

# **DEEP LEARNING APPROACHES FOR BREAST CANCER DETECTION IN THERMOGRAPHIC IMAGES**

## **A Thesis**

Submitted in partial fulfillment of the requirements for the

Award of the degree of

**DOCTOR OF PHILOSOPHY**

In

**Electronics and Communication Engineering**

By

**Sonalee P. Suryawanshi**

**(41800305)**

Under the Supervision of

**Dr. Bhaveshkumar C. Dharmani**

Professor

(Domain: Signal & Image Processing)

School of Electronics and Electrical Engineering

Phagwara, Punjab, India



**L** OVELY  
**P** ROFESSIONAL  
**U** NIVERSITY

---

*Transforming Education Transforming India*

**LOVELY PROFESSIONAL UNIVERSITY  
PUNJAB**

**December, 2024**



## CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled **“DEEP LEARNING APPROACHES FOR BREAST CANCER DETECTION IN THERMOGRAPHIC IMAGES”** in partial fulfilment of the requirement for the award of the degree of **Doctor of Philosophy** and submitted to the Department of Electronics and Communication Engineering of Lovely Professional University, Phagwara (Punjab) is an authentic record of my work carried out during the period from January 2018 to June 2023 under the supervision of **Dr. Bhaveshkumar C. Dharmani**, Associate Professor, School of Electronics and Electrical Engineering, Lovely Professional University.

Sign

A handwritten signature in black ink, appearing to read "Sonalee P. Suryawanshi", written over a horizontal line.

**Sonalee P. Suryawanshi**

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

**Date: 19/12/2024**

A handwritten signature in black ink, appearing to read "Bharmani", written over a horizontal line.

**(Dr. Bhaveshkumar C. Dharmani)**

Supervisor

The Ph.D. Viva-Voce Examination of has been held on \_\_\_\_\_

Signature of Supervisor

Signature of External Examiner

## ACKNOWLEDGMENT

---

First, I feel great pleasure in acknowledging my deepest gratitude to my revered guide and mentor **Dr. Bhaveshkumar C. Dharmani**, Associate Professor, School of Electronics and Electrical Engineering, Lovely Professional University, Punjab, India, under whose guidance, motivation, and vigilant supervision I succeeded in completing my work. They both infused the enthusiasm to continue this work into me.

I am also ever grateful to my Husband, **Mr. Amit H. Ukey** for his generous help. Also, my sincere thanks are due to all the faculty members and non-teaching staff of the School of Electronics and Electrical Engineering, Lovely Professional University, Punjab for availing me all the necessary facilities and cooperation for this work.

Words are inadequate to express my heartfelt gratitude to my affectionate **parents** and my **husband** who have so much confidence in me and by whose efforts and blessings I have reached here. Not to forget my lovely **son Avant** and **daughter Sanvi** who have been my constant driving force for completing this work.

I find it hard to express my gratitude to the **Almighty** in words for bestowing upon me his deepest blessings and providing me with the most wonderful opportunity in the form of life of a human being and for the warmth and kindness he has shown upon me by giving me life's best.

At last, I wish to express my heartiest thanks to my **friends** for their support, love, and inspiration.



(Sonalee P. Suryawanshi)

## ABSTRACT

Breast cancer kills many women worldwide. Early detection reduces breast cancer fatalities. An automated cancer detection method can identify abnormal breast tissue earlier. Due to noise and distortion, misdiagnosis might harm health. Detecting breast cancer faster is difficult. Breast cancer images have many qualities that can help locate the disease. However, diagnosing cancer's key traits might be difficult. This complicates breast cancer detection. Breast cancer early detection has used several classification and optimization techniques in recent years. However, correct identification took time and effort.

Recent years have witnessed the proliferation of research projects aimed at identifying breast images that have been damaged by cancer. Since there was interference in the environment, the efficacy of the methods that were available was insufficient. In this article, many screening procedures and segmentation techniques are discussed in order to identify the tumour at an earlier stage. Yet, a precise and trustworthy diagnosis is required for early identification.

