the tree simple not that simple which may produce more errors that's why we have a tweak 'λ' to knob the complexity and no. of errors

Now we can come to our original topic that is random forest, so what is random forest?

It is nothing but the forest of decision trees combined together to give the results and the prediction of final outcome will be based on the

the majority score. Now the question arises why the decision trees will produce the different results for a particular data, well in order to get the different results from the decision tree we use the randomization in data and features so that we might get outliers which will affect some of decision trees but not all the decision trees. Now the randomization can be achieved by using the two algorithms that is data bagging and feature selection

Bagging: (bootstrap aggregation)

According to this algorithm if we have the 'm' data points and selecting 'k' data points out of them but the selection is done with replacement it means one data point can be selected multiple times therefor it might be a possibility some of the data points may not come it means it gives the preference to some of the data points and some of the data points get missing

Feature selection: Along the bagging this is the second step so in a feature selection we are not going to train our tree on all the features instead the first decision tree is train on some features randomly selected. Now the question arises how many features we are going to choose?

- In order to train our first decision tree we can choose 'k' features out of 'N' features
- When we train other decision trees we can now choose 'k' features out of 'N' this is without replacement

Suppose we have features f1, f2 ,……f10 then if we have chosen 4 features f1,f3,f5,f7 so in order to train other decision tree the four features to choose will be completely different from the previously chosen

Generally in order to select the k features out of ‘N’ features we choose the number generally k=$\sqrt{N}$

Generally no. of decision trees taken in random forest is not limited to just 3 or 4 but it actually 8, 9 or 10.

The crux points which can be concluded from the decision tree and the random forest is that decision trees performing better on training but not good for testing On the contrary random forest perform worse on training but better on testing.

2.4.2 Naïve Bayes

This classifier is based on the probability where the prediction of the class is done by using Bayesian probability. Naive classifier is a collection of algorithms where each one has the common principle that is every pair of feature to classify is independent of each other for example let us consider the following dataset

Outlook	Temperature	Humidity	Windy	Play golf
Rainy	Cool	High	False	Yes
Sunny	Hot	Normal	True	No
Rainy	Cool	Normal	False	Yes
Rainy	Mild	High	True	No
Overcast	Mild	Normal	False	No
Sunny	Hot	High	True	Yes
Overcast	Cool	Normal	False	Yes
Overcast	Hot	High	True	Yes
Sunny	Mild	High	False	No

Table-9

The dataset is divided into two parts that is feature matrix and response vector which contains the output of the feature value. For this case outlook, temperature, humidity and wind are the features and the play golf is our output so we use Bayesian probability for predicting the class of our output

Bayes theorem

P (A/B) = P (B/A) P (A)/P (B)

For multi features we assume that features are independent of each other

For i^{th} class

$P(y=a_i/X=x) = P(X=x/y=a_i)\, P(y=a_i)/P(X=x)$

Where

$P(X=x/y=a_i) = \prod_{i=1}^{n} P(X = xi\, /(y = ai))$ for 'n' features

It may possible sometimes for some class that the data its probability for the particular set of features is 0 in that we would be using the Laplace correction

2.4.3 KNN Classifier

K nearest neighbor classifier this algorithm will classify our test data class on the basis of distance it can be Euclidean or man-ha-tan based on the majority voting

