LIVER CANCER DETECTION AND CLASSIFICATION BASED ON OPTIMUM FEATURE FUSION

Project report submitted in partial fulfillment of the requirements for the Degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

By

GARIMA MALHOTRA (161027)

KUSHAGRA SRIVASTAVA (161033)

MANTHAN CHAUHAN (161117)

UNDER THE SUPERVISION OF

DR. SHRUTI JAIN

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT

May, 2020

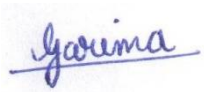
TABLE OF CONTENTS

DECLARATION iv
ACKNOWLEDGEMENT v
LIST OF ACRONYMS AND ABBREVIATIONS vi
LIST OF FIGURES vii
LIST OF TABLES viii
ABSTRACT ix
CHAPTER 1 10
1.1 IMAGE PROCESSING 10
1.2 BIO-MEDICAL IMAGE PROCESSING 10
1.3 LIVER 11
1.3.1 STRUCTURE OF LIVER 12
1.3.2 FUNCTIONS OF LIVER 13
1.3.3 CLINICAL SIGNIFICANCE 14
1.3.4 LIVER REGENRATION 22
1.3.5 LIVER TRANSPLANTATION 22
1.4 MOTIVATION 23
1.5 SOFTWARE USED 23
1.5.1 MATLAB 23
1.5.2 Microsoft Office Picture Manager 23
1.5.3 ImageJ 24
CHAPTER 2 26
2.1 RESEARCH PAPERS 26
CHAPTER 3 29
3.1 IMAGE MODALITIES 30
3.1.1 CT SCAN 30
3.1.2 MRI 30
3.1.3 COMPUTER AIDED DESIGN 30
3.2 IMAGE FUSION 30
3.3 FEATURE EXTRACTION 33
3.3.1 TYPES OF FEATURE EXTRACTION 37
3.3.2 STATISTICAL BASED FEATURES [43] 39
3.3.3 HARALICK TEXTURE FEATURES [44] 40

3.4 CLASSIFICATION 46
CHAPTER 4 49
4.1 Image Fusion and SURF 49
4.1.1 Dataset 1 49
4.1.2 Dataset 2 51
4.1.3 Dataset 3 52
4.1.4 Dataset 4 53
4.1.5 Dataset 5 54
4.1.6 Dataset 6 55
4.1.7 Dataset 7 57
4.1.8 Dataset 8 58
4.2 CLASSIFICATION RESULTS 59
CONCLUSION AND FUTURE SCOPE 69
REFERENCES 70
PUBLICATION 74
PLAGIARISM REPORT 75

DECLARATION

We hereby declare that the project work entitled '**Liver Cancer Detection and Classification Based on Optimum Feature Fusion'** submitted to the Department of Electronics and Communication at **Jaypee University of Information Technology, Solan** is a record of an original work done by us. This project work is submitted as a part of partial fulfillment under the guidance of **Dr. Shruti Jain** and we have not submitted this work elsewhere for any other degree or diploma.



Garima Malhotra
161027

Kushagra Srivastava
161033

Manthan Chauhan
161117

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Dr. Shruti Jain
Associate Professor
Jaypee University of Information Technology
Waknaghat, Distt. Solan,
Himachal Pradesh.

Date: 25/05/2020

ACKNOWLEDGEMENT

We take this opportunity to express our gratitude to our supervisor Dr. Shruti Jain for her insightful advice, motivating suggestions, invaluable guidance, help and support in successful completion of this project and also for her constant encouragement and advice throughout our Bachelors program.

The in-house facilities provided by the department throughout the Bachelors program are also equally acknowledgeable. We would like to convey our thanks to the teaching and non-teaching staff of the Department of Electronics and Communication Engineering for their invaluable help and support.

Garima Malhotra (161027)
Kushagra Srivastava (161033)
Manthan Chauhan (161117)

LIST OF ACRONYMS AND ABBREVIATIONS

CT	Computed Tomography
MRI	Magnetic Resonance Imaging
OCT	Optical Coherence Tomography
PET	Positron Emission Tomography
SPECT	Single Photon Emission Computed Tomography
FMRI	Functional Magnetic Resonance Imaging
HAV	Hepatitis A Virus
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HCC	Hepatocellular Carcinoma
AFP	Alpha-fetoprotein
PSC	Primary Sclerosing Cholangitis
CAD	Computer Aided Diagnosis
PACS	Picture Archiving and Communication System
DWT	Discrete Wavelet Transform
PCA	Principal Component Analysis
IHS	Intensity Hue Saturation
CBIR	Content Based image retrieval
GLCM	Gray-Level Co-Occurrence Matrix
SURF	Speeded Up Robust Features
SIFT	Scale-invariant Feature Transform
kNN	k-Nearest Neighbour
SVM	Support Vector Machine
GBM	Gradient Boosting Algorithm

LIST OF FIGURES

Figure 1.1: Liver Image [3] **11**
Figure 1.2: Liver Structure [4] **12**
Figure 1.3: Liver Blood Supply Function [5] **13**
Figure 1.4: Liver Diseases [6] **14**
Figure 1.5: Stages of Liver Damage [9] **18**
Figure 1.6: Cancer Cells [11] **20**
Figure 3.1: Methodology Flow Diagram **29**
Figure 3.2: Image Fusion Example [31] **31**
Figure 3.4: PCA Block Diagram showing P1I1+P2I2 [34] **32**
Figure 3.5: Behaviour of Skewness [38] **34**
Figure 3.7: Level of Kurtosis [40] **36**
Figure 3.8: Gaussian partial derivative in xy **43**
Figure 3.9: Gaussian partial derivative in yy **43**
Figure 3.10: Showing Up-Scaling **44**
Figure 3.11: SURF Orientation **45**
Figure 3.12: Summation of dx and dy **45**
Figure 4.1: Fusion and SURF of Dataset 1 [50] **49**
Figure 4.2: Fusion and SURF of Dataset 2 [51] **51**
Figure 4.3: Fusion and SURF of Dataset 3[52] **52**
Figure 4.4: Fusion and SURF of Dataset 4 ..... **53**
Figure 4.5: Fusion and SURF of Dataset 5 ..... **54**
Figure 4.6: Fusion and SURF of Dataset 6 ..... **55**
Figure 4.7: Fusion and SURF of Dataset 7 ..... **57**
Figure 4.8: Fusion and SURF of Dataset 8 ..... **58**
Figure 4.9: Heat map of SVM Classifier **60**
Figure 4.10: Heat map of Adaptive Classifier **61**
Figure 4.11: Heat map of Decision Tree Classifier **62**
Figure 4.12: Heat map of Naive Bayes Classifier **63**
Figure 4.13: Heat map of Gaussian Naive Bayes Classifier **64**
Figure 4.14: Heat map of k-NN Classifier **65**
Figure 4.15: Heat map of Stochastic Gradient Descent Classifier **66**
Figure 4.16: Heat Map of Logistic Regression Classifier **67**

LIST OF TABLES

Table 4.1: Parameters for Dataset 1 ..... **50**
Table 4.2: Parameters for Dataset 2 ..... **51**
Table 4.3: Parameters for Dataset 3 ..... **53**
Table 4.4: Parameters for Dataset 4 ..... **54**
Table 4.5: Parameters for Dataset 5 ..... **55**
Table 4.6: Parameters for Dataset 6 ..... **56**
Table 4.7: Parameters for Dataset 7 ..... **57**
Table 4.8: Parameters for Dataset 8 ..... **59**
Table 4.9: Result of Classification **68**

ABSTRACT

It is rightly said ‘Precaution is better than cure’. There are many diseases which do not show any signs or symptoms. As a result the doctor finds it difficult to identify these diseases thus increasing complexity in cure and operations.

To this, we identified a solution to this problem. Image fusion, popularly used in medical can become a way to solve this issue. Images from CT Scans and MRI for other diseases can be fused together to identify if there is any sign of other disease.

In this project we are specifically working on liver and its disease. We will collect the data in the form of images from CT Scans and MRI and then fuse it together using different software to get a new result. Newly obtained images will help in identification of any other sign. This sign can be used to identify diseases apart from the known. Thus identification and treatment for unrecognized disease will become possible.

CHAPTER 1

INTRODUCTION

The liver is a large, meaty organ that sits on the right side of the belly. It performs a major role in a lot of vital body functions and any dys-functioning causes a lot of acute medical conditions. So in this project we will go through the liver, its structure, diseases, and the ways how we can detect it's dys-functioning and cure them at an early stage.

1.1 IMAGE PROCESSING

In applied Sciences, the use of algorithm to carry out image processing on images is the way of digital image processing. In order to be helpful to the computer file it gives permission to a large number of algorithms. Also the signal distortion and noise are some of the issues that can be avoided within the whole process.

In order to drive out some important information, image processing is the way. These days, amongst fast growing tech is image processing. Also amongst engineering science it accounts for main analysis space [1]. There are three important steps in image processing:

With the help of image acquisition tools the image is imported;

The image is analyzed and manipulated;

Report supporting image analysis or the outcome where the result is altered scan.

Analogue and digital image process are the two types of methods in image processing. For the onerous written matter for example images, Analogue image process is applied. Digital image process can facilitate in handling of the pictures by victimization of software used in a PC. The before processing, improvement, and show extraction of information are the three phases where everyone sorts the information ought to bear whereas victimization technique.

1.2 BIO-MEDICAL IMAGE PROCESSING

For diagnostic and therapeutic functions, medicine imaging is employed to capture the photographs. By the use of new Sensors and computer tech, snapshots of in vivo physiological processes and physiology are collected. Medicine imaging decided the use of whether Computed picturing scan or X-rays, magnetic resonance imaging or MRI, ultrasound, lightweight (endoscopy, OCT) or hot prescription drugs (nuclear medicine: SPECT, PET) to assess this status of associate body tissues and organs may be accustomed watch a sick person for a large amount of time for treatment and diagnosis assessment.

Proficiency referred to as the X-rays which facilitate processes like a biopsy and angiogram was the technique which digital imaging gave rise to as the CT scan which intern allowed doctors to observe period of time x-rays on a computer. Current analysis here specializes in what is reaped through useful image processing and isn't restricted towards straightforward anatomical reference image processing. CT scan and also tomography become particularly vital when there is coronary failure or it is suspected that the patient has coronary failure. Research workers are victimization useful tomography (MRI) differing kinds of head injuries or abdomen injures and also strokes.

With the improvement in technology, PET scans are the ones that use a hot tracker to check blood flow, chemical elements use and changes on a metabolic level. These scans help research workers in order to check, say for example, the brain working while a patient is under depression supporting the activity on a chemical level within the brain. The replacement space of analysis which will be using molecules and human cells while not requiring for a cell test or diagnostic test is known as the optical molecules imaging technology. This process helps us detect malady processes like cancer which can be noticed even before effects can be seen on a patient.

1.3 LIVER

The left lobe and right lobe are the two massive sections of the liver. The intestines and exocrine gland is under the liver and also the vesicle is placed below the liver. In order to digest, absorb and method food, the liver and other organs mentioned above work together [2]. The liver set within the right high quarter circle of our abdomen, which is under our diaphragm. Aside from this, different metabolic processes embody the regularization of polysaccharide store, disintegration of RBC's and therefore the forming of endocrines.

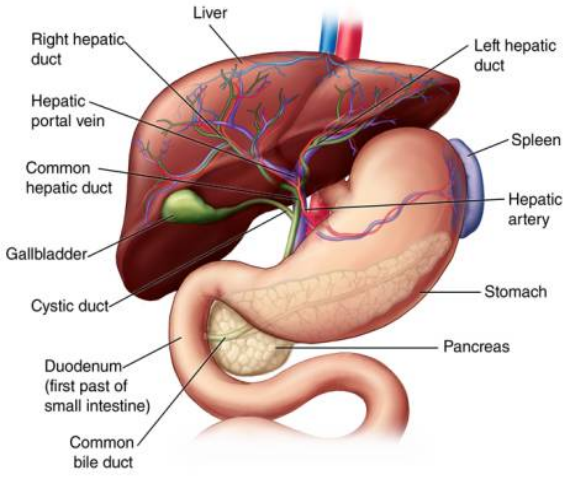


Figure 1.1: Liver Image [3]

The liver is an adjunct biological process that forms a digestive fluid, a fluid comprising sterol, digestive fluid acids. Associate in nursing a basic compound that helps within disintegration of fat. Petrifaction through the mixing or emulsification of lipids, digestive fluids come in play. Vesica, stores digestive fluid that is made by the liver that later enraptured to the tiny viscus to finish the method of digestion. Liver has extremely specialized tissues that consist of principally hepatocytes incorporates a good diversity of high content organic chemistry response, and also the deduction and breakdown of little sophisticated particles, several are unit essential for crucial functions.

1.3.1 STRUCTURE OF LIVER

The liver is chromatic, has a rubbery texture and weighs around 1.4 to 1.6 kg. It is located towards the left side of the abdomen and below the diaphragm. The sole organ bigger and heavier than the liver in human body is skin. The shape of liver is triangular and it has the lobes: a smaller left lobe and comparatively smaller left lobe. A ligament named coronary ligament separated the two lobes and also keeps it attached to the diaphragm.

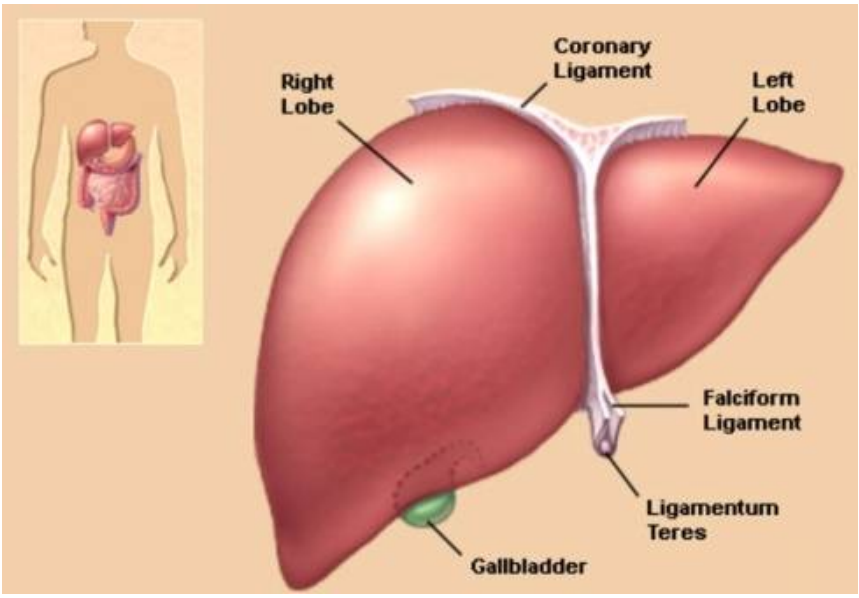


Figure 1.2: Liver Structure [4]

The artery and vena portae are the two massive blood vessels that our liver is connected to. Oxygen-rich blood is carried by the artery from the celiac complex body part, and vena portae transports blood made in digestible aliments from the complete channel also conjointly from the exocrine gland and the spleen. Liver sinusoids or liver veins are subdivided from the blood vessels that then cause subdivision.

1.3.2 FUNCTIONS OF LIVER

Liver cells divide the different functions of the liver. Liver is accountable for over five hundred different functions that are together with different organs. The liver conjointly accounts for nearly concerning two hundreds of body gas consumption.

BLOOD SUPPLY TO THE LIVER: The liver plays a very important part in our body as there are around five hundred different works that the liver is in charge of. Even with the advanced technology we have not yet created an artificial organ or organ that can copy all the functions of the human liver. Liver quantitative analysis allots some of the functions which are a treatment that is experimental in nature used in case of liver failure.

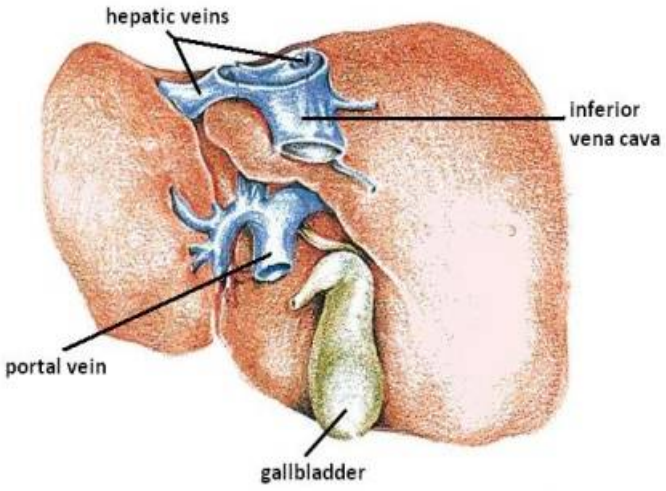


Figure 1.3: Liver Blood Supply Function [5]

The blood flows from the liver sinusoids and discharge into the mid vein of every lobe. The liver veins coalesce into the central veins that exit the liver into the inferior vein.

SYNTHESIS: Lipid metabolism, carbohydrates, proteins and amino acids are the parts where liver plays a significant role in. Liver executes many functions in carbohydrates metabolism: The liver combines and store away nearly one hundred grams of animal starch through glycogenesis and animal starch is formed. By playacting glycogenolysis which is the disintegration of carbohydrates into aldohexose, the liver than releases glucose into our blood. The synthesis of aldohexose that sure give suck, amino acids, or glycerin is called gluconeogenesis which liver is in charge of. Glycerin is manufactured by fatty and liver cells by breaking down of fat which is then used by liver for gluconeogenesis. Many roles in supermolecule metabolism are played conjointly by the liver like: sterol synthesis is performed lipogenesis, and triglycerides production, and also lipoproteins that are in bulk are synthesized within the liver. Digestion is mainly carried out by liver, because of the

production and excretion of gall which is a xanthous liquid needed for emulsion of fat. From the diet, it also conjointly facilitate within the concentration of vitamin K.

BREAKDOWN: One among foremost vital operation of the Liver is that it's chargeable for alternative hormones and mainly for the disintegration of insulin. Through glucuronidation the liver breakdowns bilirubin, this then helps its secretion into gall. The breakdown and secretion of the much waste merchandise is also done by liver. A key role is played by liver in modifying or disintegrating cytotoxic substances such as methylation also most healthful merchandise in a very method referred to as drug metabolism. This generally ends up in toxication, once the metabolic activities is additional cytotoxic than its predecessor. The toxins are conjugated for excretion to happen in urine or bile which is preferable. The conversion of ammonia into urea is a part of the urea cycle that is converted by liver and therefore organic compound is releases from the urine.

OTHER FUNCTIONS: Glycogen, Vitamin D (1-4 month supply), Vitamin A, Vitamin K, Vitamin B12 (3–5 years' supply), copper and iron amongst the large number of substances stored by the liver. The liver is chargeable for immunological works also i.e. the mononuclear scavenger cell system. Albumen which is also produced by liver is the foremost teeming super molecule gift in serum.

The liver synthesizes angiotensinogen that may be an internal secretion that's chargeable for raising the pressure of human once activated by proteinase, an accelerator that is being discharged once the urinary organ senses low pressure.

1.3.3 CLINICAL SIGNIFICANCE

The only important organ that supports almost every organ in some way is liver. Liver also has the tendency to fall for many diseases because of its important location and multi functioning.