The diagnosis of breast cancer by thermographic imaging is widely regarded as the most effective method for both mass screening and early identification of the disease. In the present study, a unique deep learning method is proposed. This approach makes use of beta entropy-based features and a newly constructed Hybrid Harris Hawks Jaya Optimization (HHHJO) algorithm for hyper parameter tuning in deep neural networks. In recent years, entropy-based characteristics have become more prevalent in the field of medical diagnostics. The beta divergence method has been shown to have the virtue of being resilient against outliers. This has seen significant use in a variety of statistical calculations, as well as deep learning. Yet, the beta entropy-based characteristics have never previously been investigated for their use in medical diagnostics. Deep learning algorithms have shown their usefulness in a variety of applications, particularly in the detection job. The effectiveness of a deep learning algorithm is very dependent on the hyperparameter tuning, which is traditionally performed with human skill via a system of trial and error. Various concatenated features were offered to classification stage and classification was performed with the help of Optimized DNN and some parameter like learning rate of DNN, epochs count

of DNN and hidden neuron count of DNN was tuned by utilizing developed HHHJO for maximizing the accuracy to offer effective classification rate. The HHO method is used for global search, while the Jaya optimization is used for local search. The Hybrid Harris Hawks Jaya Optimization (HHHJO) algorithm is derived in this article. At both the segmentation and the categorization phases, the resulting method is put to use to make an estimate of the best possible collection of hyperparameters. As was to be predicted, the newly developed method using beta entropy as a feature was able to produce an increase in breast cancer diagnosis accuracy that was **more than 5% when compared** to the state-of-the-art accuracy acquired using a standard dataset. In conclusion, the suggested method has the potential to be used for the identification of breast cancer via thermographic images with better degrees of accuracy.

# CONTENTS

Particulars	Page No.
<b>Chapter 1</b>	
<b>Introduction</b>	<b>1</b>
1.1 Background of the Breast Cancer Diagnosis	3
1.2 BC- Breast Cancer Statistics	4
1.3 Signs and symptom of Breast Cancer	6
1.4 Literature on Diagnosis Tool	8
1.5 Inspiration behind Choosing Research Area	9
1.6 Research Objectives	12
1.7 Research Methodology	13
1.8 Research Contribution	14
1.9 Thesis Data Flow	15
1.10 Summary	15
<b>Chapter 2</b>	
<b>The Problem Statement and Literature Survey</b>	<b>16</b>
2.1 Review on Available Literature	16
2.2 Feature Extraction for Breast Cancer Detection	21
2.3 Region of Interest (ROI) Based Breast Cancer Diagnosis	24
2.4 Optimization Techniques for Breast Cancer Detection	26
2.5 Classifiers for Thermography Breast Cancer Identification	27
2.6 Deep Learning Approach for Classification	30
2.7 Summary	35
<b>Chapter 3</b>	
<b>Beta Entropy from Beta Divergence</b>	<b>36</b>
3.1 Introduction	36
3.2 Information Definition – Entropy Definition – Shannon Entropy	37
3.2.1 Entropy	37
3.2.2 Shannon Entropy	38
3.2.3 Properties of Shannon Entropy	38
3.3 KL Divergence from Entropy	40
3.3.1 Applications of KL Divergence	41
3.4 Alternative Entropy Definitions	41

Particulars	Page No.
3.4.1 Minimum entropy, Maximum entropy	41
3.4.2 Sample entropy	43
3.4.3 Approximate entropy	44
3.4.4 Alpha entropy	46
3.5 Density Power divergence or Beta divergence	48
3.6 Deriving Beta entropy from Beta divergence	50
3.7 Summary	52
<b>Chapter 4</b>	
<b>Hybrid Harris-Hawk Jaya Optimization Algorithm</b>	<b>54</b>
4.1 Introduction	55
4.2 HHO Algorithm	55
4.3 JAYA Algorithm	59
4.4 Novel HHHJO Algorithm Proposal	60
4.5 Summary	63
<b>Chapter 5</b>	
<b>Overall Solution</b>	<b>65</b>
5.1 Optimized Deep Neural Network for the breast cancer detection in Thermographic images.	65
5.2 Input Image Pre-processing	68
5.3 GVF (Gradient Vector Flow) -based breast segmentation with optimized abnormality segmentation for breast cancer detection: GVF-based Breast Segmentation	69
5.4 Optimized FCM-based Abnormality Segmentation	70
5.5 Enhanced thermogram breast cancer detection using weight-optimized deep neural network: Feature Extraction	72
5.6 Basic DNN Model	78
5.7 Optimized DNN-based Detection	79
5.8 Conclusion	82
<b>Chapter 6</b>	
<b>Result and Analysis</b>	<b>83</b>
6.1 Experimental Set-up	83
6.2 Efficiency Metrics	84
6.3 Thermogram Dataset Description	85
6.4 Convergence Examination of Established Model with Diverse Optimization Algorithms	94