Euclidean distance can be given as

$$d_{ecl}=\sqrt{\sum_{i=1}^{n}(X1(i)-X2(i)}$$

$$d_{man}=\sum_{i=1}^{n}|X1(i)-X2(i)|$$

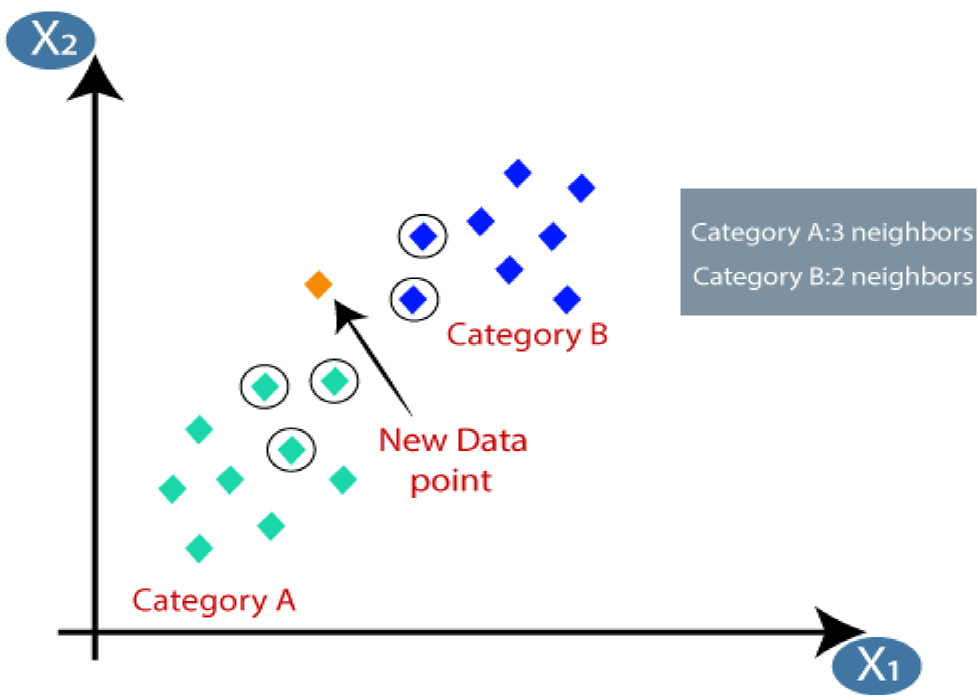


Fig-x

Our predicting the class of the new data point in the given we can use k-NN where k=1, 3, 5…. If we choose k=1 then its most- closest neighbor will be identified its class it is not generally preferred to choose 1 as it is prone to overfitting because then decision boundary will be very much complex so the optimum k is always desired.

While applying the k-NN to our test data we must do the feature scaling as it may possible that for example our data set has two features and one feature set is containing very high value compared to other then the first feature can completely overpowers the effect of other.

Cross validation: Which distance metric to choose either Euclidean or man-ha-tan to produce the good results that again the problem so the solution of the problem is given by the cross validation concept it states that lets split our data into ‘k’ parts out of which ‘k-1’ is used for training and test on the remaining then applying the same to all the data points

The value of k must be chosen to be optimum as if k value is large then it will lead to the under fitting when it is low then it will lead to overfitting

Curse of dimensionality

In most of the classifier if you have more number of features then this generally means we have more data and its eventually leads to better accuracy butter k-NN is little different from the others as if you very

Irrelevant data or correlated data this might lead to lot of problems that is double calculation of the same data which can lead to the error therefore in order to avoid this we can do two things to reduce this problem

1) Assigning of weights to our features
$$\sum_{i=1}^{n} w(i)(X1(i) - X2(i))^2)$$

For assigning the optimum weights one can use gradient descent with respect to weight

2) Feature selection this is also known as backward elimination in this method we first choose a set of features not all features and have a loop for all features here we first check the accuracy by including the feature and then by excluding that feature so the decision whether to retain the feature or to exclude the feature will depends upon the accuracy

Pros and cons of KNN

Pros: It is very easy to understand and code and it works for the multiclass as well

Cons: Testing time is large as it had to compare the testing data with all the training data points.

If training data split is biased that is one class has lot more data points

Then the other class, then KNN tends to get biased as well.

2.4.4 SVM (support vector machine)

The basic intuition behind SVM is to increase the confidence interval margin so that we can clearly identify the two classes

For example in case of logistic regression we have the transfer function

$$h_{\theta x} = 1/_{1+\ e^{-\theta x}}$$

This function is hypothesis function its graph is plotted by taking h_θ on the y axis and θx on the x axis which identifies the class 1 for $h_\theta > 0.5$

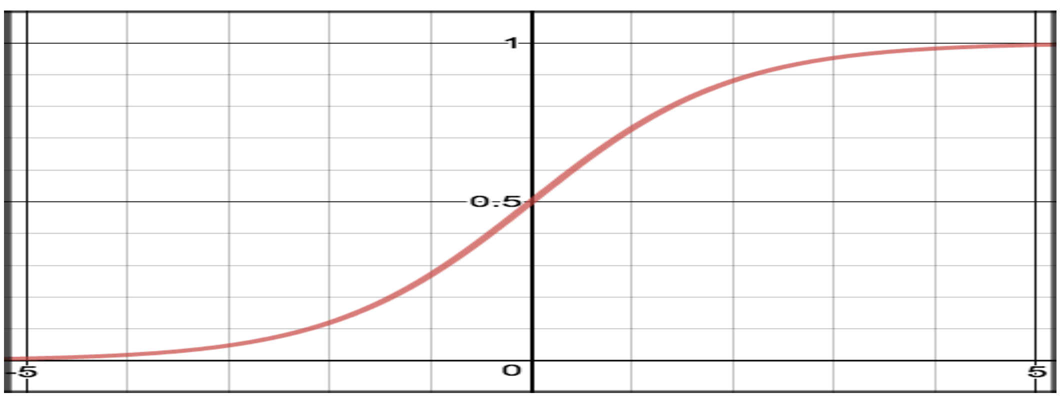


Fig xi

And class 2 for $h_\theta<0$ so the purpose of SVM is to widen the space in order to distinguish between the two classes for example it can like for x1>1 it belongs to class 1 and for x2<-1 it belongs to class 2