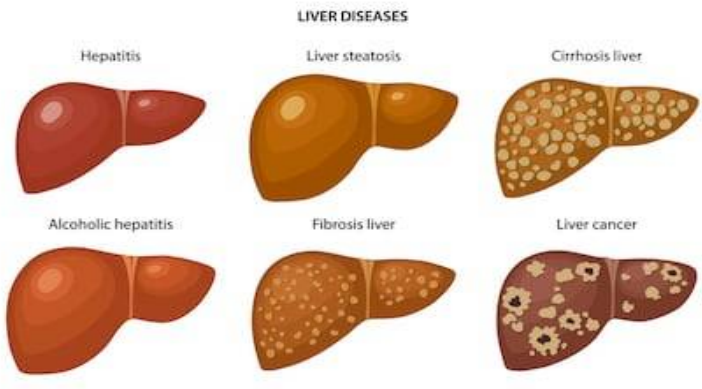


Figure 1.4: Liver Diseases [6]

Some common and hazardous diseases related to the liver are:

HEPATITIS: There are the viruses that likely cause hepatitis like hepatitis A,C,B. It is also known as the inflammation of the liver. It can be caused through noninfectious reasons also, such as allergic reactions, drugs, heavy drinking or getting extra fat.

Hepatitis A is caused because of infection with the hepatitis A virus (HAV). This type of hepatitis is usually transmitted by consuming water or food contaminated by a person infected with hepatitis A.

Hepatitis B is transmitted through contact with infectious body fluids, such as blood, vaginal excretions, or semen [7], containing the hepatitis B virus (HBV). Injection drug use, having sex with an infected partner, or sharing razors with an infected person increase your risk of having hepatitis B.

Hepatitis C is transmitted via direct contact with infected body fluids which is typically through sexual contact and injection drug use. HCV is among the most common blood borne viral infections in the country of United States.

CAUSES OF HEPATITIS:

There are few causes because of which a person may suffer from Hepatitis which includes causes like excessive use of alcohol and also because of one's immune.

a) Other toxins and alcohol

(i) Liver damage can be caused by excessive drinking of alcohol. The cells of your liver can directly get injured by consuming high amounts of alcohol. In prolonged duration, permanent injuries can be caused which are irreversible and can turn into liver failure and cirrhosis.

(ii) Over dose or over use of medications and consumption of poison are amongst other toxic reasons of hepatitis.

b) Auto-immune system response: When our own immune system errors the liver as a hostile organ, it starts attacking it. The liver's functions are also often hindered due to ongoing inflammation which can be gentle and in some cases severe too. Men are much more prone to it the women.

SYMPTOMS: There may be no symptoms in the starting if you have transmitting form of hepatitis, like hepatitis B or C which are chronic. There are certain symptoms from which you can identify whether you have the disease or not. These symptoms include:

i. Tiredness
ii. Symptoms that are like flu
iii. Dark yellow or red urine
iv. Stomach ache
v. Constipation
vi. Loss of craving
vii. Light colored stool or pale in color
viii. Light yellow eyes and yellow pale skin which indicate jaundice.
ix. Loss in weight that can't be explained

These symptoms or signs may be too difficult to notice because chronic hepatitis develops slowly.

HEPATITIS DIAGNOSIS:

If a person is suffering from hepatitis, then the disease can be diagnosed by many ways like Ultrasound, Liver Biopsy, Blood Tests and many more ways are also there such as:

i. **Physical exam:** If there is any kind of pain or tenderness while the doctor presses gently on your abdomen during a physical examination that might not be a good sign. Liver enlargement is also checked by the doctor that whether the live is the normal size or not. The doctor will also notice if you have yellow skin and eyes.

ii. **LFT:** Blood samples are used in liver function tests to determine how efficiently and effectively your liver is working. Abnormal results of these tests may be the first sign that there is a problem, especially if you don't show any signs in the physical exam of liver disease. High liver enzyme levels may show that your liver is damaged, stressed, or not functioning properly.

iii. **More blood tests:** In case of unnatural liver function tests, to detect the source of the problem the doctor will likely ask for some more tests. Hepatitis virus then can be checked by these tests.

iv. **Ultrasound:** To take a closer look at the organs and at our liver, the ultrasound test is done. The use of ultrasound waves can be seen here to create a scan of the abdomen. It can show us:

   a) Abdominal fluids
   b) Liver tumor, damaged liver or enlargement of liver
   c) Abnormalities of the gallbladder

v. **Liver biopsy:** By taking a sample of tissue from the liver the liver biopsy can be done. It doesn't require any surgery and can be done through the skin with the use of a needle. The doctor while taking a biopsy sample takes the help of an ultrasound to guide its way through. To determine if there is any infection or inflammation that has affected the liver, the doctor takes this test. The areas that appear irregular in the liver can also be tested by liver biopsy.

HEPATITIS COMPLICATIONS: Chronic hepatitis B or C will usually result in additional serious malady or health issues. People having hepatitis B or C are at risk for diseases because the virus affects the liver. These diseases include:

i. Cancer of the liver
ii. Cirrhosis
iii. Liver disease that is chronic

CIRRHOSIS: The terrible degradation of the liver and poor liver operate that can be felt or observed at the starting stages of chronic liver disease is known as the cirrhosis of the liver[8]. The scarring or degradation of the liver is mostly due to prolonged abuse of alcohol or some other viral infections.

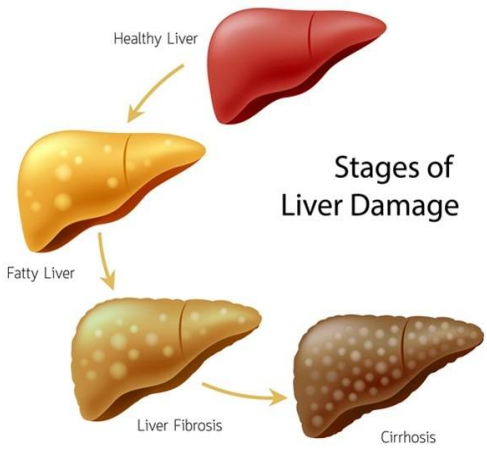


Figure 1.5: Stages of Liver Damage [9]

CAUSES OF CIRRHOSIS: The prolonged abuse of alcohol and hepatitis-c infection is amongst the main causes of cirrhosis, and some other causes of Cirrhosis are:

i. Obesity is additionally a reason behind cirrhosis of the liver, though it's not as general as alcoholism or hepatitis C.
ii. By exposure towards infected blood or infected needles and through unprotected sexual intercourse, Hepatitis-C can be contracted. Through piercings inclusive of tattooing, needle sharing, intravenous drug abuse it is possible to being open to infected blood.

There are other reasons from which cirrhosis can happen:

i. Hepatitis B: This will cause liver damage and inflammation which will result in cirrhosis of the liver.
ii. Hepatitis D: Cirrhosis of the liver can also be caused by hepatitis D. People who are already suffering from hepatitis B are more prone to it.
iii. A drug reaction sometimes causes inflammation: reaction liver disease might have a genetic cause.
iv. Damage to the digestive juice ducts that operate to empty bile: Biliary cirrhosis is one of the examples.
v. Disorders that have an effect on the ability of the body to handle some nutrients like copper and iron: 2 such instances are {|iron-storage malady |iron overload | bronzed diabetes| pathology} and disease of the Wilson's.

vi. Medicine: Medications as well as prescription and over-the-counter medicine like Tylenol, and a few other drugs, will result in liver cirrhosis.

vii. Cirrhosis Symptoms: When the liver is unable to cleanse our blood, produce clotting proteins, toxin breakdown, and assimilation of fat-soluble vitamins and fat the symptoms of cirrhosis seem to appear.

These symptoms are:

i. Bleeding from the nose

ii. Jaundice

iii. Scratchy skin

iv. Spider like arteries or veins under the skin

v. Fatigue

These are some of the more dangerous symptoms:

i. Swelling in the abdomen

ii. Difficulty processing clearly and confusion

iii. Impotence

iv. Swelled legs

DIAGNOSIS OF CIRRHOSIS: There are few tests which helps the doctor in the diagnosis of Cirrhosis such as:

i. Endoscopy

ii. MRI scan

iii. CT of the abdomen

iv. Ultrasound of the liver scan

v. Biopsy of the liver

CIRRHOSIS COMPLICATIONS: If the disease does not get cured or diagnosed then it may lead to some other diseases and complications such as:

i. Problem with consuming medicines.

ii. kidney failure

iii. liver cancer

LIVER CANCER: Primary or secondary are the two types a liver disease is usually classified. The disease that begins inside the cells of the liver is known as primary liver disease. When cancerous cells from some other adjoining organ unfold or hit the liver, the secondary stage liver disease develops [10]. In contrast to alternative cells within the body, cancer cells can separate from wherever the beginning of the cancer is with the help of the blood and the vascular system, the cells can also travel to other areas of the body. Cells that have turned cancerous eventually start growing on other organs also.

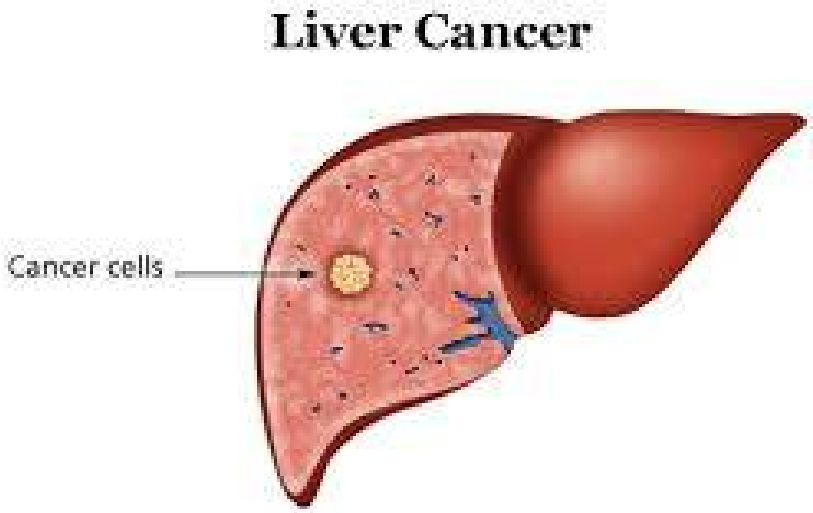


Figure 1.6: Cancer Cells [11]

The most common kind of liver disease, hepatoma, nearly always happens once cirrhosis of the liver is gift.

Hepatocellular Carcinoma

The most commonly found liver cancer is Hepatocellular carcinoma or HCC and it is also known as hepatoma. This liver cancer counts as 75% of all the liver cancers. Other body parts such as intestines, stomach, and pancreas are also prone to the spread of the cancer.

PROMINENT FEATURES OF LIVER CANCER: Most of the patients do not feel any symptoms in the start. Prominent features of liver cancer include:

i. Tenderness and abdominal pain

ii. Pale stool

iii. Jaundice

iv. Headaches or nausea

v. Constant Vomiting

vi. Bleeding or bruising easy

vii. Constant tiredness

LIVER CANCER DIAGNOSIS: With a bodily examination and medical history, the liver cancer diagnosis starts .you have to tell your doctor whether you have a prolonged abuse of alcohol or any viral infection such as hepatitis B and hepatitis C.

There are various methods, procedures and tests for liver cancer:

i. Liver Biopsy

ii. By measuring levels of liver proteins, enzymes and bilirubin in the blood, the LFT's will help the doctor to find about the health of the liver.

iii. The signs of liver cancer include the bearing of alpha-fetoprotein or AFT in the blood. When the babies are still in the yolk sac that is before they are born, this protein is normally only produced in the liver of those babies. Normally after the birth, AFT production stops.

iv. We can get imaging of the liver and other organs present in the abdomen that are very detailed with the help of MRI and CT scans. This in turn allows us to pinpoint the tumor, its current dimensions and check if it has affected other organs or not.

LIVER CANCER TREATMENT: various factors determine the treatment of liver cancer like:

i. the location, size and number of the tumors

ii. functioning of the liver

iii. If liver has turned cirrhotic

iv. If the cancer is in contact with other organs and has it spread or not

Some other Liver Diseases are:

i. **Failure of the liver:** Genetic diseases, infection, and large amounts of alcohol results in failure of the liver.

ii. **Gallstones**: Hepatitis infection and bile duct can happen if a gallstone gets stuck in the bile duct.

iii. When iron to set down in the liver, damaging it, it is known as hemo-chromatosis.

iv. **Primary Sclerosing Cholangitis (PSC):** This is a disease in which bile duct gets destroyed as it is a serious disease. Currently there is no cure for it and also the cause is not clear yet, whereas this condition is suspected have an immune response.

v. **Fatty liver disease:** This disease is mostly seen in people with bad alcohol habits and obese people. The condition is called non-alcoholic fatty liver disease if the reason is not alcohol abuse.

1.3.4 LIVER REGENRATION

The human liver can re-grow very fast as long as the liver is healthy and evolution has ensured it so because of its importance and its functions. From human beings to almost all animals this ability is seen in all the vertebrates. The automatic growing organ that can grow back on itself is only liver. If a minimum of 25 % of the tissue remains, it can recreate itself and heal fully [12]. The most noteworthy aspect of the liver is that it can re-grow to its previous ability and size reclusive of any kind of functioning loss even during its growing back. Its seen rats, that remaining liver tissue can re-grow to its original size within 4 to 6 days even if two third of the liver is removed. This process takes a bit longer in humans but the growing back of liver can be observed within 7 to 14 days which is an amazing milestone if we take in account the complexity and size of the human liver.

1.3.5 LIVER TRANSPLANTATION

When the patient is left with no other option and has liver failure which has become irreversible then only liver transplantation is considered. In case of chronic liver problems that lead to hepatitis C, cirrhosis, autoimmune hepatitis, the transplant is done in usually these cases. Lesser observed is that liver transplant is done for fulfillment of liver failure.

Usually when a donor dies from lethal brain injury [13], in that case liver allografts for transplants are available. The method where a piece of a living person's liver is cut (hepatectomy) and put-upon to replace the entire liver of the recipient is called living donor transplant. In order to serve as a liver allograft for an infant or small child only 20% of an adult’s liver is needed.

1.4 MOTIVATION

Some diseases show certain signs and symptoms while some do not. These are difficult to detect. Till the time signs are observed these become impossible to be cured. It is better to operate such diseases at an early stage rather than trying to find cures when it's been too late.

This project will help us to find certain diseases which are difficult to detect using combined results of others which shows some signs and symptoms.

1.5 SOFTWARE USED

The various software used in the implementation of our project are MATLAB, Microsoft Office Picture Manager, ImageJ, Anaconda:

1.5.1 MATLAB

Matrix Laboratory full form of MATLAB which is developed by Math Works, is a proprietary programming language and multi-paradigm numerical computing environment [14]. MATLAB allows plotting of functions and data, matrix manipulations, creation of user interfaces, implementation of algorithms, and interfacing with programs written in other languages.

Although MATLAB was first introduced for mathematical computing, another toolbox uses the MuPAD engine that gives access to symbolic computing capabilities. Simulink, an additional package, adds model-based design for dynamic and embedded systems and graphical multi-domain simulation.

1.5.2 Microsoft Office Picture Manager

MOPM was named Microsoft picture library in the start but the name was changed when a raster graphics editor was induced in 2003 until 2010. Microsoft Photo Editor was introduced in Office97 to OfficeXP was replaced by Microsoft Office Picture Manager.

It includes redaction features like crop, colour rectification, resize, rotate and flip [15]. In order to organize images, through locate pictures command, users can automatically or manually keep adding shortcuts to folders in a hierarchical file system that eliminates the need to form some class for image scans or to shift them to another location. This can be done using a shortcut introduced in Picture Manager. In order to share images from email,

directly to a share point library or to an intranet location, picture manager permits it from outlook, power point, excel and word, it allows images to be shared.

1.5.3 ImageJ

Java based imaging program named as imageJ was developed by NIH and the Laboratory for optical and computational instruments (University of Wisconsin, LOCI). ImageJ 1x is the first version that is first programmed in the domain in public. The projects and ImageJ2 like ImgLib2, SciJava and SCIFIO related to ImageJ2 are BSD-2 licensed and permissive. Here this program was created with the help of an architecture that supplies malleability with the help of registrable macros [16]. Java compiler and ImageJ's built-in editor can be used to develop analysis, custom acquisition and processing plugins. Many image processing and analysis problems can be solved using user-written plug-ins, from radiological image processing to, 3-D live-cell imaging to, automated hematology to, systems multiple imaging and system data comparisons.

With Java 5 or later virtual machine, ImageJ can be run as a downloadable app, on any computer or as an online applet. We can download it from mac OS, Windows, Sharp Zaurus PDA, Mac OS and Linux. There is a free source code available for ImageJ.

1.5.4 Anaconda

The software that targets to clarify package deployment and management is Anaconda which is a distribution that is of open source also software that is free and is of the R program and python coding for computing such as sciences related to data, predictive analytics, etc. Languages such as MAC OS, windows and Linux are suitable for this open source distribution. Travis Oliphant, Peter Wang discovered Anaconda, Inc. in 2012 which maintains it. Conda a package management system manages the package versions in Anaconda. As it was useful for its own that's why this package manager was boundless as a different open source package. Miniconda is the small version of it that includes conda, also only a small number of packages that are different are depends on python.

1.6 OUTLINE OF REPORT

Till now, in the first chapter we read about the liver and its various characteristics such as its structure, functions, diseases and many more. Also we learnt about the different software's used in the implementation of our project. In the upcoming second chapter, we have mentioned various research papers that we have studied during the making of our project. We tried to gather most of the information from these papers. Further chapter three includes all the information about every step involved in the whole process of liver cancer detection and classification. In this chapter we learned about image fusion, various techniques of image fusion, feature extraction, surf features and finally about the classification using machine learning algorithms. The last chapter, chapter four covers all our results which we got after the implementation. It shows the different datasets and the techniques applied on them to get the classification results. Finally the conclusion and future scope is mentioned.

CHAPTER 2
LITERATURE REVIEW

In this chapter different research papers are studied on the different topics related to our project.

2.1 RESEARCH PAPERS

1. Sorensen, K *et.al*. (2017): In this research paper there are five different periods in which the history of metaheuristics is defined. We have clearly learned how we can clear up the development in this work .if we do not focus on papers, but by trying to find the crucial change of events that this work has deal with.

2. James, A.P. *et.al.* (2014): In this research paper we learned that with the help of a single or many image modalities in order to increase the image quality and decrease randomness, image fusion is the method where these images are registered or combined. By the help of this paper we got to know that image fusion has helped us greatly in the understanding of medical image fusion [19].

3. H.M. Taylor, P.R. Ros (1998): Our understanding from this research paper is that for optimal detection and characterization of focal or diffuse liver disease, it is very important to obtain the most appropriate imaging test in the correct clinical setting. Also, access to clinical information and medical history is, therefore, essential. Moreover, the understanding with currently available modalities for imaging the liver allows the best use of the technical advances in ultrasound imaging, CT scanning, MR imaging technology and contributes to improved diagnostic accuracy [20].

4. Dutta At., Dubey A. (2019): This paper helped us detect liver cancer by using image processing technique. Also we learned that image processing is a technique used for processing of images with the help of mathematical operations. It uses any of the form of signal processing. Let’s say if the input is an image or video then the output is also an image or a set of image. This technique is also used in medical applications for various detection and treatment. In this paper, it has been used to detect cancer cell of the liver. Here ostu's method is used for enhancing the MRI image and watershed method is used to segment the cancer cell from the image [21].