<b>Particulars</b>		<b>Page No.</b>
6.5	Overall performance analysis of developed model with diverse classifier	97
6.6	Summary	105
<b>Chapter 7</b>		
<b>Conclusion, Future Work and Discussion</b>		<b>106</b>
7.1	Conclusion	106
7.2	Future Work	107
7.3	Discussion	107
<b>References</b>		<b>109</b>
<b>Anneuxre</b>		<b>120</b>



## LIST OF TABLES

<b>Table No.</b>	<b>Title</b>	<b>Page No.</b>
2.1	Landscapes and challenges of conventional thermogram-based BC recognition prototypes	32
2.2	Structures and encounters of conventional thermogram-based cancer recognition prototypes	33
6.1	Statistical feature analysis with developed thermogram-based breast cancer detection model over classifier	99
6.2	Statistical and textural feature analysis executed on the established breast cancer detection model with already existing classifiers	100
6.3	Statistical, texture and entropy feature analysis on developed thermogram breast cancer detection model over classifiers	101
6.4	Statistical, texture, entropy and beta-entropy analysis over the breast cancer identification model with several classifiers	102
6.5	Overall performance analysis on developed breast cancer detection model with classifier approaches	104
6.6	Statistical feature analysis with developed thermogram-based breast cancer detection model over classifier	105

## LIST OF FIGURES

Figure No.	Title	Page No.
1.1	A chart showing various races, sexes of cancer diagnosis as per stats of SEER 21 2013–2017. Credit: National Cancer Institute	4
1.2	Rate of New Cases and Deaths per one lakh women and % of fresh Cases by Age: Breast Cancer	5
1.3	Statistics on Breast Cancer in India, Broken Down by Age Group	11
2.1	Cells in blood flow or the lymph system spread breast cancer	18
2.2	Classification of various B-Cancers reported in various countries time to time	19
2.3	Breast Cancer Imaging	20
2.4	Basic block diagram of Deep Neural Network	34
4.1	Flow diagram of developed HHHJO	62
5.1	Architectural view of developed thermogram-based breast cancer detection model	67
5.2	Diagrammatic representation of optimized FCM-aided breast cancer segmentation	72
5.3	Developed ODNN-based breast cancer detection model	81
6.1	Samples of breast cancer-affected and healthy images from thermogram dataset	86
6.2	Resultant abnormality segmented images from the Optimized FCM technique	88
6.3	Evaluation on proposed thermogram-based breast cancer detection model with multiple classifier over “(a) accuracy , (b) F1-score, (c ) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NVP, (h) precision, (i) sensitivity & (j) specificity”	93
6.4	Convergence examination on thermogram-based breast cancer detection using proposed model with multiple classifier using (a) GLCM features, (b) 1st and 2nd order features, (c ) LBP features and (d) Entropy features	97

## LIST OF ABBREVIATIONS

Expansion	Abbreviation
Multimodal Firefly Optimization Technique	MFFO
Statistical Maximum Likelihood Optimization and Curvilinear Support Vector Machine Method	SMLO-CSVM,
Population Rescaled Differential Evolution with Weighted Boosting Method	PRDE-WB
Hybrid Harris Hawks Jaya Optimization	HHHJO
Digital Signal Processing	DSP
Computer-Aided Design	CAD
Generalized Maximum Likelihood	GML method
Peak Signal-To-Noise Ratio	PSNR
Mean Square Error	MSE
False Positive Rate	FPR
Mega-Trend Diffusion	MTD
Deoxyribonucleic Acid	DNA
Stacked Sparse Auto Encoder,	SSAE,
Magnetic Resonance Imaging	MRI,
Sequential Quadratic Programming	SQP.
Ant Colony Optimization	ACO
Feed-Forward Artificial Neural Network Classifier	FFANN
Support Vector Machine	SVM.
Weighted Area Under the Receiver Operating Characteristic Curve Ensemble	WAUCE
Mammographic Image Analysis Society Database	MIAS
Drug Information Tracking Integrator	DITI
General Value Functions	GVFs