So we can modify the hypothesis function by taking the log of hypothesis in order to find the cost function of SVM

Cost=log ($h_{\theta x}$)

=log ($1/_{1+\ e^{-\theta x}}$) ; y=1

Cost=log (1- $1/_{1+\ e^{-\theta x}}$) ; y =0

The overall cost is given by

Overall cost=y_i $cost_1$ ($h_{\theta x}$) + (1-y_i)* log (1- $h_{\theta x}$) + λ/2 $\sum \theta$^2 or it can be further written as

Overall cost=c y_i $cost_1$($h_{\theta x}$) + (1-y_i)log(1- $h_{\theta x}$)+1/2 $\sum \theta$^2 where c=1/λ

a very high value of 'c' leads to overfitting and very low value of 'c' leads to under fitting so optimum value of 'c' is desired to get the optimum results

3.2 Deep learning algorithms

Convolutional neural networks

It is a type of artificial neural network in which connecting pattern between the neurons is inspired by the organization of animal visual cortex

Working of CNN is classified by the 4 layers

1) convolutional layer
2) Re-Lu layer
3) Pooling
4) Fully connected layer

CNN compares the images piece by piece the pieces that it looks are called features so what we do in the CNN is taking small patches of images known as features of the image

Convolutional layer: Here we will move the feature/filter to every possible position in the image with 4 simple steps

Step1: line up the feature and the image

Step2: Multiply each image pixel by the corresponding pixel value

Step3: Now after multiplying the feature image pixel with the corresponding image pixel we need to add them up

Step4: and finally divide by the total number of pixel in the image

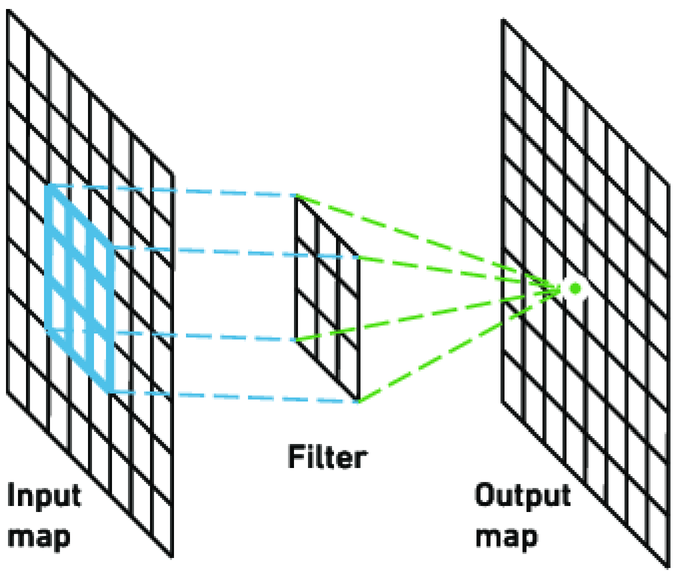


Fig xii

(Convolutional layer)

Re-Lu layer: It is basically a activation function which only activates a node if the input is above a certain threshold otherwise it outputs zero. Re-Lu layer function can be given as