5. Kunio Doi, (2008): In this research paper we learn that Computer-aided diagnosis (CAD) has become one of the major research subjects in medical imaging and diagnostic radiology. The motivation and philosophy for early development of CAD schemes are presented together with the current status and future potential of CAD in a PACS environment. With CAD, radiologists use the computer output as a "second opinion" and make the final decisions. CAD is a concept established by taking into account equally the roles of physicians and computers, whereas automated computer diagnosis is a concept based on computer algorithms only. So this paper helps us understand computer aided diagnosis in medical imaging [22].

6. Preeti Sharma, Tapan Jain (2014): In this research paper we learned about the Discrete Wavelet Transform technique. In this paper, hybrid watermarking scheme using SVD and DWT were introduced, where the watermark was embedded in the singular values of the red component of the cover image's DWT sub-bands and then combined with the other two i.e. green and blue components to yield the watermarked image. The experimental results which we got depicted the improved imperceptibility and robustness under attacks and preserved copyrights by using this technique [23].

7. S. Jain, M. Sachdeva, *et.al* (2019): This research paper helped us to learn about one of the image fusion technique i.e. Intensity Hue Saturation. We learned that for sharpening, IHS is the most commonly used fusion technique. It has become a standard procedure in image analysis for colour enhancement, feature enhancement, improvement of spatial resolution and the fusion of disparate data sets [24].

8. K. V. Mardia (2014): This paper helped us to know that by extending specific studies on robustness of the t statistic, the measures of kurtosis and skewness are developed. Desirable functions are shown present by these measures. A multidimensional normal population is used for the derivation of asymptotic distributions of the measures for samples. A test is proposed of multivariate normality. An event of non-normality on

measure a single sample Hotelling's T* test is studied through empirical observation by the use of these works therefore it is discovered that Hotelling's T* test is more sensible to the measure of skewness than to the measure of kurtosis.

9. H. Bay, T. Tuytelaars, et.al. (2006): In this paper, we a novel scale- and rotation-invariant interest point detector and descriptor, coined SURF (Speeded Up Robust Features) is presented. It approximates or even outperforms previously proposed schemes with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster. We basically learned that SURF is used to get the points of interest in an image [26].

10. Robert M. Haralick , K. Shanmugam , et.al. (1973): This research paper helped us to know that the texture is one of the important characteristics used in identifying objects or regions of interest in an image, whether the image be a photomicrograph, an aerial photograph, or a satellite image. It describes some easily computable textural features based on gray-tone spatial dependancies, and illustrates their application in category-identification tasks. These results indicate that the easily computable textural features probably have a general applicability for a wide variety of image-classification applications [27].

11. J.S. Raikwal, Kanak Saxena, (2012): In this paper, after implementation we found that K-NN is a quite a good classifier but when we apply this algorithm over textual data (Nominal data) it’s all performance parameter varies according to the size of dataset. K-NN performs poor results as the size of data set increases. SVM is a complex classifier and here we implement leaner kernel. We found that the accuracy and other performance parameters are not too much dependent over dataset size. It is a best fit classifier for text mining (contain mining) [28].

CHAPTER 3

METHODOLOGY

Various image fusion techniques can be used for the image fusion. Here, different image modalities are fused and feature extraction is done, so that, on the basis of these features, classification can be done. Surf features are extracted to know the interest points. In this chapter, there is detailed information about every topic used in the project.

The following flow chart in the figure 3.1 shows the various steps involved in the classification.

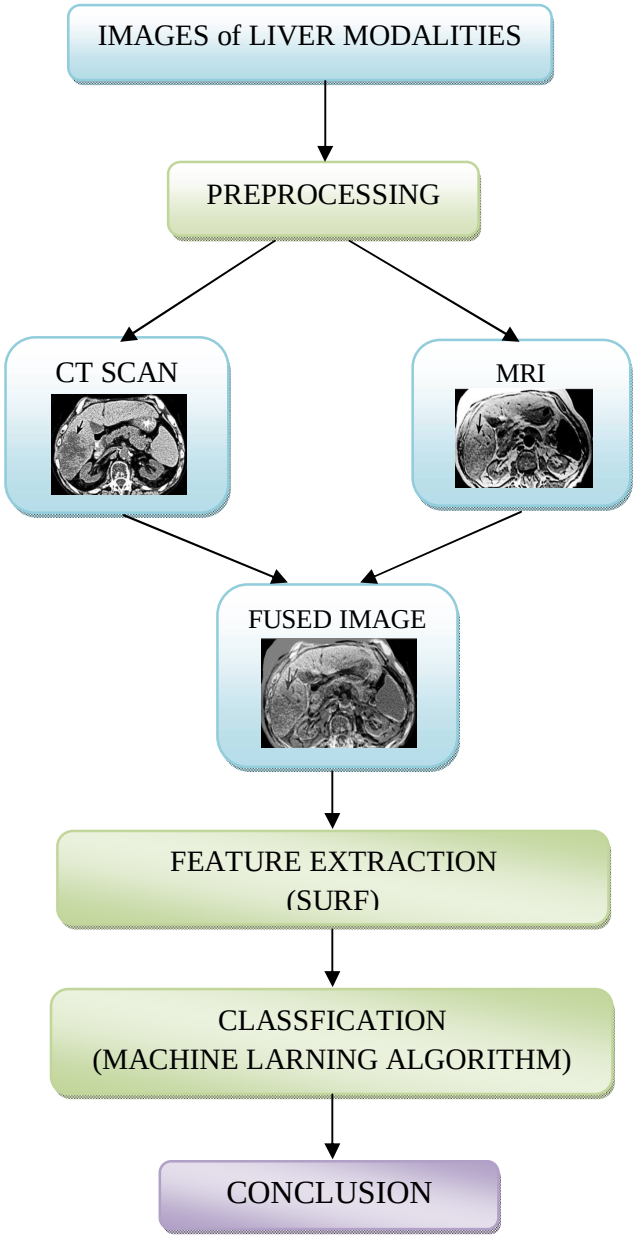


Figure 3.1: Methodology Flow Diagram

3.1 IMAGE MODALITIES

3.1.1 CT SCAN

For assessing the liver and the structures related for lesions and tumours, bleeding, infections, un-explained abdominal pain, or other conditions, specifically when other type of examinations, such as physical examination, X-rays, and ultrasound is not decisive, Computed Tomography of the liver is done.

CT scan is used to even differentiate between the obstructive and non-obstructive jaundice of the liver. It is also used for providing guidance for aspiration of tissue from the gall bladder or liver.

3.1.2 MRI

In diagnosing focal hepatic masses, Magnetic resonance imaging (MRI) is mostly used over ultrasound, Computed Tomography (CT) or any other imaging modality. At an early stage of the liver tumour, it is crucial to spot and stage the tumour (to select patients who will benefit from curative liver resection, and avoid unnecessary surgery). For an optimal preoperative evaluation of the liver, contrast-enhanced MRI is generally considered the state-of-the-art method. Though liver MRI without contrast administration is not sufficient for most liver tumour diagnoses, but still it’s appropriate for cholelithiasis.

3.1.3 COMPUTER AIDED DESIGN

To aid in the creation, optimization, modification or analysis of a design, the use of computers (or workstations), Computer-aided design (CAD) is used [29]. It is used to make better the quality of design, increase the productivity of the designer, also improve the communications to create a database for manufacturing. For manufacturing operations, CAD software output is usually in the form of electronic files

3.2 IMAGE FUSION

The collection of all the useful details from multiple images into a single one is known as image fusion. That single image which we get after image fusion is more detailed and even more accurate than any single source image [30]. Image fusion is not only used to lessen the information but to also construct images that are more understandable and appropriate for the machine and human perception.

In computer vision, the process of combining all to the point details from two or more than two images into a single image is known as Multi-sensor Image fusion. It is done to get a more detailed and informative image than any of the input images.

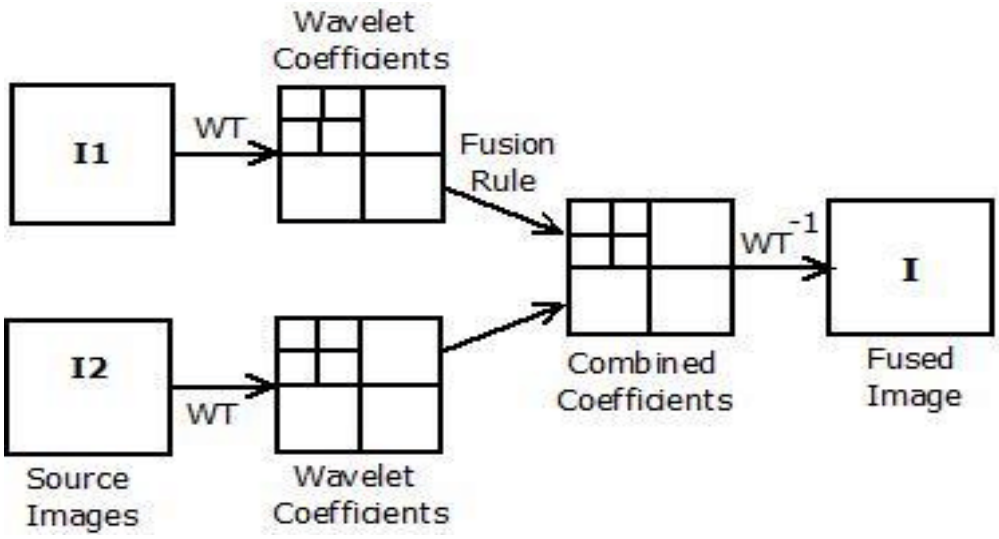


Figure 3.2: Image Fusion Example [31]

Some of the various methods and techniques of image fusion discussed here are DWT, IHS and PCA [32].

a) **Discrete Wavelet Transform (DWT):** "Time – frequency localization is allowed by Wavelet transforms". As the word wavelet means "small wave" therefore it's analysis is used to evaluate the signals with finite energy functions whose duration is really short. So basically the signal is changed into a representation presenting it in a better form. "Continuous wavelet transform is the transform in which the scale and position is varied very smoothly whereas discrete wavelet transform is when the scale and position are changed in discrete steps."

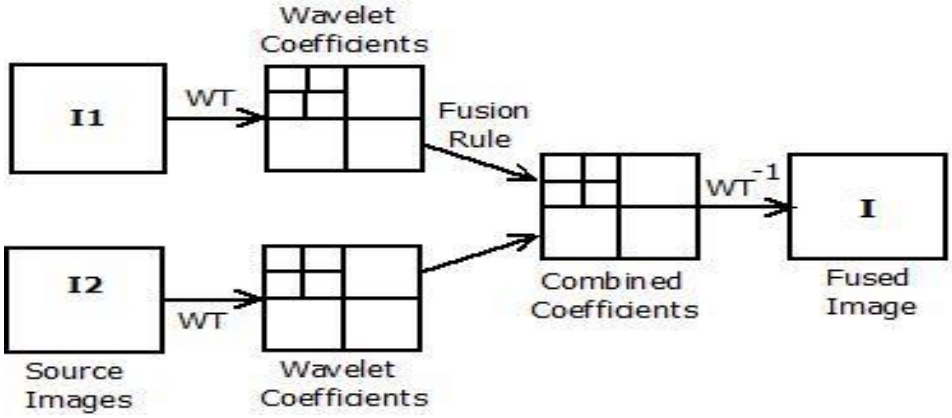


Figure 3.3: DWT Block Diagram [33]

There are basically three categories of image fusion-

(i) Feature level

(ii) Decision level

(iii) Pixel level

In pixel level, the images used for fusion hold the new measured variables, also the algorithms are really efficient and are easy to implement, and therefore mostly the image fusion applications that are used are pixel level based.

b) **Principal Component Analysis (PCA):** When the number of non-mutually related variables changed from mutually related variables such a procedure is called PCA. It figures out the small, optimum nature of the code. The variance that it left accounts for every other component and also the 1st part accounts for as much as of the variance in the data as possible. Applications where huge amount of data is to be analysed, there PCA is implemented which is also a linear transformation. In a way to show the same and different things without much of data loss PCA comes into play in case of pattern matching and data firmness.

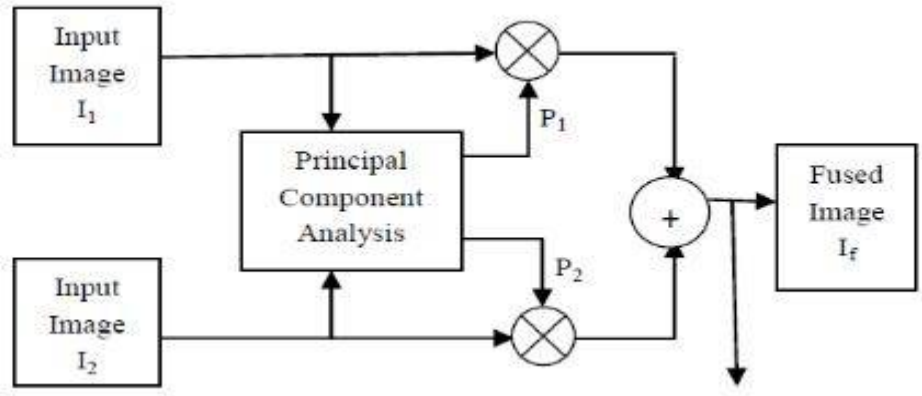


Figure 3.4: PCA Block Diagram showing $P_1I_1+P_2I_2$ [34]

c) **Intensity Hue Saturation (IHS):** This formula is a simple method that is used in image fusion. True colour space RGB is what it is based on. For sharpening, IHS is the most commonly used fusion technique. It has become a standard procedure in image analysis for colour enhancement, feature enhancement, improvement of spatial resolution and the fusion of disparate data sets. In the IHS space, spectral information is mostly reflected on the hue and the saturation [35]. From the visual system, one can conclude that the intensity change has little effect on the spectral information and is easy to deal with. For the fusion of the high-resolution and multispectral remote

sensing images, the goal is ensuring the spectral information and adding the detail information of high spatial resolution, therefore, the fusion is even more adequate for treatment in IHS space.

3.3 FEATURE EXTRACTION

Feature extraction is referred to as the method where like a proportion pruning which creatively shows compelling features of the image as a dense feature vector. When the size of image are too big also, feature representation that is dense is needed to quickly finish the tasks such as repossession and matching the image, here extraction of features is useful. For solving similar problems related to computer vision such as recognition and object detection, face detection, image retrieval based on content and recognition, the feature's detection, extraction, and matching are often joined together.

Types of Distribution in extraction of features [36]:

i. Symmetric Distributions
ii. Non- Symmetric Distributions

Symmetric Distributions: The distribution is symmetric when the probability or relative frequency is similar at same distances from the point of symmetry. Mean is the centre of distribution that is the point of symmetry. Point of symmetry can be referred to as α. In order to represent that the point of distribution is symmetric at the mean, the median and mean are equal and occur at alpha. So mode also matches to the median and mean when the dispersion is symmetrical and uni-modal.

Normal distribution is amongst the very common symmetric distributions. Symmetric are also a lot of other distributions. This article's very good to understand all types of distributions that are symmetric. In order to see how symmetry is lost we should focus on normal distributions.

Non-Symmetric Distributions: Non-symmetric distributions consist of kurtosis and skewness.

SKEWNESS: Amongst the main characteristic of a normal distribution, skewness is extend of deviation or distortion from the same curve that is an important thing. In order to calculate lack of symmetry in data distribution skewness can be used. In case of one two tails that are against each other helps us find the extreme values. The data with skewness of 0 have a symmetrical distribution [37]. There are two types of Skewness: Positive and Negative. When the tail on the left side of the distribution is smaller or slimmer than the tail on the right side then it is known as positive skewness. Here the mode is less than the median and the mean. Also when the tail on the left side of the distribution is longer or fatter than the tail on the right side it is known as negative skewness. And here the mode will be greater than the mean and the median. In figure 3.5, this behaviour is shown.

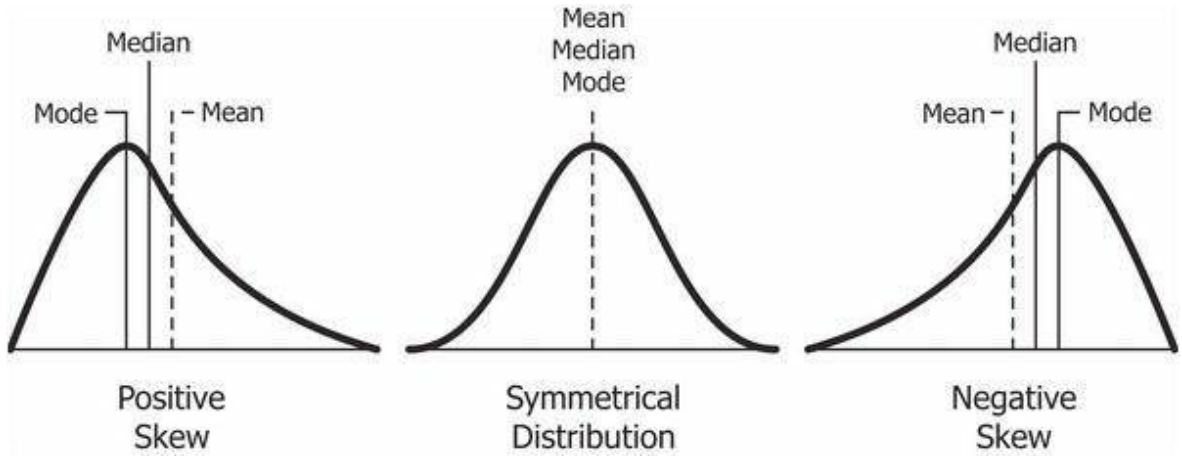


Figure 3.5: Behaviour of Skewness [38]

Here it is shown that mode and the median are not the same as mean and also bring in some answers for analysis of data. Such distributions can be dealt later on.

In case for uni-variate data the skewness can be measured as X_1, X_2,..., X_N, is the formula for skewness:

$$Skewness = \frac{\frac{\sum_{i=1}^{N}(X_i - \overline{X})^3}{N}}{s^3} \qquad (1)$$

Where N is the number of data points, $\overline{X}$ is the mean and s is the standard deviation. In case of calculating the skewness only.

Fisher-Pearson coefficient of skewness is the name for equation (1). The most commonly used way to calculate skewness is this only.

Thumb rule refers to:

i. If the value of skewness lies in between -0.5 and 0.5 than the data will probably be symmetric.
ii. Data will be moderately positioned when the value will lie in-between -1 and -0.5 or between 0.5 and 1.
iii. The data is highly positioned if the skewness is lesser than -1 or greater than 1.