Expansion	Abbreviation
Flow Cytometry	FCM
Gray Level Co-Occurrences Matrix	GLCM
Artificial Neural Networks	ANNs
Deep Neural Networks	DNN
Local Binary Patterns	LBP
Optimal Deep Neural Network	ODNN
Computed Tomography,	CT
Hierarchical Lossless Segmentation	HLS
Rectified Linear Unit	ReLU
Multilayer Perceptron	MLP
Extreme Learning Machine	ELM
Sound Navigation And Ranging	Sonar
Single Photon Emission Computed Tomography Or	SPECT
Positron Emission Tomography Or PET Scans	PET
Semi-Supervised Learning	SSL
Convolutional Neural Networks	CNN.
Regions Of Interest	ROI
A Computer-Aided Diagnosis	CAD
Case-Based Reasoning	CBR
Adaptive Neuro-Fuzzy Inference System	ANFIS
Particle Swarm Optimization-Support Vector Machine Model	PSO-SVM
Support Vectors	SV
Classification And Regression Tree	CART
Neural Networks	NN
Breast Cancer	BC

Expansion	Abbreviation
Cranio-Caudal	CC
Medio-Lateral Oblique Mammography Images	MLO
Machine Learning	ML
Wisconsin Diagnostic Breast Cancer	WDBC
Oppositional Improvement Based Tunicate Swarm Algorithm	OI-TSA
Harris Hawks Optimization	HHO
Jaya-Algorithm	JA
Grey -Valued Fourier Transform	GVFS
Optimized FCM	OFCM
Improved TSA	ITSA

# CHAPTER 1

## INTRODUCTION

---

As the most frequent form of cancer in females, breast cancer must be diagnosed at an early stage to have any chance of being treated effectively. Mammography is the gold standard for screening for breast cancer, however detecting tiny tumours may be challenging for radiologists. To detect breast cancer, an X-ray image of the breast is taken, and this procedure is known as a mammogram. Qualified medical experts depend heavily on mammograms to detect breast cancer in its earliest stages. Routine mammograms are the most reliable method doctors have for identifying breast cancer in its early stages. In rare cases, a mammography may not identify breast cancer for an entire year. By automatically detecting suspicious areas in mammograms, image processing methods may improve the precision of breast cancer diagnosis. The breast lump or mass or tumour feel is depending on its cause, location, and growth and symptoms of breast tumour that changes from person to person. Lumps in the breast do not cause any pain hence they are not noticeable. Symptoms of breast cancer includes persistent changes in the shape of the breast, increase in size of the breast, changes in the appearance of the breast, pain, changes in the nipple of the breast. Invasive breast cancer has a normal breast cancer symptom while ductal carcinoma in situ rarely causes symptoms that nipple discharge or may feel lump and it can be detectable by the mammogram. Thankfully, there is software designed specifically for this purpose. Several image processing methods are applicable to the diagnosis of breast cancer. There are two primary groups to which these methods belong: Feature extraction: This involves identifying and extracting features from mammograms that can be used to classify tumours as benign or malignant. Common features extracted from mammograms include texture, shape, and size.

### **Advantages of Image Processing for Breast Cancer Detection in Present Times**

When compared to more conventional methods of breast cancer diagnosis like mammography, image processing techniques have some benefits. Some of the benefits are:

Increased accuracy: Image processing techniques can increase the accuracy of BC detection by automatically identifying suspicious regions in mammograms. This can assist to lower the amount of false-positive and false-negative results, which in turn can lead to earlier breast cancer diagnosis and treatment.

Reduced cost: Image processing techniques can be used to screen large numbers of mammograms quickly and efficiently, which can help to reduce the cost of breast cancer screening.