$$F(x) = \begin{cases} 0 & \text{if } x<=0 \\ 1 & \text{if } x>0 \end{cases}$$

X	F(x=X)	Output
-1	F(-1)	0
-5	F(-5)	0
3	F(3)	3
1	F(1)	1

Table -X

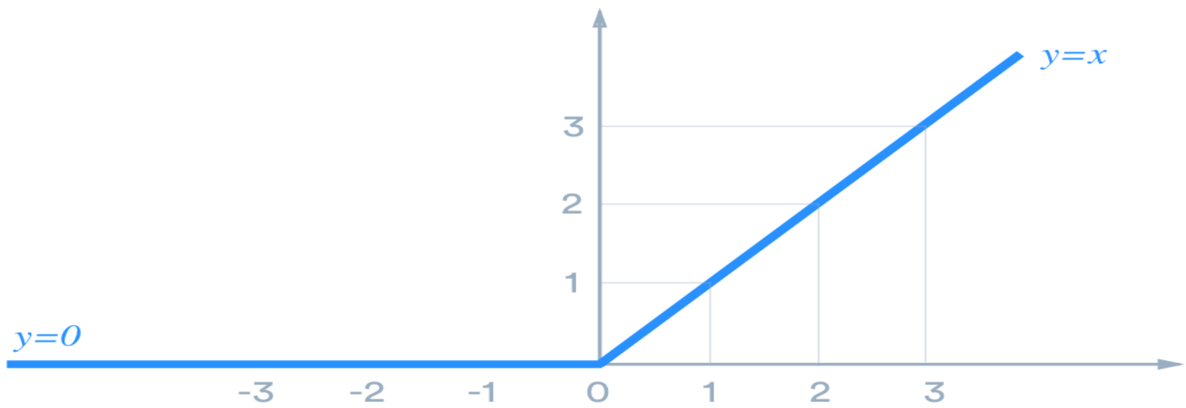


Fig xiii

Pooling Layer: In this layer we shrink the image stack into the smaller size by doing these simple steps

1) pick a window size usually it is 2 or 3
2) pick a stride (usually 2) is taken
3) walk your window across the filtered images
4) from each window take the max value

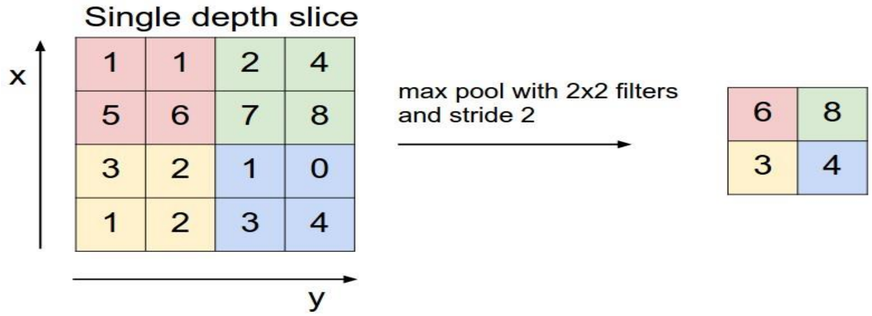


Fig xiv

Fully connected layer: This is the final layer where actual classification happens here we take our filtered and shrink images and put them into a single list and after that score is evaluated for the different features from the sum of stack values of the particular feature the more closest the summation score with specific feature value is identified as the particular feature