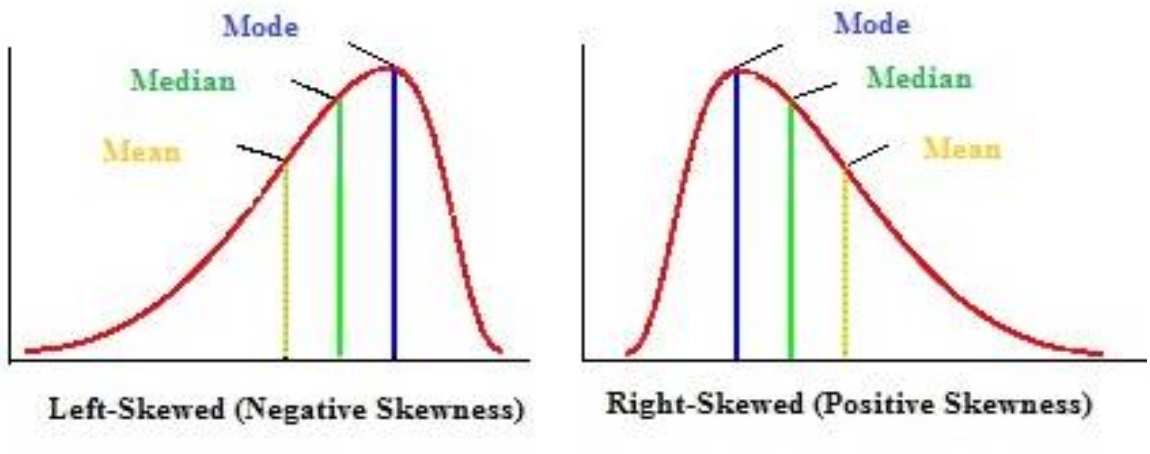


Figure 3.6: Level of Skewness [39]

KURTOSIS: It is not really the peaked-ness but kurtosis handles the length of tails in the distribution. In order to describe the ultimate values in 1 tail versus the 2nd tail , kurtosis is used. It is in reality present in the distributions and is the rate of outliers.

Measuring Kurtosis:

In case of uni-variate dataset: $Y_1, Y_2, \ldots, Y_N$, kurtosis's method is:

$$Kurtosis = \frac{\frac{\sum_{i=1}^{N}(Yi-\overline{Y})}{N}}{s^4} \qquad (2)$$

When the kurtosis is high, then, it is needed to find out that how there is so much deviation from the path. Existence of errors could mean that there are outliers present. In case of

transfers that are related to banking, an outlier may show an activity that is fraudulent in nature. Also its dependence on domain shows us how to deal with the outliers. We should look over always.

Low kurtosis in the data signifies that there is a lack of outlier tails that are light. There is a need to find out and cut the dataset of results that are unrequited even if there is low kurtosis.

Bad kurtosis is how much?

There is a certain level above or below which Kurtosis is going to affect the body which is presented in Figure 3.7.

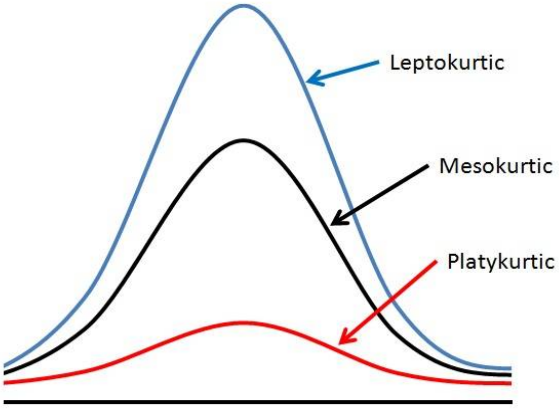


Figure 3.7: Level of Kurtosis [40]

Mesokurtic: The serious values that we have are same as the normal distribution characteristic as kurtosis static is very same to the normal distribution. We use this understanding of ours to have the distribution that is normal and standard of kurtosis as 3.

(3<kurtosis) Leptokurtic: In this we have longer distribution, fatter tails, and sharper and heighten peaks compared to the Mesokurtic. This shows that the information is very much out of deviation path. The horizontal axis of the graph is stretched across the outliers that represent most of the data that is shown in skinny reach that is it gives the skinniness as a distribution that is leptokurtic in nature.

(3<Kurtosis) Playkurtic: In comparison to the distribution that is normal we have shorter distributions and thinner backside in this. Mesokurtic is higher and slender than the peak. This means that there is an outlier deficiency. As the normal distribution is more as compared to the extreme values.

Kurtosis Characteristics:

If the curve is normally or even abnormally shaped, then the measures of kurtosis will help us to identify it. When the curve is highly arched at the mean with the short tails, then it is leptokurtic while flatter than the normal curve with a lower peak and longer tails is platykurtic.

KEY DIFFERENCES BETWEEN KURTOSIS AND SKEWNESS:

The fundamental differences between kurtosis and skewness are [41]:

i. Kurtosis is defined as the relative pointed-ness of the standard bell curve in the frequency distribution whereas skewness is defined as the characteristic of a frequency distribution that ascertains its symmetry about the mean.
ii. The measure of degree of tailed-ness is known as kurtosis in frequency distribution while the measure of the degree of lop-sidedness is known as skewness.
iii. Kurtosis tells us that whether the data we have is either flat or peaked with respect to the probability distribution and on the other side skewness tells us about the lack of symmetry, i.e. both left and right sides of the curve are unequal, with respect to the central point.
iv. Also kurtosis explains that how much sharp and tall the central peak is while skewness tells us how much and in which direction, the values deviate from the mean.

3.3.1 TYPES OF FEATURE EXTRACTION

Some of the mostly used methods to extract features from the images are as follows [42]:

1. Transform features
2. Spatial features
3. Colour features
4. Edge and boundary features
5. Shape features
6. Texture features

1. Transform Features

Frequency domain information of data is obtained from transformation of an image. Zonal filtering is used for extracting the transform features of an image which is known as the feature mask. Here, feature mask is an aperture or a slit. High frequency components are usually used for edge and boundary detection and angular slits are used for orientation detection.

2. Spatial Features

Grey level, spatial distribution and amplitude characterize the spatial features of an object. One of the most important and simplest features of the object is amplitude. It enables discrimination of bones from tissues and also shows the absorption characteristics of the body masses in X-ray images.

3. Colour Features

Object things visual attribute is colour which emerges from the light transmitted or reflected or emitted. The extension to vector-signals from scalar-signals of the colour signals is from a mathematical viewpoint. Histogram of the image can be used to get the colour features of an image. There is one disadvantage of colour histogram and that is, we can get equal colour histogram of various things if they have the same combination of colours. According to Platt and Goetz (2004), the colour features are still very useful for many biomedical imaging applications such as content-based image retrieval (CBIR) systems and cancer cell detection and many more. For the computation of colour histogram every image is added to the collection and is analysed in CBIR. During the search time, the user can either submit complementary images from which a histogram is to be computed or can specify the desired proportion of each colour. Either way, we get the result as those images whose colour histograms are matched with those of the query very closely.

4. Edge and Boundary Features

The areas in an image with strong intensity contrast are known as edges. A great variation in the picture quality can be created by a jump in the intensity from one-pixel level to the next. So edge detection basically preserves important information of an image while filtering out the unimportant information thereby reducing the amount of data. An edge doesn't have width. It is scale-dependent and contains other edges at a certain scale. The common properties of an object such as shape, area and perimeter can be measured easily only if the image's edges are identified accurately. Hence, for boundary segmentation and estimation in the scene, edges are used.

5. Shape Features

The profile and physical structure is referred as shape of an object. Shape features are most commonly used for making measurement of shapes, recognizing objects and finding and matching shapes. Perimeter, area, orientation and moment are some of the processes used for computing shape feature extraction. External boundary abstracting from other properties such as content, colour and material composition, as well as from the object's other spatial properties is used to determine the shape of an object.

6. Texture Features

In 2012, Guiying Li, a Chinese scholar defined that texture is a recurring pattern of information or data or arrangements of the structure a regular interval. From a universal perspective, texture alludes to surface qualities and presence of properties which given by shape, thickness, size, course of action, extent of its rudimentary parts. Feature Extraction is known as the stage to collect all such features through texture analysis process. Because of the implication of surface data, surface element extraction is a key function in different image processing applications like remote detecting, clinical imaging and substance based picture recovery.

3.3.2 STATISTICAL BASED FEATURES [43]

First order statistics, Second order statistics and higher order statistics are basically three different types of Statistical based features:

First Order Histogram based Features

First Order histogram gives different factual highlights, for example, four measurable snapshots of the quality histogram of a picture. These rely just upon solitary pixel esteems and not on the affiliation or co-occasion of neighbouring pixel esteems. Mean, variance, skewness and kurtosis are some features based on Ist Order Histogram

1. Mean (m): Mean value describes the mean intensity value and it likewise gives the commitment of individual pixel intensity for the whole picture and variance is regularly used to discover how every pixel shifts from the neighboring pixel (or focus pixel).

$$Mean = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}\left(\mathrm{TR_{i,j}}\right)}{\mathrm{M} * \mathrm{N}} \qquad (3)$$

2. Standard Deviation: It is a broadly utilized for the estimation of changeability or decent variety utilized in insights. It also shows the variation or dispersion exists in the input image, in terms Image Processing.

$$Standard\ Deviation = \sqrt{\frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(\mathrm{TR_{i,j}}-\mathrm{mean})^3}{\mathrm{M}*\mathrm{N}}} \qquad (4)$$

3. Skewness: It measures the asymmetry of the probability distribution of a random variable that is real valued.

$$\mathrm{Skewness} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(\mathrm{TR_{i,j}}-\mathrm{mean})^3}{\mathrm{M}*\mathrm{N}*\mathrm{SD}} \qquad (5)$$

4. Kurtosis: It measures of the probability distribution's shape of a random variable that is real valued.

$$\mathrm{Kurtosis} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(\mathrm{TR_{i,j}}-\mathrm{mean})^4}{\mathrm{M}*\mathrm{N}*\mathrm{SD}^4} - 3 \qquad (6)$$

3.3.3 HARALICK TEXTURE FEATURES [44]

Haralick extracted thirteen texture features from Gray-Level Co-Occurrence Matrix for an image. Energy, Entropy, Inverse Difference Moment, Contrast and Directional Moment are the important features for classifying the image into water body and non-water body. In the year 1995, Flavio Parmiggiani and Andrea Baraldi have discussed five statistical parameters which are energy, contrast, entropy, Inverse Difference Moment and Directional Moment, which are considered the most relevant among the texture features which were proposed by Haralick in 1973.

Let a and b are the coefficients of co-occurrence matrix, $M_{a,b}$ is the element in the co-occurrence matrix with a and b the coordinates and the dimension of the co-occurrence matrix is N.

Entropy: It is measure of the randomness of the elements of the image and also of image information content, which is explained as the average uncertainty of information source. Entropy in any image is also defined as corresponding states of intensity level which individual pixels can adapt.

$$\mathrm{Entropy} = \frac{\sum_{a=0}^{M}\sum_{b=0}^{N}(\mathrm{TR_{a,b}})^2}{\mathrm{M}*\mathrm{N}} \qquad (7)$$

Energy: It measures the uniformity of any image as well as it also measure the extent of pixel pair repetitions. Energy value will be higher whenever the pixels in the image are similar. Energy is a relative term in image processing. It is used in typical object segmentation or detection tasks where it is constituted as an Energy minimization problem.

One can characterize energy as a capacity that would catch the arrangement we want and perform gradient-descent drop to register its least worth, bringing about a solution for the picture division. In optimization, the energy is commonly used as a synonym “Measure of Error” or for "distance from the ideal model ".

$$\text{Energy} = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} M^2\,(i,j)} \qquad (8)$$

SURF (Speeded-Up Robust Features):

For assessment of images, local and similarity invariant representation, SURF method is used as it is a robust and fast algorithm. Actual applications like object tracking and recognition through box filters which is being used by the swift computation of operators and it is also the prime emphasis of SURF technique [45]. The provocation of locating point resemblances between two images with similar object or scene is a part of various computer vision applications. Few applications are image retrieval and registration, object recognition and camera calibration retrieval.

SURF is made up of two steps, initial step comprises of Feature Extraction which is followed by Feature Description.

1. Feature Extraction

Interest points from an image are detected with the use of Hessian matrix approximation

Integral images

In 1984, the idea of Summed-Area Table or we can say of Integral Image was introduced. Summation of values (pixel values) in an image or a rectangular subset of a grid has been made easy because of this and also was considered as an effective and a fast method. So basically, it was used to calculate the average intensity within a given image.

$$I_{\sum}(i) = \sum_{a=0}^{a \le i} \sum_{b=0}^{b \le j} I(a,b) \qquad (9)$$

This allowed us for high speed computation in box type convolution filters. The entry of an integral image $I_{\sum}(x)$ at a location $x = (x, y)^T$ defines the sum of all the pixels in the input image I within a rectangular area formed by the origin and x.

We require four additions for I_{Σ} to be calculated, for the calculation of the sum of the intensities over any upright, rectangular sections.

Hessian matrix-based interest points

Because of hessian matrix's accuracy rate and good performance in computation time, it has a great applicability in SURF. Therefore more than using any different quantity, SURF depends on the computation of the determinant of the Hessian matrix for choosing its scale (Hessian-Laplace detector) and location. When a pixel is given to us then its hessian matrix is like:

$$H(i,\sigma)=\begin{bmatrix} L_{ii}(i,\sigma) & L_{ij}(i,\sigma) \\ L_{ij}(i,\sigma) & L_{jj}(i,\sigma) \end{bmatrix} \qquad (10)$$

For adaptations to any scale, image by the Gaussian kernel is filtered, so given a point X=(x,y), the Hessian matrix H(x, σ) in x at scale σ is defined as:

$$H(f(i,j)) = \begin{bmatrix} \frac{\partial^2 f}{\partial i^2} & \frac{\partial^2 f}{\partial i \partial j} \\ \frac{\partial^2 f}{\partial i \partial j} & \frac{\partial^2 f}{\partial j^2} \end{bmatrix} \qquad (11)$$

Where:

For the image I in point x, Lxx (x, σ) is the convolution of the Gaussian II-order derivative, similar for Lxy (x, σ) and Lyy (x, σ). Though being optimal in scale-space analysis, the Gaussians are required to be cropped and discretized resulting in the declination of repeatability under image rotations around the odd multiples of π/4. This powerless point holds for Hessian-based finders when all is said in done. All things considered, the locators/detectors despite everything do well, and the slight lessening in general execution doesn't exceed the advantage of quick convolutions brought by means of the discretization and cropping.

For determinant calculation of Hessian matrix, convolution is done with II-order derivative which follows the convolution with the Gaussian kernel. After Lowe's success with LoG approximations (SIFT), approximations (second-order derivative and convolution) are pushed by Surf and even

further with box filters [46]. Very low arithmetic cost used independently of size can estimate the Gaussian II-order derivative and this is the reason for the fastness of SURF.

1. Gaussian partial derivative in xy:

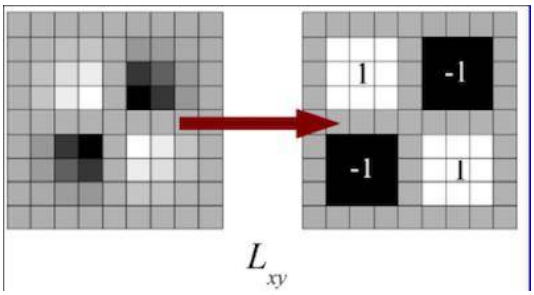


Figure 3.8: Gaussian partial derivative in xy

2. Gaussian partial derivative in yy:

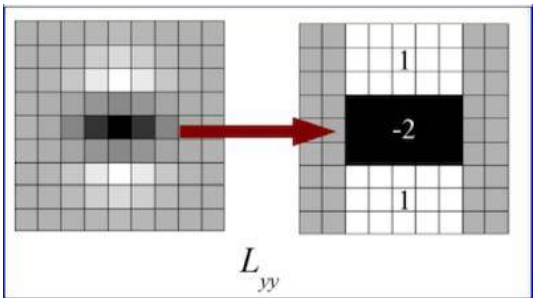


Figure 3.9: Gaussian partial derivative in yy

In the above pictorial representation, approximations for Gaussian IInd order derivative is shown by 9x9 box filters with $\sigma = 1.2$ and d_{xx}, d_{yy}, and d_{xy} are used for denoting these approximations. The determinant of the approximated Hessian can be represented by below equation:

$$\det(\mathcal{H}_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \qquad (12)$$

w=0.9 (Bay's suggestion)

Picture pyramids are used to complete scale spaces. So as to accomplish a more significant level of the pyramid, the pictures are over and again smoothed with a Gaussian and in this way sub-examined. Attributable to the utilization of box filter and fundamental pictures, surf does no longer need to iteratively apply a similar filter to the yield of a some time ago filtered layer yet then again can apply such filter of any size at the very same speed legitimately on the first picture, and even in parallel. In place of reduction of image size iteratively, the examination of scale space is done through up-scaling of the filter size such as 9×9, 15×15, 21×21, 27×27. So for every new octave, the size of the channel increments as multiplied twice at the same time, the inspecting/examining

between times for the extraction of the key focuses (σ) can be served additionally which award the up-scaling of the channel at consistent cost. In order to localize key points in the pictorial representation and over scales, a non maximum suppression in a 3 × 3 × 3 neighborhood is applied.

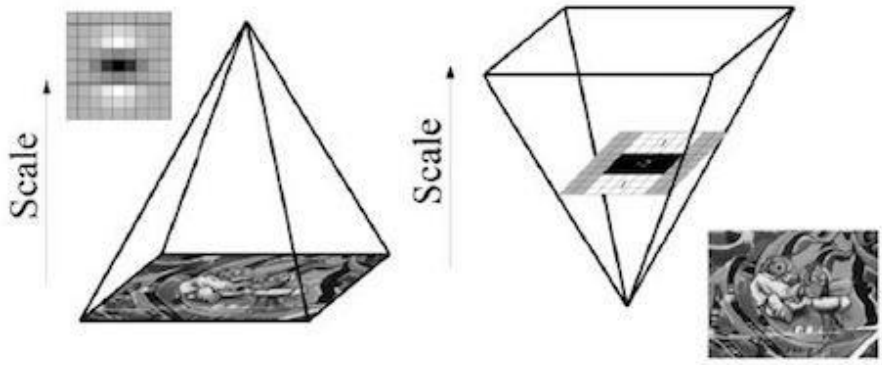


Figure 3.10: Showing Up-Scaling

Instead of reducing the image size (shown in image at left) iteratively, the use of integral images allows the up-scaling of the filter at constant cost (shown in image at right).

2. Feature Description

Two steps procedure is followed in the creation of SURF descriptor. Initial step comprises- to fix a reproducible direction dependent on data obtained by roundabout region near the key point. Next step is to build a squared territory which is in alignment with the chosen inclination and take out the SURF descriptor from it.

Orientation Assignment

SURF attempts to recognize a reproducible inclination for the key-points, so as to be invariant to rotation. To accomplish this:

1. SURF initially ascertains the Haar-wavelet reactions in y and x-directions. It is found in a round neighborhood of span 6s near the key point, where s is scale at which key point was distinguished. Also, the sampling is picked to be s as it is dependent on scale, and at present scale s, the wavelet reactions are registered. In that manner, the wavelets size is large at high scales. Consequently basic pictures are utilized again for quick filtering.

2. At that point the summation of horizontal and vertical wavelet reactions are determined in a checking territory, the scanning direction (include π/3 in that) is then changed, and

re-determined, until orientation with largest value of sum is obtained. For feature descriptor this is considered as the main orientation.