Improved patient experience: Image processing techniques can be used to create 3D models of the breast, which can help radiologists to better visualize tumours\ and make more accurate diagnoses. This can help to reduce anxiety and stress for patients who are perusing a breast cancer diagnosis.

### **Challenges in Image Processing for Breast Cancer Detection**

While processing has many potential benefits for detecting breast cancer, there are also many obstacles that must be overcome. Problems may be seen in:

Data scarcity: There is a limited amount of data available for training and testing image processing algorithms for breast cancer detection. This is for the reason that breast cancer is a relatively infrequent disease, and it can be difficult to obtain high-quality mammograms of tumours.

Variability in mammograms: Mammograms can vary significantly in terms of quality and contrast, which can make it difficult for image processing algorithms to identify tumours.

Heterogeneity of tumours: Tumours can vary significantly in the form of magnitude, contour, and grain, which can make it difficult for image processing algorithms to classify tumours as benign or malignant.

Anyhow the use of image processing methods in the screening for breast cancer is an exciting new development. However, a numerous obstacle must be overcome before these methods may be extensively used. The development of more precise and reliable image processing algorithms for breast cancer diagnosis requires further study. In addition to the challenges mentioned above, there are also ethical considerations that

need to be taken into account when using Digital signal processing for breast cancer diagnosis. For example, it is vital to ensure that the privacy of patients is protected and that their images are not used without their consent. Despite the challenges and ethical considerations, image processing is a promising new technology that has the power to increase the premature detection and management of breast cancer.

### **1.1 Background of the Breast Cancer Diagnosis**

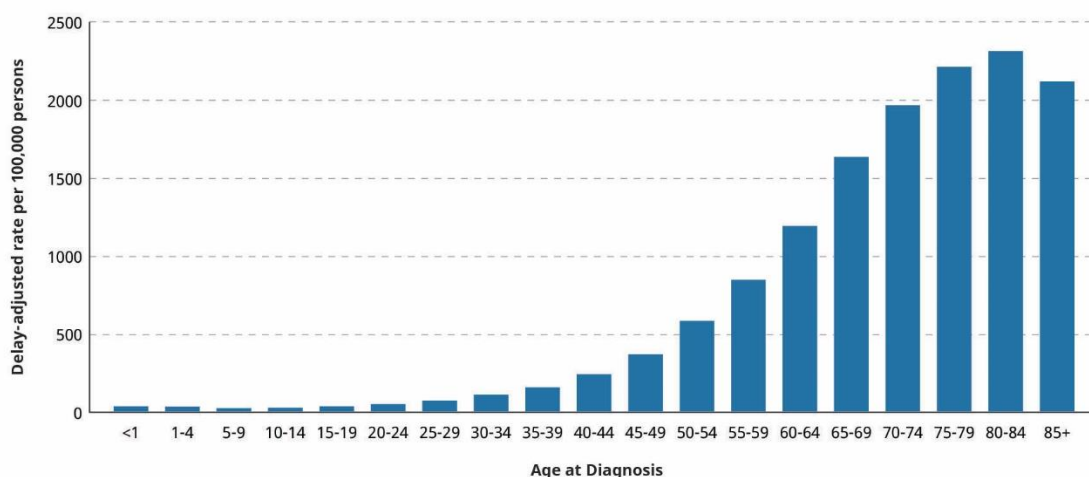
The most common kind of malignancy seen in women worldwide is breast cancer. The breast cancer mortality rate among women is rising overall, and this trend can be seen across all cancer types. At any time, any tissue inside the breast has the potential to develop cancer, which could lead to the amputation of the affected breast portion. Cancer cells most commonly originate in the breast duct; if they spread to other parts of the breast, such as the lobule or tissue, this occurs very rarely. Almost eighty percent of women diagnosed with breast cancer have a form known as invasive ductal carcinoma. This form of the disease indicates that the malignancy has spread to other breast tissues in the immediate vicinity. Breast cancer typically begins in the duct area and spreads by breaching the duct wall and reaching the fatty tissue, both of which increase the severity of the disease. Ultimately, this can cause the malignant cells to spread to other parts of the body through the lymphatic system.