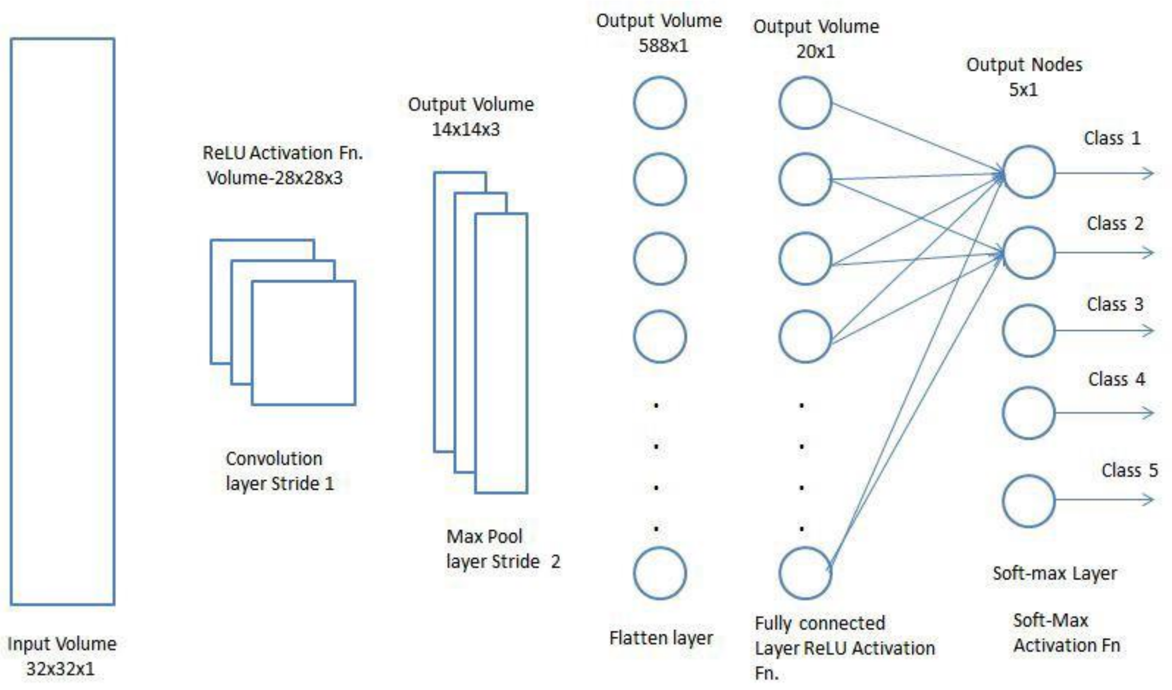


Fig xv

Chapter-4

RESULTS COMPARISON

Machine learning Algorithms	Accuracy% results(without using PCA)	Accuracy% results(with using PCA)
Random forest	40%	25%
SVM with linear kernel	30%	15%
KNN	30%	10%
Multinomial Naïve bayes	35%	20%
Multilayer perceptron	10%	20%

The purpose of applying the PCA to our dataset is to check that how well our algorithms performed if the redundant features are removed accuracy is reduced with less feature in all cases with exception of multilayer perceptron.

Implemented code along with the dataset are being sent along with the thesis.

OBSERVATIONS AND CONCLUSIONS

Observations are made after going through the papers that fingerprint recognition using deep learning algorithm provide much better classification then the machine learning algorithms with effective preprocessing technique combined with deep learning can reduce the training time without hampering the accuracy. Although fingerprint recognition accuracy can be affected due to several factors such as environmental conditions, age, illness, personal cause etc as discussed [13].The finger vein recognition on the contrary does not encounter any such problem and could achieve a very high accuracy of more than 99percent[12]

Since fingerprint using deep learning can enhance the accuracy but has some anomalies as discussed so finger vein technology on the other hand prevent such limitations and also enhance the accuracy to a much great extent and also the since vein pattern is not visible so it is very difficult to forge. So the future work can be on the finger vein recognition with deep learning architecture and on a larger database.

REFERENCES

[1] Babatunde, I. G., Kayode, A. B., Charles, A. O., & Olatubosun, O. (2012). Fingerprint image enhancement: Segmentation to thinning.

[2] Borra, S. R., Reddy, G. J., & Reddy, E. S. (2016, March). A broad survey on fingerprint recognition systems. In 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (pp. 1428-1434). IEEE.

[3] Weldon, T. P., Higgins, W. E., & Dunn, D. F. (1996). Efficient Gabor filter design for texture segmentation. Pattern recognition, 29(12), 2005-2016.

[4] Lampinen, J., & Oja, E. (1995). Distortion tolerant pattern recognition based on self-organizing feature extraction. IEEE Transactions on Neural Networks, 6(3), 539-547.

[5] Hong, L., Wan, Y., & Jain, A. (1998). Fingerprint image enhancement: algorithm and performance evaluation. IEEE transactions on pattern analysis and machine intelligence, 20(8), 777-789.

[6] Oulhiq, R., Ibntahir, S., Sebgui, M., & Guennoun, Z. (2015, October). A fingerprint recognition framework using artificial neural network. In 2015 10th international conference on intelligent systems: theories and applications (SITA) (pp. 1-6). IEEE.

[7] Ali, M. M., Mahale, V. H., Yannawar, P., & Gaikwad, A. T. (2016, February). Fingerprint recognition for person identification and verification based on minutiae matching. In 2016 IEEE 6th International Conference on Advanced Computing (IACC) (pp. 332-339). IEEE.

[8] Zhou, J., Chen, F., & Gu, J. (2008). A novel algorithm for detecting singular points from fingerprint images. IEEE transactions on pattern analysis and machine intelligence, 31(7), 1239-1250.

[9] Ali, A., Khan, R., Ullah, I., Khan, A. D., & Munir, A. (2015, October). Minutiae based automatic fingerprint recognition: Machine learning approaches. In 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (pp. 1148-1153). IEEE.

[10] Pandya, B., Cosma, G., Alani, A. A., Taherkhani, A., Bharadi, V., & McGinnity, T. M. (2018, May). Fingerprint classification using a deep convolutional neural network. In 2018 4th International Conference on Information Management (ICIM) (pp. 86-91). IEEE.

[11]Shrein, J. M. (2017). Fingerprint classification using convolutional neural networks and ridge orientation images. In 2017 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-8). IEEE.

[12] Liu, W., Li, W., Sun, L., Zhang, L., & Chen, P. (2017, June). Finger vein recognition based on deep learning. In 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA) (pp. 205-210). IEEE.

[13] Rathod, V. J., Iyer, N. C., & Meena, S. M. (2015, October). A survey on fingerprint biometric recognition system. In 2015 International Conference on Green Computing and Internet of Things (ICGCIoT) (pp. 323-326). IEEE.