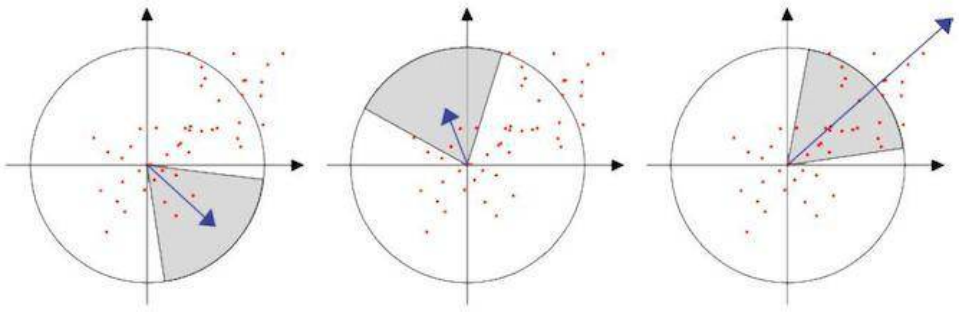


Figure 3.11: SURF Orientation

Descriptor Components

Extraction of the descriptor is done by the following procedure:

1. Initial step comprises of developing a square district revolved around the key point and situated alongside the orientation which was obtained earlier. The size is 20s for this window.

2. At that point the region is separated routinely into more compact 4 × 4 square sub-areas. Some of basic highlights are processed at 5×5 consistently divided sample focuses, for each sub-area. In vertical direction *dy* is called as the Haar wavelet response and in horizontal direction *dx* is the Haar wavelet response (filter size 2s) for the reasons of simplicity. To increase the robustness towards geometric deformations and localisation errors the responses *dy* and *dx* are first weighted with σ = 3.3s (Gaussian) with center at the key point.

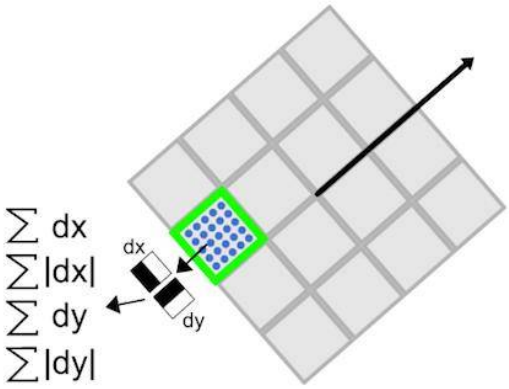


Figure 3.12: Summation of dx and dy

Then, the summation of wavelet responses *dy* and *dx* are done over each sub area and form an initial set of entries to the feature vector. So as to convey the knowledge regarding polarity of intensity shifts, we additionally extract the summation of the absolute values of the responses, |*dy*| and |*dx*|. Therefore, for its underlying intensity structure V = (∑ dx, ∑ dy, ∑|dx|, ∑|dy|), each sub-area has a 4-D descriptor vector v. So we get result as a descriptor vector for all 4×4 sub-areas of length 64. The reason why SURF is faster than SIFT is that: in SIFT, our descriptor is the 128-D vector.

3.4 CLASSIFICATION

Classification of new observations is done on the basis of the input data. It is a supervised learning approach**.** There are various machine learning algorithms that can be categorized into three categories [47]. Firstly, we have **Supervised learning** in which prediction of the result is done from a predictors set. And from these, the generation of function takes place which map the given input to the required output. Till we get the required level of accuracy of the model, the training process is continued. Some of its examples are k-NN, Decision tree, Regression etc. Second is the **Unsupervised learning** in which no prediction of the target variable takes place. In this, we split the customers into various groups for a particular intervention. One of the examples is Clustering Algorithm. The third category that we have is the **Reinforcement learning.** In this algorithm, definite decisions are made by the machine. Here, the machine plays an important role; therefore training of the machine is very important which is done by trial and error. Knowledge from past experiences is gained by the machine to give accurate results in future.

There are so many algorithms in Machine learning which are used for various purposes. Some of the algorithms are:

a) Linear Regression: It approximates the continuous variables. By establishing a best line, the dependent and independent variables establish a relationship. The linear Equation represents the best fit line known as Regression line, where Dependent variable is represented by Y, Independent Variable is represented by X, a and b represents slope and intercept respectively.

$$Y = a*X + b \qquad (13)$$

Based on minimization of the summation of square of difference of distance between the regression line and data points, the coefficients (*a,b)* are derived.

b) **Logistic Regression:** Used for the estimation of discrete values (binary values) on the basis of given set of independent variable, it is not a regression algorithm but a classification. In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function. Hence, it is also known as logit regression.

c) **Support Vector Machine (SVM):** In n-dimensional space, SVM plot every data item as a point. Here n defines number of features. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM [48] model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

d) **Decision Tree**: In this classifier, prediction is done by iterating the input data via a learning tree. So in this whole process, for reaching a specific category, the feature properties are compared relative to decision conditions. It splits the data into smaller units and the final result is represented in the form of a tree with leaf and decision nodes.

e) **Naive Bayes:** This classifier works according to the Bayes' Theorem. It assumes individuality amongst its predictors. In simple words it can be explained as, that all the features in this contribute individually for finding the probability. The features are not related in this.

f) **Clustering Algorithm**: This algorithm group the more similar set of objects in one cluster (group).Each clustering algorithm differs from the other. Some of them are: Dimensionality Reduction, Neural networks / Deep Learning, Probabilistic and Centroid-based algorithms.

g) **k- Nearest Neighbours (k-NN**): Equally for regression as well as classification problems, k-NN can be used. However, classification problems make an extensive use of k-NN. All available cases are stored by k-NN [49] and new cases are classified by a greater number of votes of its k neighbours. Measured using a distance function, the case which is assigned to the class is generally unvarying amongst its K adjoining neighbours. Examples of distance function are Hamming, Minkowski, Manhattan, or Euclidean distance. Hamming is used for categorical variables while others are utilized for continuous function.

h) **Random Forest**: As its name suggests, it is a group of decision trees. The tree "votes" for the class when classification is given by each tree to segregate a new object based on the attributes. The classification having majority votes is chosen by the forest. This is done for all forest trees.

i) **Gradient Boosting Algorithm (GBM):** When one has to work with ample data, GBM algorithm can be used for making predictions with greater prediction powers. A collection of learning algorithms is termed as boosting which for the improvement of robustness over a single estimator, merges the predictions of numerous base estimators. For building a strong predictor it merges several average or weak predictors.

CHAPTER 4
RESULTS AND DISCUSSION

We have collected different datasets of patients with liver diseases from the online platform to carry out our results. Here we will show the results of image fusion, feature extraction and how the classification is done to get the better results. In CT scan, contrast-to-noise ratio is improved. It is done between a normal liver and a liver with focal lesions. On the other hand, MRI doesn't use any ionizing radiation providing high lesion-to-liver contrast. Image fusion is not only used to lessen the information but to also construct images that are more understandable and appropriate for the machine and human perception. The prime emphasis of SURF technique lies in its swift computation of operators that uses box filters, therefore enabling the actual applications including object tracking and recognition. Classification of the surf images is done based on many machine learning classifiers such as k-NN, Naïve Bayes, SVM and many more.

4.1 Image Fusion and SURF

This section represents the results of image fusion and the feature extraction on the given datasets. We have taken eight datasets to show our results.

4.1.1 Dataset 1 of 54 years old male patient: Here the CT scan shows us a large irregular lesion whereas the MRI s confirms the lesion to be a focal fatty infiltration.

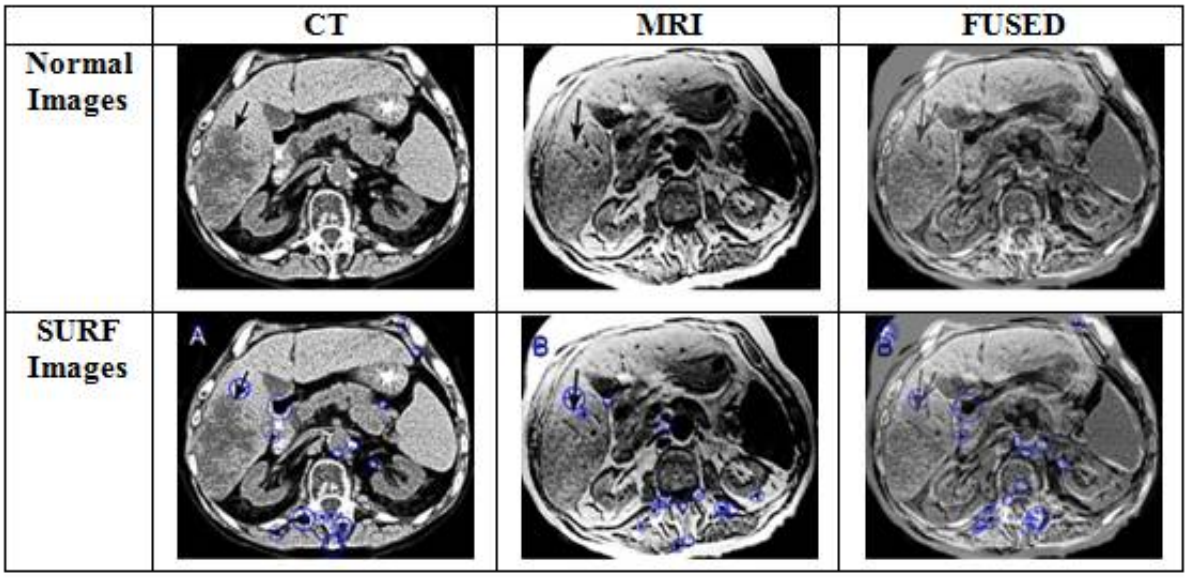


Figure 4.1: Fusion and SURF of Dataset 1 [50]

Now we did image fusion and applied SURF on the two models. A large irregular lesion is shown in the fused image which is in the right side of the liver. It is confirmed to be a focal fatty infiltration. The fused image gave us a more accurate and better result and had the details of both the images. Surf feature extraction was done to get those interest points on the basis of which we could do our classification.

Then we calculated some of the statistical based features:

Table 4.1: Parameters for Dataset 1

	CT	**MRI**	**FUSED**
SNR	12.20	12.997	11.241
PSNR	17.99	18.19	17.644
MSE	85.763	89.386	64.275
MEAN	98.926	108.071	103.715
STD. DEVIATION	85.762	89.385	64.275
VARIANCE	7355.29	7989.85	4131.36
ENTROPY	7.172	7.155	7.426
SKEWNESS	0.268	0.216	-0.006
KURTOSIS	1.652	1.627	2.327

The above table shows the various parameters calculated. For the comparison of image compression quality, we use peak signal to noise ratio and also the mean square error. Peak signal to noise ratio as the name suggests is used to calculate the peak error whereas cumulative squared error is calculated in the mean square error. With the results it is clear that the fused image has less peak error and less Mean square error as compared to other two images. The fused image shows better results than the individual images. According to the skewness values our data is fairly symmetrical and as kurtosis value is less than 3 so it is platykurtic.

4.1.2 Dataset 2 of a 45-year-old man: Here the CT scan of the patient doesn't clearly show any focal liver lesion as they are very small to be detected by CT scan. On the other hand, multiple focal enhancing liver lesions that are consistent with multifocal HCC are revealed by MRI.

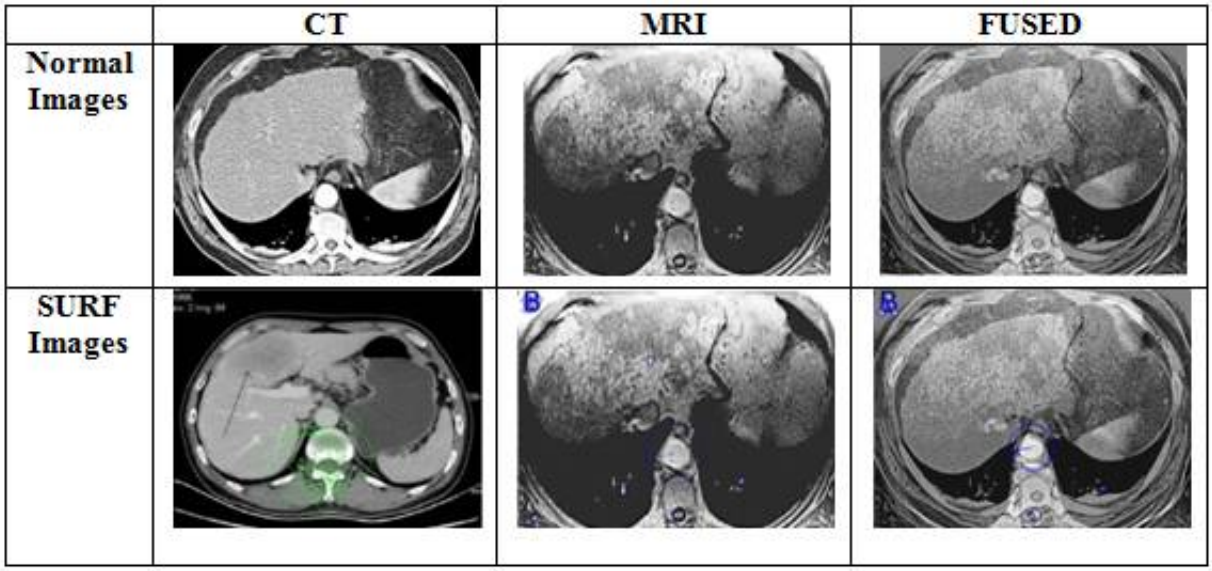


Figure 4.2: Fusion and SURF of Dataset 2 [51]

The fused image here showed us that there are multiple liver lesions with multifocal HCC. If we had not combined the two and had only relied on one of the scans, we might not have gotten the answer. Here also fused image gave us a better result and had helped us find the diseases. The blue circles here show us the surf features extracted for the classification.

Then we calculated some of the statistical based features:

Table 4.2: Parameters for Dataset 2

	CT	MRI	FUSED
SNR	12.986	12.159	11.629
PSNR	18.063	17.891	17.617
MSE	82.365	70.856	59.009
MEAN	115.939	111.13	113.56
STD. DEVIATION	82.365	70.865	59.003
VARIANCE	6784.11	5020.67	3481.43
ENTROPY	7.505	7.532	7.771
SKEWNESS	0.035	0.533	0.024
KURTOSIS	1.703	1.949	2.302

From the results of the characteristics of the images it is clear that the fused image has less peak error and less mean square error as compared to other two images of CT scan and MRI imaging and gave us the better result and better understanding of the image and the disease. According to the skewness values the CT image is fairly symmetrical, MRI image is moderately skewed whereas the fused image which we get is again fairly symmetrical and as kurtosis value is less than 3 so it is platykurtic.

4.1.3 Dataset 3 of 55 year old male patient: Single focal liver metastasis in the medial segment of left lobe is clearly shown in the CT scan of the liver. On the other hand, multiple lesions are shown by MRI in both liver lobes.

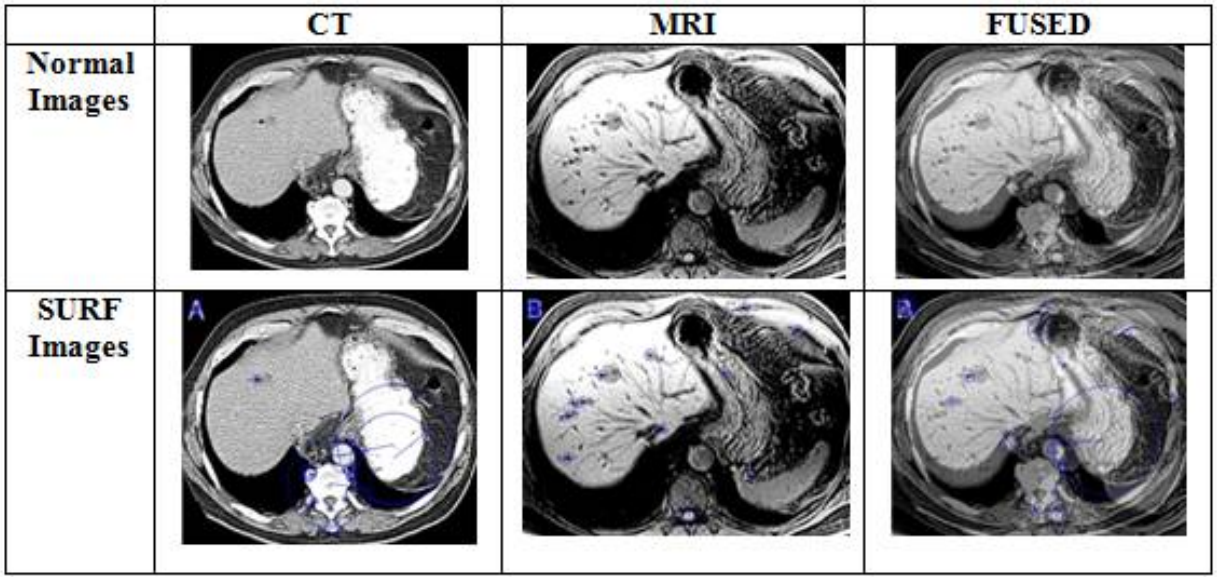


Figure 4.3: Fusion and SURF of Dataset 3[52]

Now we fused the two images using image fusion technique and received the details from both the modals i.e. from CT scan and MRI which provided information about the soft-tissue characteristics and also vascularity of the lesions which helps us determine from the scan that on the left top side part of liver we can see multiple lesions with dark spots around the veins which are not normal in a healthy liver, so at the end we can say that the fused image is more accurate and provides all the information from both the scans. The table also shows the surf images which shows our interest points or we can say the main points which will be used further for the conclusion.

Then we calculated some of the statistical based features:

Table 4.3: Parameters for Dataset 3

	CT	MRI	FUSED
SNR	13.503	12.892	12.064
PSNR	18.313	18.181	17.749
MSE	91.724	88.788	69.458
MEAN	114.317	106.261	112.849
STD. DEVIATION	91.723	88.788	69.457
VARIANCE	8413.31	7883.47	4824.40
ENTROPY	7.089	7.136	7.825
SKEWNESS	0.097	0.214	0.03
KURTOSIS	1.527	1.655	1.881

Result of image characteristics clearly shows that the fused image has less mean and peak error as compared to other two images of CT scan and MRI imaging and thus gives us better view of the images and its details by which it can be more accurate in detection of any kind of problem or disease. According to the skewness values our data is fairly symmetrical.

4.1.4 Dataset 4 of a 60-year-old man: In the hepatic left lobe segment there is a low density nodule present there. The CT image shows a clear enhancement in the arterial phase while the enhanced MRI shows a typical enhancement pattern in the liver.