Because of this, it is important for clinical professionals to undergo screenings for the identification and recognition of cancer at an early stage. To improve the chances of survival, appropriate prognostic measures are essential. These measures should minimize the need for breast removal as well as the adverse effects of chemotherapy and radiation. Since there is a risk to human life, we must use the most precise methods possible to identify potentially harmful cancerous cells using breast imaging. Computer-aided detection software is an examination technique used to assist radiologists in locating malignancies, which will reduce the number of incorrect predictions. This technique involves re-reading the image, which enables the radiologist to see details more clearly. As a result, the development of a computer-aided tool that may assist oncologists has been a topic of significant interest.

One of the developing subspecialties in the realm of medical applications is digital image processing DIP which is another branch of DSP digital signal processing technology. Acquiring images, storing them, processing them, communicating the



results, and displaying them are all crucial steps in the image processing workflow. Imaging procedures in medicine are beneficial for providing an early diagnosis as well as early illness detection. This thesis examines the possibility of locating and categorising breast cancer using a computer-aided strategy. Such a method might provide the radiologist with assistance in locating tumours in the breast using images obtained from mammograms and thermal imaging. The features that were extracted from a picture utilising a computer-aided design (CAD) system that was constructed may be used to classify the image as either benign or malignant. This determination is made based on the characteristics. This chapter includes a summary of the broad scope of the thesis, as well as its research aim, thesis statement, contributions, and organisational structure. It is the first chapter in the thesis.



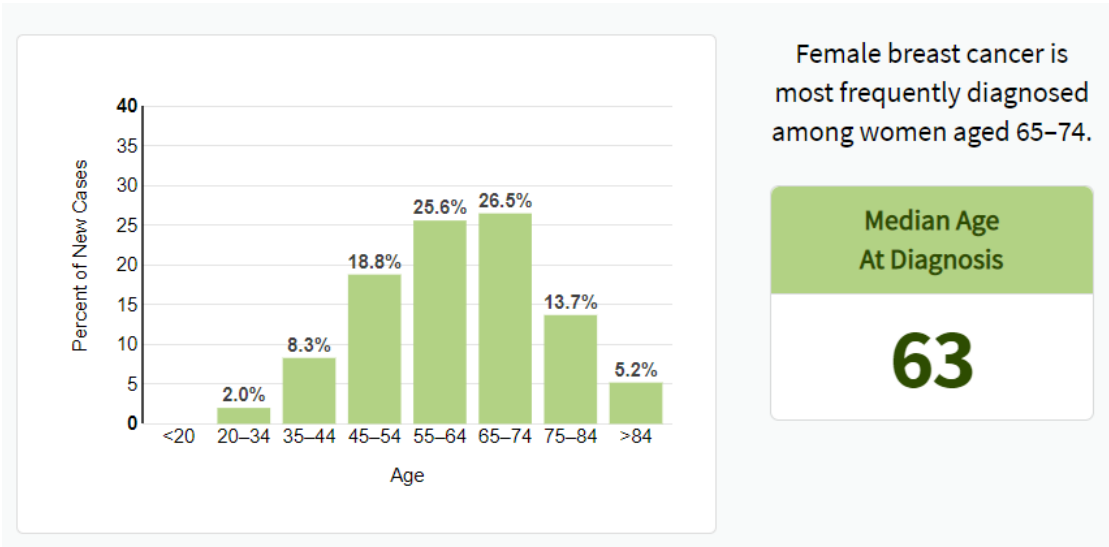
**Fig 1.1:** A chart showing various races, sexes of cancer diagnosis as per stats of SEER 21 2013–2017. Credit: National Cancer Institute [8].

It is seen from figure Fig.1.1, Growing older is the another most significant risk factor for developing cancer in general as well as for several specific forms of cancer. The overall prevalence rates for cancer ride up gradually as individuals age, from a rate of less than 25 incidents per lakh people in age groups under the age of twenty years to a rate of approximately 350 per one lakh people in age categories 45–49 to a rate of more than thousand cases per one-lakh people in age groups sixties and older.

## 1.2 BC- Breast Cancer Statistics

The development of female chest cancer, like the growth of other cancer, is caused by the interaction of a genetically vulnerable host and a factor in the

surroundings (external factor). The division of normal cells stops when it has occurred the required number of times. They do this by affiliating with other cells in the tissue, which allows them to remain in situ. When cells lose the capacity to stop proliferating, adhere to other cells, remain where they belong, and die when it is appropriate, this may lead to the development of carcinoma. A tumour in the breast, a change in the look of the breast, puckering of the skin, rejection of breast milk, liquid pouring from the nipple, a newly inverted nipple, or a small area of skin that is red or scaly are some of the warning signals that may point to breast cancer. Other warning indications include a change in the shape of the breast, a rejection of breast milk, fluid spilling from the nipples, and a change in the form of the breast (Ali Salem Ali Bin et al. 2017) [9].



At a Glance

Estimated New Cases in 2022	287,850
% of All New Cancer Cases	15.0%
Estimated Deaths in 2022	43,250
% of All Cancer Deaths	7.1%

5-Year Relative Survival
90.6%
2012–2018

Fig 1.2: Rate of New Cases and Deaths per one lakh women and % of fresh Cases by Age: Breast Cancer [10].

Survival rates are high among women detected with BC as shown in the Fig 1.3. On the other hand, it's possible that older women who are diagnosed with the condition have a higher risk of passing away from it than younger women do. In the America, female this cancer ranks as the 4th largest origin of mortality in the diseases overall. According to data collected between 2016 and 2020, the annual mortality rate was at 19.6 per one-lakh women as shown in the Fig 1.3 [8].

### **1.3 Signs and symptom of Breast Cancer**

The existence of a lump in the breast that is different in consistency from the breast tissue that surrounds it is the most common sign that a woman has breast cancer. More than eighty percent of the time, a malignant tumour will be discovered by a woman when she feels a lump with her fingertips. On the other hand, mammograms have the ability to detect breast cancer in its earlier stages (Boyd, N. F. et al. 2007) [13]. The presence of lumps in the lymph nodes that may be seen under the arms is another possible indicator of breast cancer. These nodes are located under the armpits.

Other signs of BC, in addition to the presence of a bulge, may comprise breast matter that has become thicker than just the remained of the breast, one side that has become bigger or lesser, a nipple that has changed stance or structure or that has become reversed, epidermis that has become puckered or cellulite, a reddishness on or around a chest-nipple, expulsion from body nipple/s, physical discomfort in a portion of the breast-position or underarms, and inflammation. The sensation of ache in the chest (also known as "mastodynia") is an inconsistent indicator of whether or not a woman has BC, but it may point to other concerns related to breast health (Boyd, N. F. et al. 2007) [13][14] (Gage, M. et al. 2012)[16]. It is considered that 5–10% of all cases are caused by genetics as the key factor. Women whose mothers were detected with the cancer previously the 50's had a more than one and a half times the chance of developing the disease in females whose parents were diagnosed at age 50 or later have a greater than one and a half times the risk (Gage, M. et al. 2012) [16].

The levels of hormones in a woman's body are one of the most noteworthy contributors to her hazard of rising spontaneous breast cancer. Estrogen hormone is a factor that contributes to breast cancer. Throughout puberty, menstrual cycles, and pregnancy, this hormone is responsible for stimulating the growth of breast tissue. During in the menstrual episodes, cell growth is brought on by an oestrogen and

progesterone discrepancy. In addition, oxidative metabolites of oestrogen have been shown to increase the risk of DNA alterations and degradation.