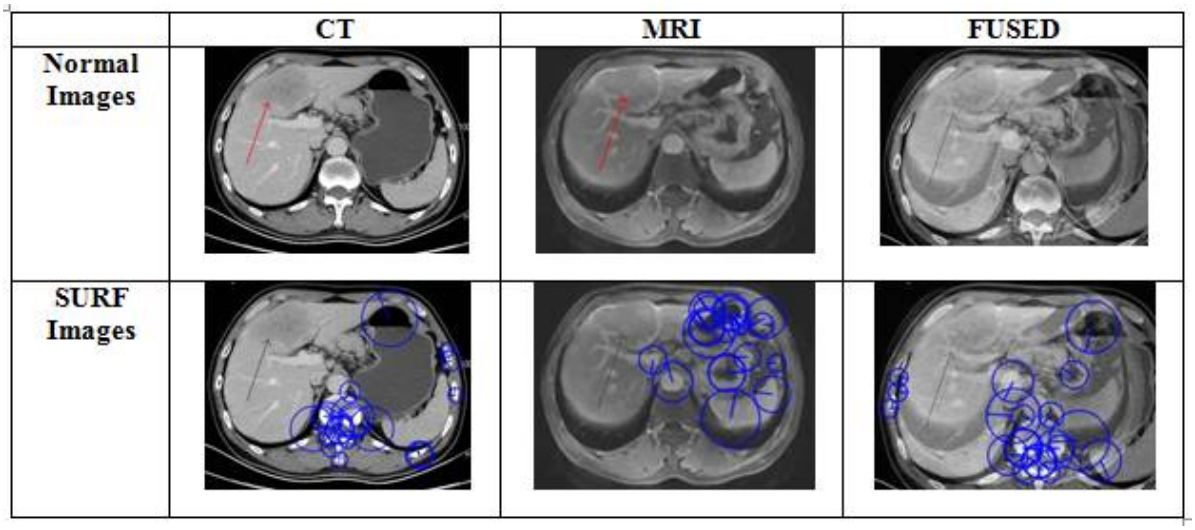


Figure 4.4: Fusion and SURF of Dataset 4

Now we fused the CT and MRI images and as a result we got the image which shows the clear enhancement. Also we got the surf images as before.

Then we calculated some of the statistical based features:

Table 4.4: Parameters for Dataset 4

	CT	MRI	FUSED
SNR	10.934	8.799	11.182
PSNR	17.614	17.429	17.494
MSE	70.575	36.068	55.506
MEAN	94.791	87.26	110.093
STD. DEVIATION	70.575	36.068	55.506
VARIANCE	4980.846	1300.946	3080.95
ENTROPY	7.019	6.709	7.513
SKEWNESS	0.010	0.081	-0.395
KURTOSIS	1.843	1.819	2.298

From the table it is clear thatthe fused image has less peak and mean square error as compared to other two images of CT scan and MRI and thus giving us better results. According to the skewness values our data is fairly symmetrical and as kurtosis value is less than 3 so it is platykurtic.

4.1.5 Dataset 5: Here we have the Preoperative CT-scan and MRI of the liver showing hyper vascularized hepatic lesions with central necrosis and associated hematoma.

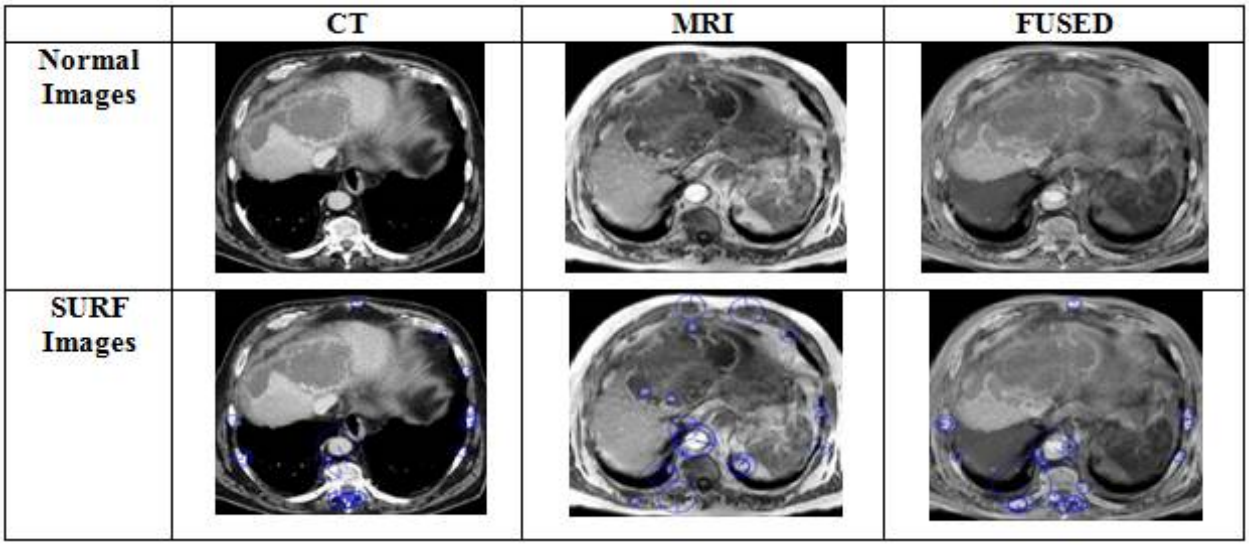


Figure 4.5: Fusion and SURF of Dataset 5

When we fuse the CT and MRI images, the resultant image that we get shows the better results as in a much more clear view of the hepatic lesions. Also the extracted surf features gives us the interest points based on which we will do the classification.

Then we calculated some of the statistical based features:

Table 4.5: Parameters for Dataset 5

	CT	MRI	FUSED
SNR	10.011	12.232	10.139
PSNR	17.694	17.84	17.467
MSE	73.528	67.517	50.308
MEAN	75.369	115.397	97.463
STD. DEVIATION	73.528	67.517	50.308
VARIANCE	5406.377	4558.667	2530.95
ENTROPY	6.639	7.50	7.193
SKEWNESS	0.615	0.129	-0.359
KURTOSIS	2.229	2.441	2.657

From the results of the characteristics of the images it is clear that the fused image has less peak error and less Mean square error as compared to other two images of CT scan and MRI imaging and gave us the better result and better understanding of the image and the disease. According to the skewness values the CT image is moderately skewed, while the MRI image and the fused image is fairly symmetrical and as kurtosis value is less than 3 so it is platykurtic.

4.1.6 Dataset 6: In the following figure the CT and MRI shows the axial view of the liver. Hepatocellular carcinoma is clearly represented through them. The MRI image shows the soft tissues of the tumor much better than the CT.

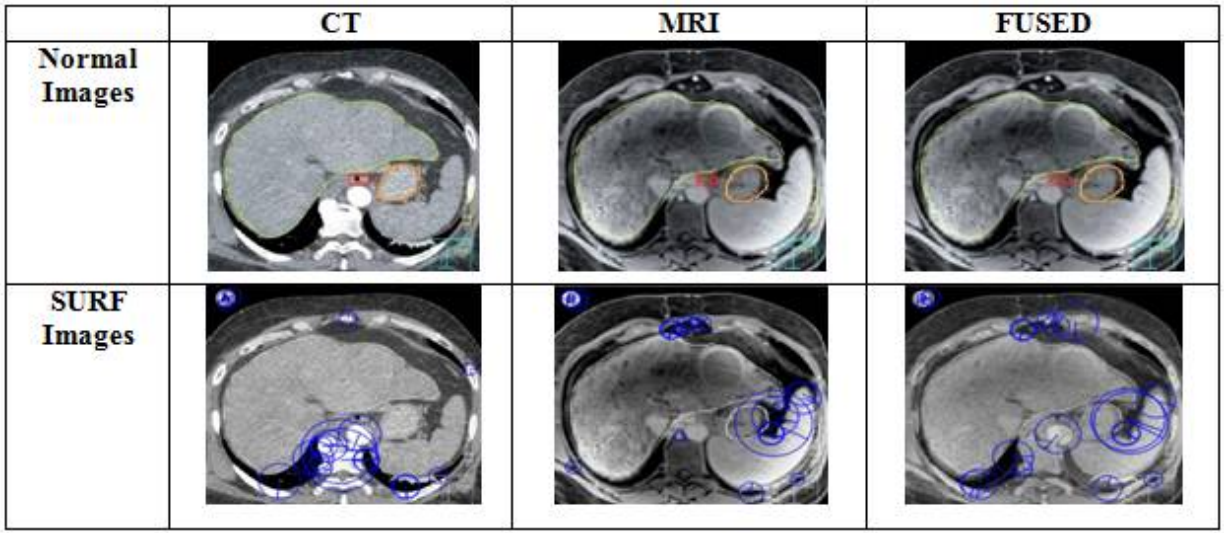


Figure 4.6: Fusion and SURF of Dataset 6

The fused image here shows the accurate results and also contains the important information of both the scans. Then we also get the surf features extracted from the CT, MRI and fused images.

Then we calculated some of the statistical based features:

Table 4.6: Parameters for Dataset 6

	CT	**MRI**	**FUSED**
SNR	12.588	10.577	11.094
PSNR	17.758	17.595	17.786
MSE	69.622	62.929	56.086
MEAN	122.167	94.679	108.522
STD. DEVIATION	69.622	62.929	56.086
VARIANCE	4847.262	3960.081	3145.70
ENTROPY	7.365	7.648	7.451
SKEWNESS	-0.283	0.29	-0.371
KURTOSIS	2.182	2.265	2.20

From the table it is clear that the fused image has given us better results Ratio as compared to other two images of CT scan and MRI. We have less PSNR (peak signal to noise ratio) and even less error (MSE). According to the skewness values our data is fairly symmetrical and as kurtosis value is less than 3 so it is platykurtic.

4.1.7 Dataset 7: Here the CT abdomen shows only the enlargement of the liver while in the MRI, we are even able to see some mild diffuse steatosis. Clearly, there is no detection of any focal liver lesions.

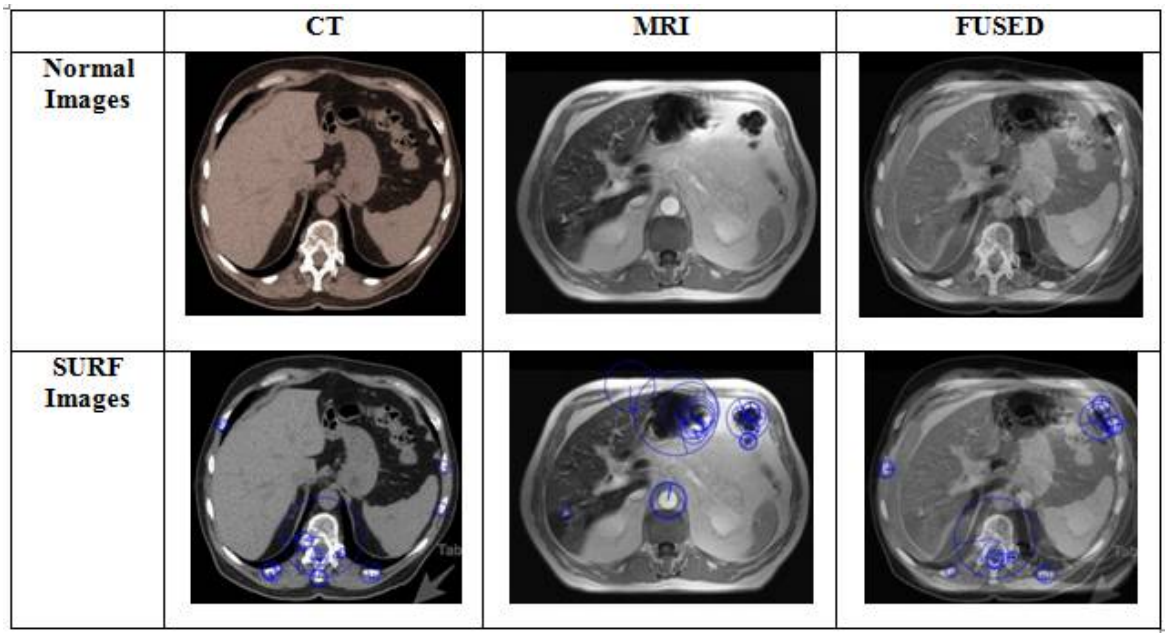


Figure 4.7: Fusion and SURF of Dataset 7

Now we fused the CT and MRI images and as a result we got the image which shows the clear enlargement. Also we got the surf images as before.

Then we calculated some of the statistical based features:

Table 4.7: Parameters for Dataset 7

	CT	MRI	FUSED
SNR	17.604	17.636	17.492
PSNR	9.012	10.209	9.189
MSE	62.906	67.141	51.582
MEAN	70.952	85.134	83.37
STD. DEVIATION	62.906	67.141	51.582
VARIANCE	3957.25	4507.99	2660.76
ENTROPY	6.76	7.011	7.185
SKEWNESS	0.568	0.421	-0.093
KURTOSIS	2.609	2.176	2.147

From the results of the characteristics of the images it is clear that the fused image has less peak error and less Mean square error as compared to other two images of CT scan and MRI imaging and gave us the better result and better understanding of the image and the disease. According to the skewness values the CT image is moderately skewed, while the MRI image and the fused image is fairly symmetrical and as kurtosis value is less than 3 so it is platykurtic.

4.1.8 Dataset 8: Multiple liver lesions of a patient are shown here through the CT and MRI images. The CT image here is not able to show the anterior, smaller lesion properly while the MRI image clearly shows the small tumor which is also marked here with a red arrow.

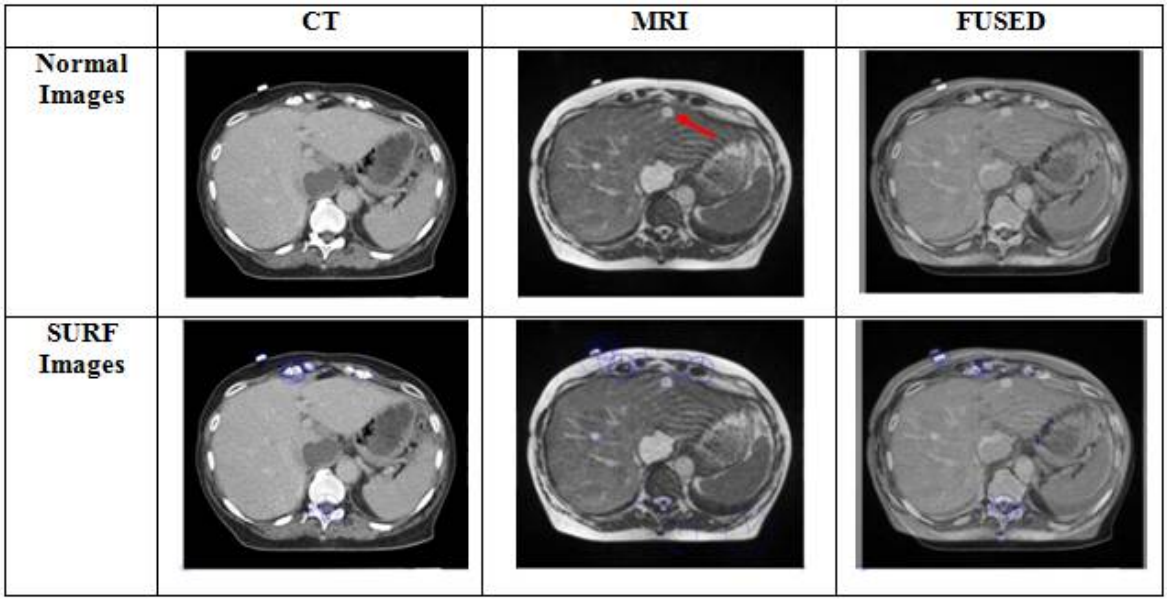


Figure 4.8: Fusion and SURF of Dataset 8

Now we did image fusion and applied SURF on the two models. The fused image gave us a more accurate and better result and had the details of both the images. Surf feature extraction was done to get those interest points on the basis of which we could do our classification.

Then we calculated some of the statistical based features:

Table 4.8: Parameters for Dataset 8

	CT	MRI	FUSED
SNR	11.364	10.36	10.033
PSNR	17.836	17.769	17.593
MSE	82.923	71.705	64.4
MEAN	88.175	81.638	85.202
STD. DEVIATION	82.923	71.705	64.40
VARIANCE	6876.336	5141.651	4147.46
ENTROPY	5.673	6.726	6.565
SKEWNESS	0.362	0.685	0.040
KURTOSIS	1.841	2.683	2.156

From the results of the characteristics of the images it is clear that the fused image has less peak error and less Mean square error as compared to other two images of CT scan and MRI imaging and gave us the better result and better understanding of the image and the disease. According to the skewness values the CT image is fairly symmetrical, MRI image is moderately skewed whereas the fused image which we get is again fairly symmetrical and as kurtosis value is less than 3 so it is platykurtic.

4.2 CLASSIFICATION RESULTS

Classification of selected features is the next stage in the process. In this whole process, classification of new observations is done on the basis of the input data. It is a supervised learning approach. This data can either be bi-class or may be multi-class**.** Here in our project we have used different classifiers to classify our extracted features. Different classifiers gave us different results with various accuracy rates.

Here, the numbers 0, 1, 2, 3, 4, 5 stands for:

0- No Detection
1- Pre-cancer
2- Tumour
3- Cancer
4- HCC
5- Fatty Liver

These are basically the diseases which we are able to identify after the all the process is done. This means that after the classification when we provide any new data, we will quickly get our results.

Now let us see what results we got from the different classifiers:

1. **SVM CLASSIFIER:** An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

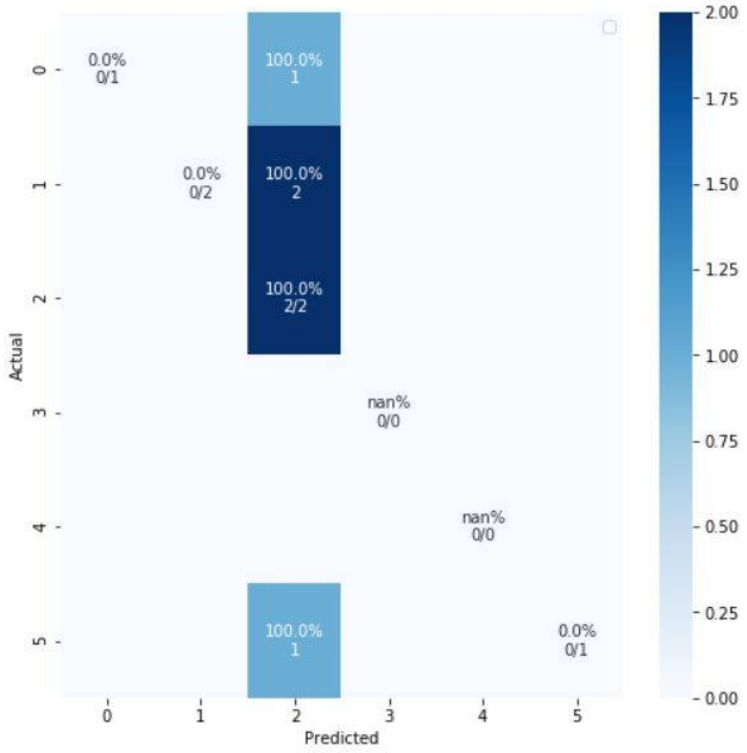


Figure 4.9: Heat map of SVM Classifier

It gives us the accuracy rate of 33%.

2. **ADAPTIVE CLASSIFIER:** This classifier integrates the parameters to find near to optimal performance. It uses the iterative method to get the results.

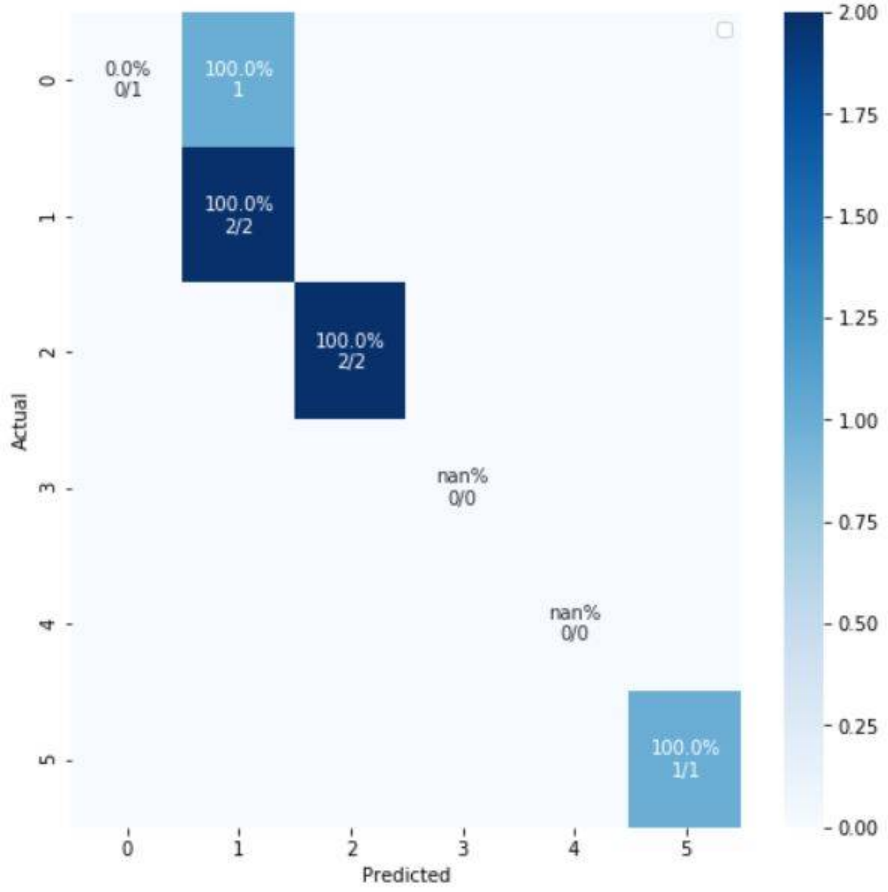


Figure 4.10: Heat map of Adaptive Classifier

This classifier gives us 83% accuracy.

3. **DECISION TREE CLASSIFIER:** In this classifier, prediction is done by iterating the input data via a learning tree. So in this whole process, for reaching a specific category, the feature properties are compared relative to decision conditions. It splits the data into smaller units and the final result is represented in the form of a tree with leaf and decision nodes.

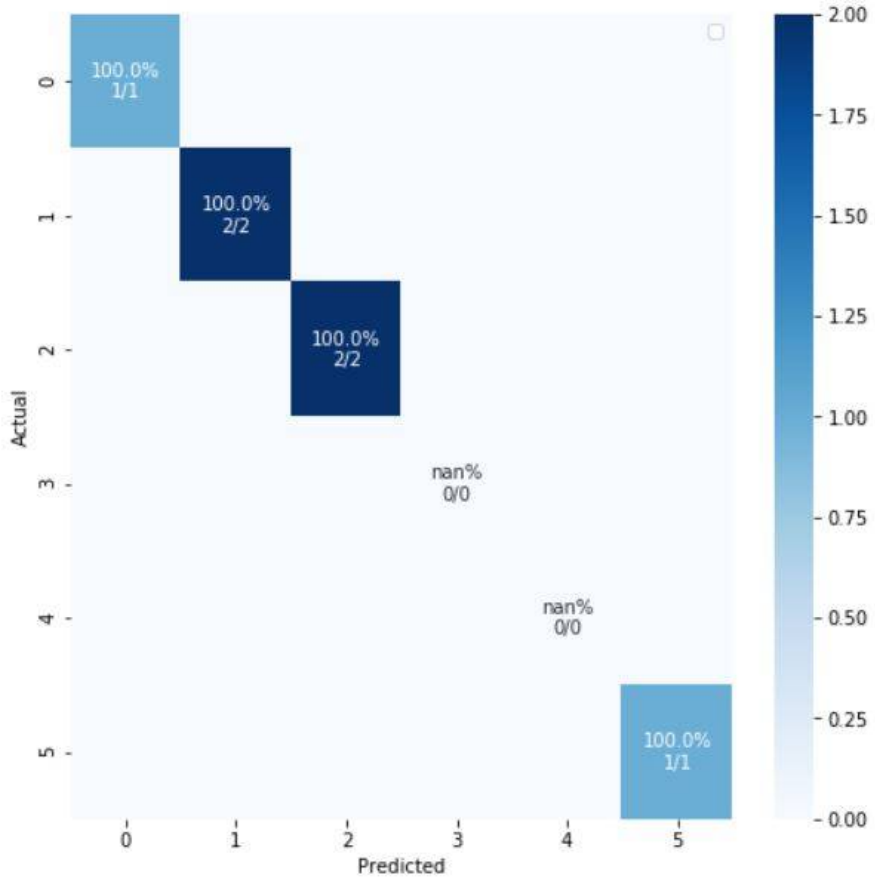


Figure 4.11: Heat map of Decision Tree Classifier

It gives us the accuracy rate of 99.9%.

4. **NAIVE BAYES CLASSIFIER:** This classifier works according to the Bayes' Theorem. It assumes individuality amongst its predictors. In simple words it can be explained as, that all the features in this contribute individually for finding the probability. The features are not related in this.

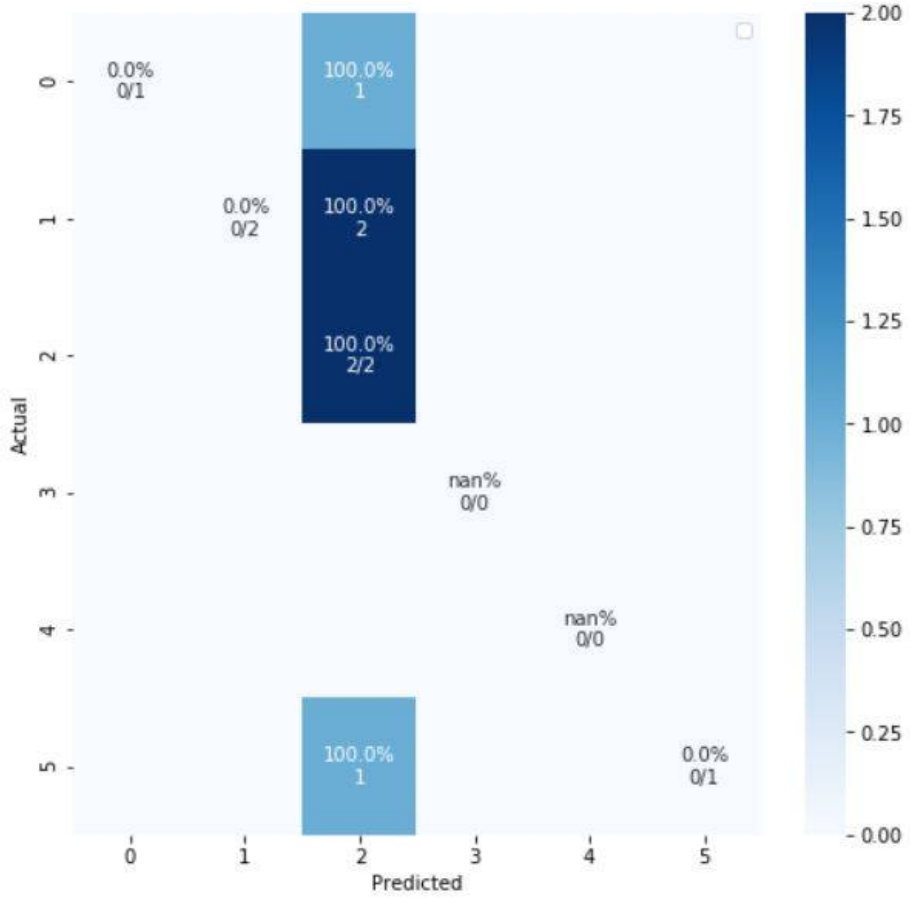


Figure 4.12: Heat map of Naive Bayes Classifier

Besides simple and easy problems, very highly complicated classification problems can also be solved by Naïve Bayes. It gives us the accuracy rate of 33%.

5. **GAUSSIAN NAÏVE BAYES CLASSIFIER**: Besides being a linear classifier, Naïve Bayes is also a probabilistic classifier. It is a supervised machine learning algorithm. Gaussian process is used to generate the big data. This generation of data is considered by the, Naïve Bayes classifier.

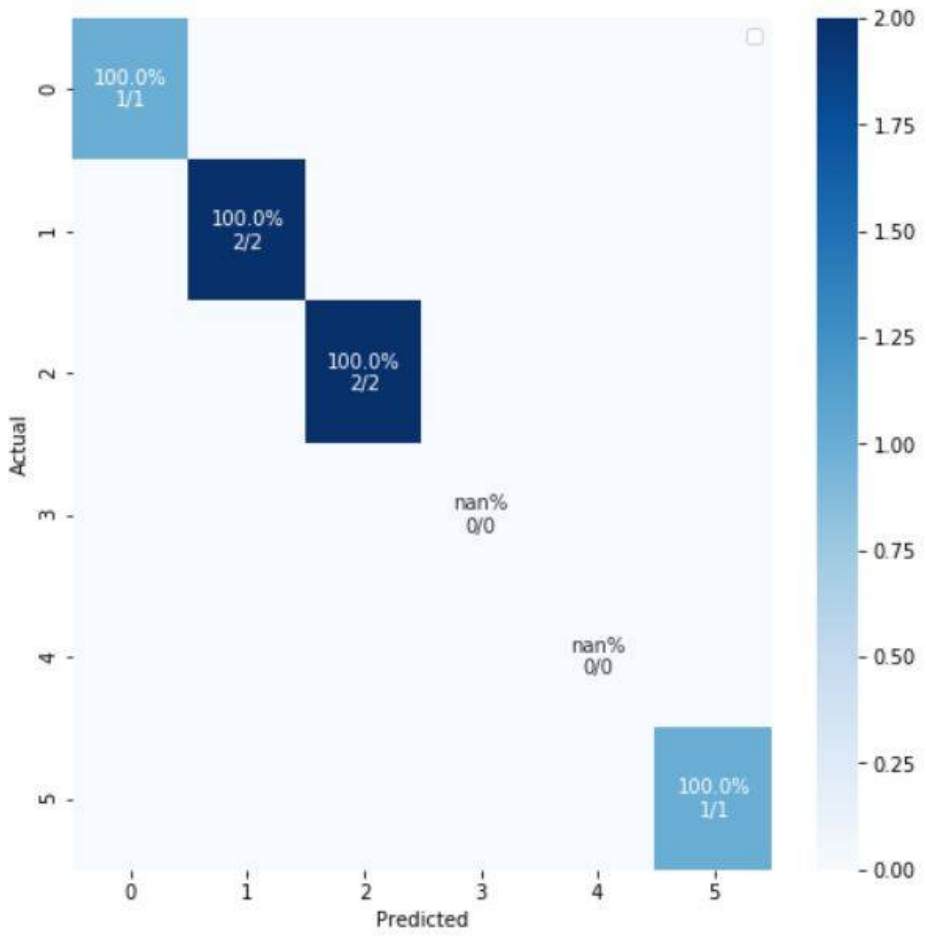


Figure 4.13: Heat map of Gaussian Naive Bayes Classifier

This classifier gives us the accuracy of 99.9%.

6. **K-NEAREST NEIGHBOR CLASSIFIER:** All available cases are stored by k-NN and new cases are classified by a greater number of votes of the k neighbours. Measurement is done using a distance function; the case which is to be allotted to the class is generally unvarying amongst its k adjoining neighbours.

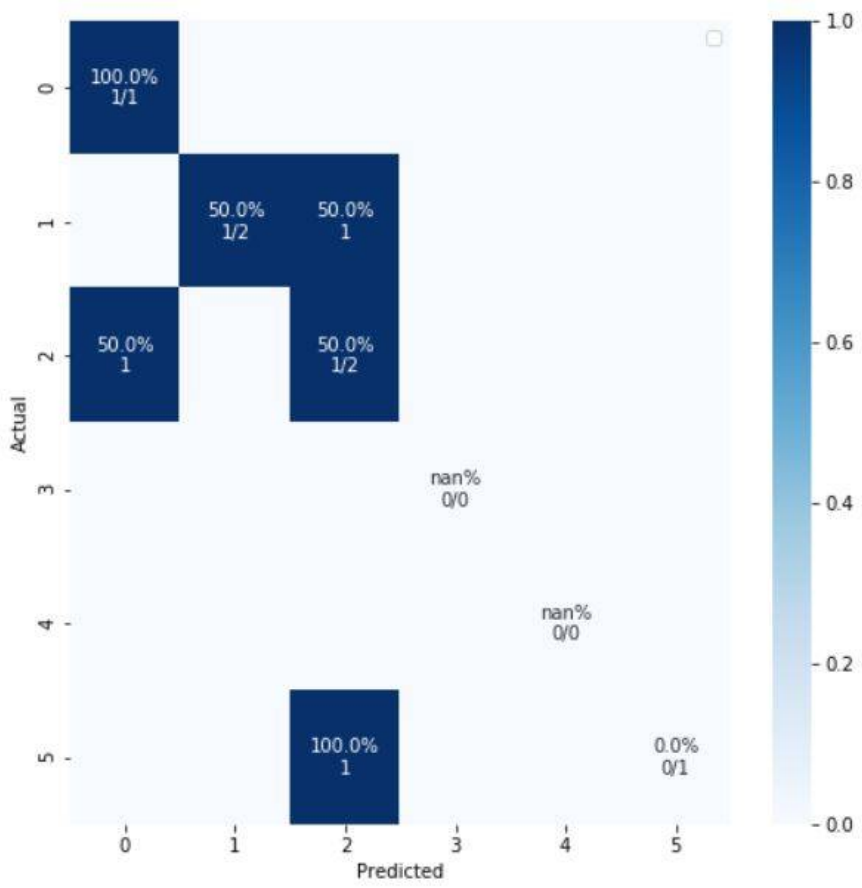


Figure 4.14: Heat map of k-NN Classifier

This classifier gives us the accuracy rate of 50%.

7. **STOCHASTIC GRADIENT DESCENT CLASSIFIER:** Discriminative learning of classifier such as (linear) SVM and Logistic Regression. For the optimization of an objective function, this is an iterative way.

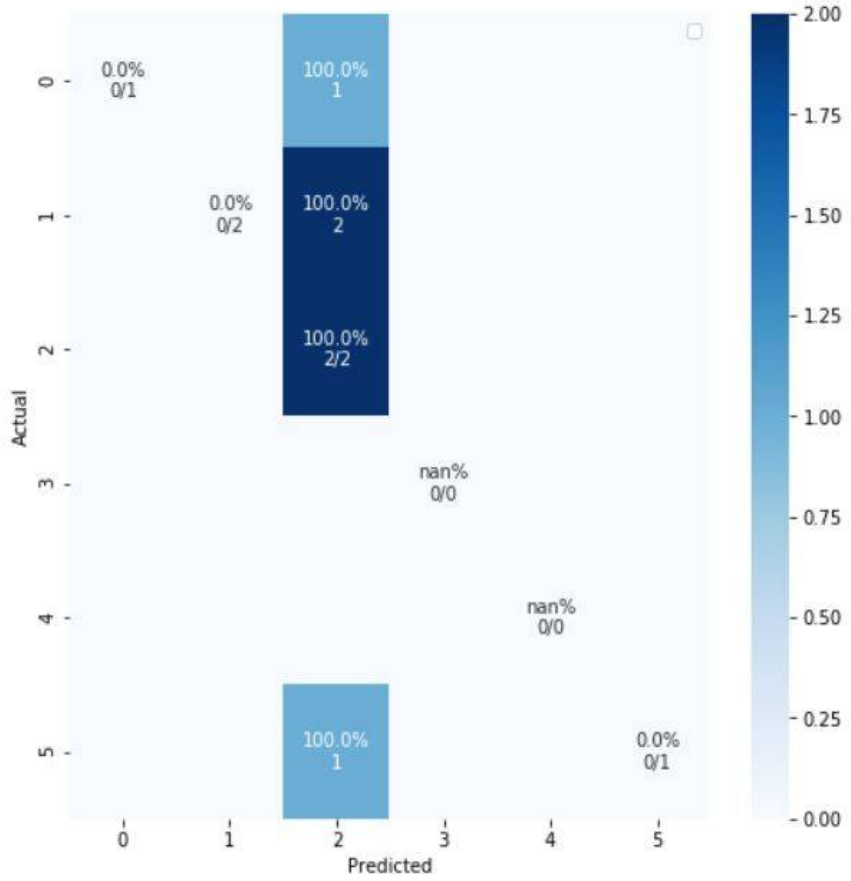


Figure 4.15: Heat map of Stochastic Gradient Descent Classifier

This classifier gives us an accuracy rate of 33%.

8. **LOGISTIC REGRESSION CLASSIFIER:** Used for the estimation of discrete values (binary values) based on the given set of individualistic variable, it is not a regression algorithm but a classification. In simple words, here probability of occurrence of an event is predicted by fitting data to a logit function or we can say logit regression**.**

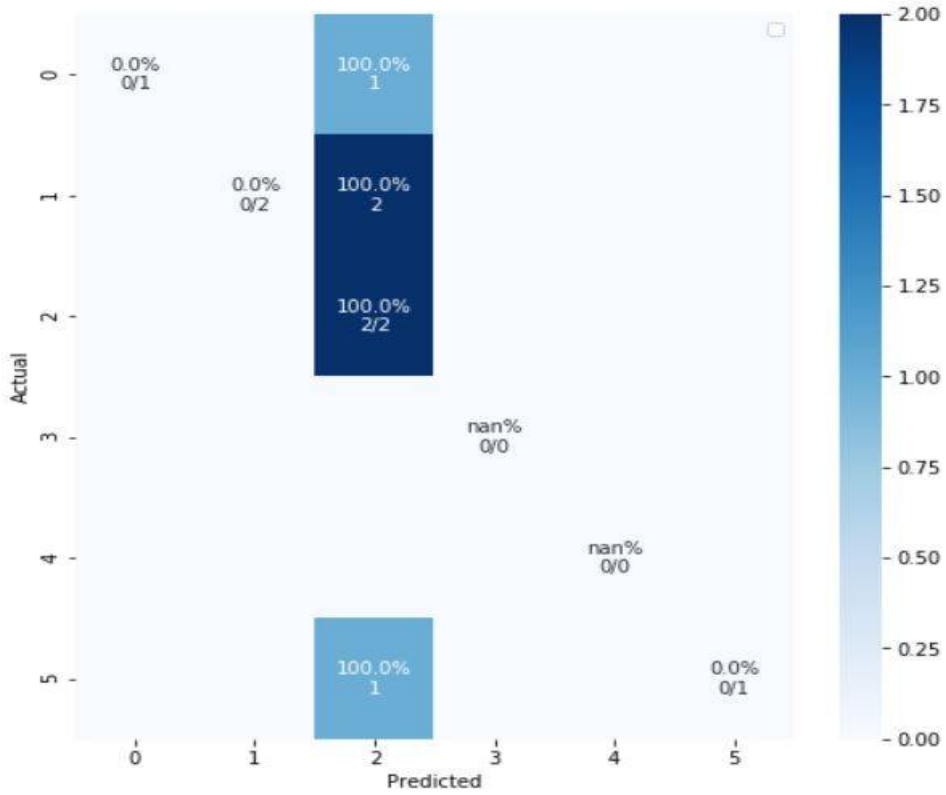


Figure 4.16: Heat Map of Logistic Regression Classifier

This classifier gives us the accuracy rate of 33%.

The accuracy rate of all the classifiers is given in the following table:

Table 4.9: Result of Classification

CLASSFIER	ACCURACY
DECISION TREE CLASSIFIER	99.9%
GAUSSIAN NAÏVE BAYES CLASSIFIER	99.9%
ADAPTIVE CLASSIFIER	83%
k-NN CLASSFIER	50%
SVM CLASSIFIER	33%
STOCHASTIC GRADIENT DESCENT CLASSIFIER	33%
LOGISTIC REGRESSION	33%
NAÏVE BAYES'	33%

The given table shows that we get the best results from the two classifiers out of eight classifiers. Decision Tree and Gaussian classifier gives the accuracy of 99.9%. It proves that when a new dataset will be entered these classifiers will tell the output almost accurate.

CONCLUSION AND FUTURE SCOPE

In this project, we have done thorough study about the liver and its characteristics and also about different modalities of the liver scan like CT Scan and the MRI. Then we performed the fusion of CT and MRI images. After that, preprocessing of both the images was done and then on the basis of CT image, we did the registration of MR image. Now from both the images and even from the fused image, we extract our points of interest. Also, some of the statistical based features such as mean, median, peak signal to noise ratio (PSNR), standard deviation and some more were calculated. All these features proved that the fused image gave us the better result; it had all the important details of the individual images. Then, we also extracted the SURF features which gave us all those important regions, on the basis of what classification could be done. And now at the end, we applied several machine learning algorithms to get the classified data. The obtained accuracy using Gaussian Naïve Bayes algorithm with 99.9% with the application of the technique for large database and improving efficiency of time, this application can be further developed in future. For the determination of the best classifier which provides better precise decisions to determine the tumor, we can implement different types of classifiers.

REFERENCES

[1] R. C Gonzalez, R E. Woods, “Digital image processing.”, Upper Saddle River, N.J ,Prentice Hall. pp. 23-28, 1954 (3rd edition.).

[2] B.R. Thapa and Anuj Walia, “Liver Function Tests and their Interpretation”, Vol. 74, pp. 663-671, July 2007.

[3] https://www.hopkinsmedicine.org/health/conditions-and-diseases/liver-anatomy-and-functions

[4] https://www.123rf.com/photo_47868520_stock-vector-medical-structure-of-the-liver-illustration.html

[5] https://courses.lumenlearning.com/boundless-ap/chapter/the-liver/

[6] https://stock.adobe.com/uk/images/liver-disease-fatty-fibrosis-alcoholic-hepatitis-hepatitis-cirrhosis-and-cancer-liver-anatomical-vector-illustration-in-flat-style-isolated-over-white-background/279143392

[7] Ian R. Wanless and John S. Lent, “Fatty Liver Hepatitis (Steato hepatitis) and Obesity: An Autopsy Study with Analysis of Risk Factors”, Department of pathology, University of Toronto. Toronto, Canada, Vol. 12, June 1990.

[8] Charles S. Lieber, Leonore M. Decarli and Emanuel Rubin, “Sequential Production of Fatty Liver, Hepatitis, and Cirrhosis in Sub-Human Primates Fed Ethanol with Adequate Diets”, Vol. 72, No. 2, pp. 437-441, February 1975.

[9] http://theconversation.com/gut-microbes-can-get-you-drunk-and-damage-your-liver-123741

[10] J. H. Oh, J. Gao J, “Fast kernel discriminant analysis for classification of liver cancer mass spectra,” IEEE/CM Transactions on Computational Biology and Bioinformatics, vol. 8 (6), pp. 15221534, 2011.

[11] http://theconversation.com/gut-microbes-can-get-you-drunk-and-damage-your-liver-123741

[12] George K. Michalopolous and Marie C. Defrancis, "Liver Regeneration”, Vol. 276, pp. 60-66, April 1997

[13] Thomas E. Starzl, Shunzaburo Iwatsu, “Evolution of Liver Transplantation”, Vol. 2, No. 5, pp. 614-636, 1982.

[14] https://en.wikipedia.org/wiki/MATLAB

[15] https://en.wikipedia.org/wiki/Microsoft_Office_Picture_Manager

[16] https://en.wikipedia.org/wiki/ImageJ

[17] https:// https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)

[18] Kenneth S¨orensen, Marc Sevaux, Fred Glove “A History of Metaheuristics∗” Vol. 1, pp. 1-20, 4 Apr 2017.

[19] James, A.P., Dasarathy, B V. "Medical Image Fusion: A survey of state of the art". *AIP Conference*, pp. 1-15, December 2013.

[20] H.M Taylor, P. R. Ros, “Hepatic imaging: An overview”, *Radiologic Clinics of North America*, Vol. 36, No. 2, 1998.

[21] Atrayee Dutta, Aditya Dubey, “Detection-of-Liver-Cancer-Using-Image-Processing”, IEEE, April 2019.

[22] Kunio Doi, “Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential”, March 2007.

[23] Preeti Sharma, Tapan Jain, “ Robust Digital Watermarking for Coloured Images using SVD and DWT Technique” ,IEEE, pp. 978-1-4799-2572-8, 2014.

[24] S. Jain, M. Sachdeva, P. Dubey and A. Vijan, "Multi-Sensor Image Fusion using Intensity Hue Saturation Technique", 3rd International Conference on Advanced Informatics for Computing Research, Solan, India on June 15-16, 2019, pp 147-157.

[25] K. V. Mardia, “Measures of multivariate skewness and kurtosis with applications”, Vol. 57, No. 3, pp. 519, 1970.

[26] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in European conference on computer vision, Springer, pp. 404-417, 2006.

[27] Robert M. Haralick , K. Shanmugam , “Textural features for image classification”, Vol. 3, pp. 610-621, November1973.

[28] J. S. Raikwal, Kanak Saxena, “Performance Evaluation of SVM and K-Nearest Neighbour Algorithm over Medical Data set” , International Journal of Computer Applications ,Vol.50 No.14, July 2012.

[29] https://en.wikipedia.org/wiki/Computer-Aided-Design

[30] A.P. Jamesa and B.V. Dasarathyb," Medical Image Fusion: A survey of the state of the art", Information Fusion, pp 1-45, 2014.

[31] https://www.researchgate.net/figure/Schematic-Diagram-of-Image-Fusion-using-Wavelet-Transform-The-wavelet-decomposition-of-an_fig1_311590630

[32] B. Mehera, S. Agrawala, R. Pandaa and A. Abrahamb, "A survey on region based image fusion methods", Information Fusion, 48, pp -120-125, 2019

[33] https://www.researchgate.net/figure/DWT-SWT-DTCWT-Based-Image-Fusion_fig3_228886260

[34] Sonali Mane, S. D. Sawant, “Image Fusion of CT/MRI using DWT, PCA Methods and Analog DSP Processor”, Int. Journal of Engineering Research and Applications, Vol. 4, Issue 2 (Version 1), , pp.557-563, February 2014.

[35] S. Jain, M. Sachdeva, P. Dubey and A. Vijan, "Multi-Sensor Image Fusion using Intensity Hue Saturation Technique", 3rd International Conference on Advanced Informatics for Computing Research, Solan, India on June 15-16, 2019, pp 147-157.

[36] Venus Bansal, Dr. Ajay Kumar Singh, “A Review Paper on Feature Extraction and General Approaches for Face Recognition”, Volume 6 Issue I, January 2018.

[37] https://github.com/learn-co-students/dsc-0-09-16-skewness-and-kurtosis-data-science)

[38] https://study.com/academy/lesson/skewness-in-statistics-definition-formula-example.html

[39] https://www.investopedia.com/terms/s/skewness.asp

[40] https://www.investopedia.com/terms/k/kurtosis.asp

[41] https://keydifferences.com/differences-between-skewness-and-kurtosis.html

[42] https://shodhganga.inflibnet.ac.in/bitstream/10603/24460/9/09_chapter4.pdf

[43] Sami Brandt, Jorma Laaksonen and Erkki Oja, “Statistical Shape Features for Content-Based Image Retrieval” , Journal of Mathematical Imaging and Vision 17, pp. 187–198, 2002.

[44] Ludvík Tesar, Akinobu Shimizu, et.al. , “Medical image analysis of 3D CT images based on extension of Haralick texture features” , Computerized Medical Imaging and Graphics, Vol. 32, pp. 513–520, 2008.

[45] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in European conference on computer vision, Springer, pp. 404-417, 2006.

[46] P. M. Panchal, S. R. Panchal, *et al*, "A Comparison of SIFT and SURF ", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 2, April 2013

[47] Aurangzeb Khan, Khairullah Khan *et al* ,“A Review of Machine Learning Algorithms for Text-Documents Classification”, Journal of advances in information technology, Vol. 1, No. 1, February 2010.

[48] Durgesh K. Srivastava, Lekha Bhambu, " Data Classification using Support Vector Machine", Journal of Theoritical and Applied Information Technology, Vol 12, No. 1, 2010

[49] Aman Kataria, M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 6, June 2013.

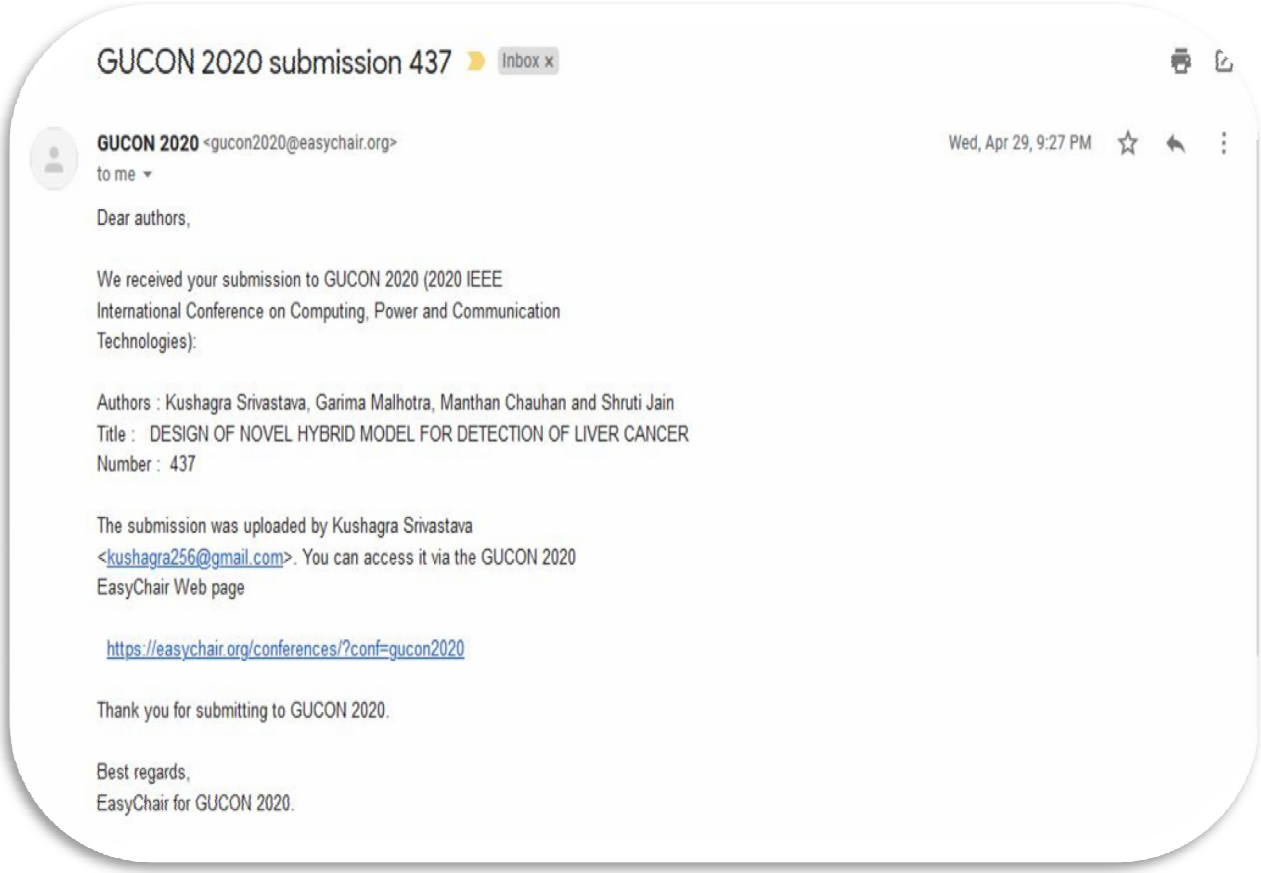
[50] https://theoncologist.alphamedpress.org/content/9/4/385/F1.large.jpg

[51] https://theoncologist.alphamedpress.org/content/9/4/385/F4.large.jpg

[52] https://theoncologist.alphamedpress.org/content/9/4/385/F3.large.jpg

PUBLICATION

K. Srivastava, G. Malhotra, M. Chauhan, S. Jain, "Design of Novel Hybrid Model for Detection of Liver Cancer", GUCON 2020 IEEE International Conference on Computing, Power and Communication Technologies, 29th April 2020.



GUCON 2020 submission 437 Inbox x

GUCON 2020 <gucon2020@easychair.org> Wed, Apr 29, 9:27 PM

to me

Dear authors,

We received your submission to GUCON 2020 (2020 IEEE International Conference on Computing, Power and Communication Technologies):

Authors : Kushagra Srivastava, Garima Malhotra, Manthan Chauhan and Shruti Jain
Title : DESIGN OF NOVEL HYBRID MODEL FOR DETECTION OF LIVER CANCER
Number : 437

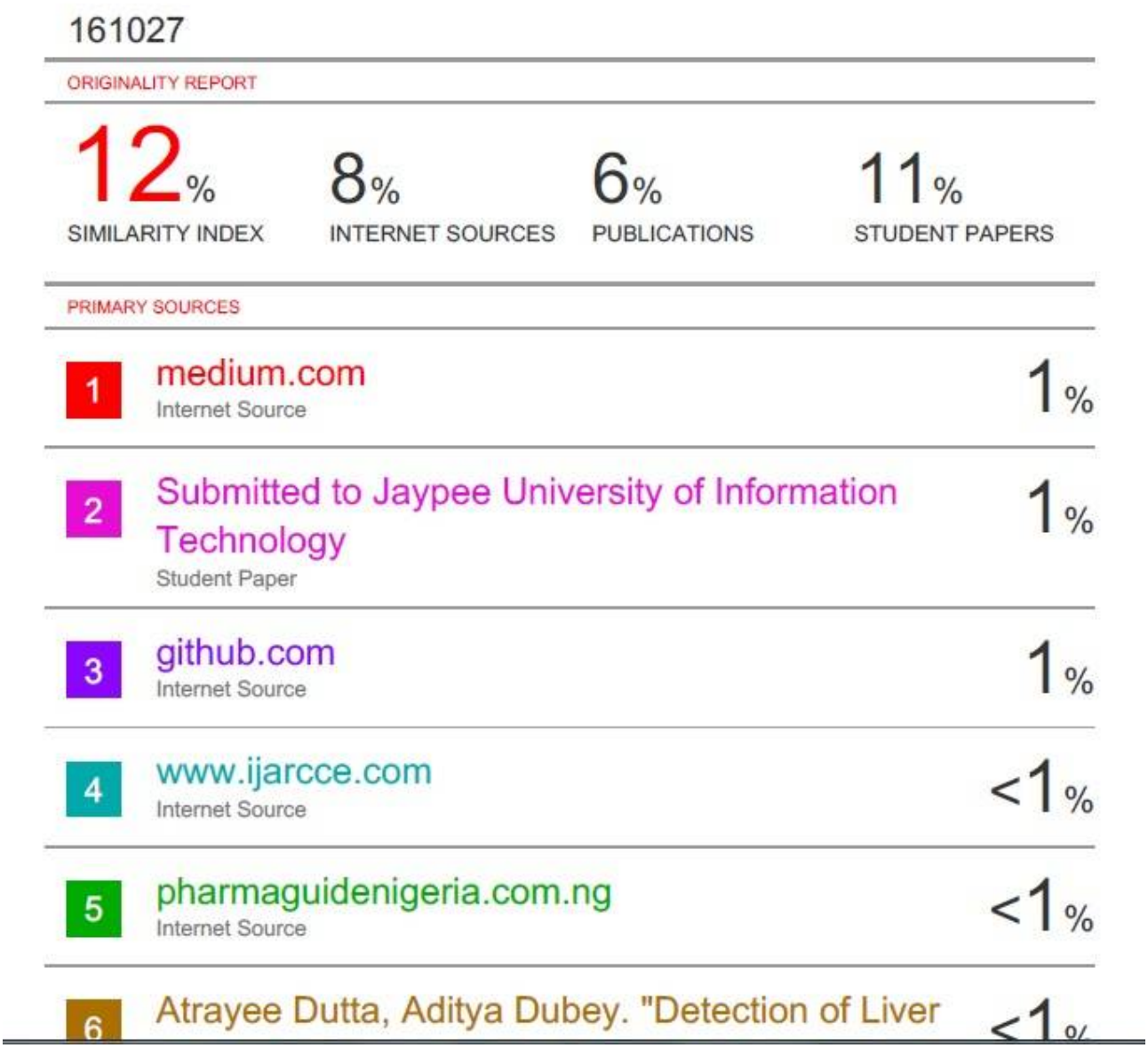
The submission was uploaded by Kushagra Srivastava <kushagra256@gmail.com>. You can access it via the GUCON 2020 EasyChair Web page

https://easychair.org/conferences/?conf=gucon2020

Thank you for submitting to GUCON 2020.

Best regards,
EasyChair for GUCON 2020.

PLAGIARISM REPORT



161027

ORIGINALITY REPORT

12%	8%	6%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	medium.com Internet Source	1%
2	Submitted to Jaypee University of Information Technology Student Paper	1%
3	github.com Internet Source	1%
4	www.ijarcce.com Internet Source	<1%
5	pharmaguidenigeria.com.ng Internet Source	<1%
6	Atrayee Dutta, Aditya Dubey. "Detection of Liver	<1%

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: 15/07/2020

Type of Document (Tick): PhD Thesis | M.Tech Dissertation/ Report | **B.Tech Project Report** | Paper

Name: Garima Malhotra, Kushagra Srivastava and Manthan Chauhan **Department:** ECE

Enrolment No: 161027, 161033 and 161117

Contact No. 9805100991, 7408336788, 9149519004 **E-mail:** garima15698@gmail.com, kushagra256@gmail.com and manthanchauhan0099@gmail.com

Name of the Supervisor: Dr. Shruti Jain, Department of ECE, JUIT, Solan

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): LIVER CANCER DETECTION AND CLASSIFICATION BASED ON OPTIMUM FEATURE FUSION

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages = 75
- Total No. of Preliminary pages = 02
- Total No. of pages accommodate bibliography/references = 04

Garima Srivastava Manthan Chauhan

(Signature of Students)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at **12** %. Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

15/7/2020 15/07/2020

(Signature of Guide/Supervisor) **Signature of HOD**

Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at plagcheck.juit@gmail.com