Patients who have the disease spread to their limbs may experience signs such as yellowing of the skin. Other symptoms may include the disease progressing to other organs. Starting menstruation at a younger age, having children later in life or not at all, reaching advanced years, having a personal or family history of breast cancer, and having a family history of breast cancer are all factors that can increase the likelihood of developing breast cancer. Other factors that can increase the likelihood of developing breast cancer include being overweight, not getting enough exercise, being addicted to alcohol, receiving hormone replacement therapy after menopause, being exposed to ionising radiation, starting menstruation at a younger age, and starting Having a personal or family history of the illness as well as having a family history of breast cancer are also additional factors that raise the risk that a person may acquire breast cancer in their lifetime [5]. An intrinsic genetic predisposition is responsible for around 5–10% of cases. This genetic predisposition may include mutations in BRCA genes, also known as the breast cancer gene, in addition to abnormalities in other kinds of genes. The cells that line milk ducts and the lobules that feed milk to these ducts are the most frequent areas where breast cancer starts. Lobules are another typical place where breast cancer starts. Other areas of the mammary gland are also potential starting points for the development of breast cancer. Cancers that develop inside the ducts are often referred to as ductal carcinomas, while cancers that begin within the lobules are typically referred to as lobular carcinomas. Ductal carcinomas are more prevalent. Both forms of cancer have the potential to be lethal. Breast cancer may be further classified into more than 18 distinct subtypes. These subtypes can also be further subdivided. Some forms of cancer, such as ductal carcinoma in situ, start off as pre-invasive lesions on some part of the body. A breast biopsy is the only method that can definitively diagnose breast cancer. This is because it is performed directly on the breast. After the suspicious tissue has been removed, it is subjected to a further inspection under a microscope. After a diagnosis has been made, more tests are performed to determine whether or not the cancer has spread to other parts of the body, as well as which therapies have the best chance of being successful in the long run.

#### **1.4 Literature on Diagnosis Tool:**

Mammography is gold standard as a breast cancer screening tool. It is one type of low-dose x-ray picture of breast used to detect breast cancer. However, mammography screening tool is not accurate and a mammography screening which show abnormal result may have a false negative or false positive rate about 1 in 10 women. It is specifically not accurate to the woman having dense breast tissue and near about 50% women undergoes screening have dense breast. There is also a risk of causing radiation induced breast cancer as younger women are more susceptible to effect of radiation compared to olden women. Clinical breast exam and self-breast exam is also used to detect breast cancer which are manual exams carried out by clinician or patient by self. Clinical breast exam performed by experienced clinician may use in detection of cancer that may not be detected by mammography.

A digital tomosynthesis is a 3D mammography which is a three-dimensional image. Tomosynthesis is typically performed along with mammography and due to this the radiation exposure to patient increases by twice. When tomosynthesis is used with the combination of mammography it increases the detection rate and reduces the false positive rate of mammography compared to mammography screening alone.

Sonography or ultrasound is another technique is used for further investigation to find suspicious area of breast in the mammograms to distinguish between cyst and solid masses [5]. Thus, sonography used in combination of mammography improves the sensitivity of mammography.

Breast MRI helps to find the size of the tumour, to find out the presence of other tumours, to check the spread of the tumour after detection of cancer, to check whether the chemotherapy is working or not and also as a part of follow up after removing the breast lump. MRI is not suggested as a single screening test for breast cancer detection because it has high false positive rate [5].

Though mammography is gold standard for breast cancer detection technique among all the available techniques, there is not available any single technique which is capable for detecting all the abnormalities of breast. Hence there is a need of more reliable technique. In many medical applications, thermography as a scanning tool has attracted attention. It is used to analyse the physiological function related to the body

temperature. It has capability of providing very important information of the breast that is used to detect early-stage tumours in the breast. As a scanning tool, thermography is safe as it is radiation, non-invasive, non-contact, as well, pain free as there is no compression of breast in between plates. The principle of thermography depends on black body radiation law which states that all objects above absolute zero temperature emits infrared radiations. Infrared thermography scanning for breasts is a screening method which searches for changes in temperature of the breast and captures them as an image. Due to the blood vessel activity and chemical in precancerous tissue, the temperature in the area surrounding the cancerous tissue is always higher than normal tissue. Hence, it needs a large amount of nutrition for their growth that result in increasing surface temperature of the breast. This indicates the presence of tumour which can only see through the thermography.

Thermography is an adjunctive technique that might potentially discover early signs of pre-cancerous and malignant cells eight to ten years earlier than other modalities for the early identification of these conditions. According to Aghdam et al. (2013) [36], finding cancer at an earlier stage is associated with a lower risk of mortality. It was predicted that areas of a patient's body that dried first would indicate underlying organ disease. This was observed by monitoring the distribution of slime on the patient. Over time, it has become evident that specific temperatures associated with certain regions of the human body are definitive indicators of both normal and abnormal physiological processes carried out by those organs.

### **1.5 Inspiration behind Choosing Research Area**

It is challenging to detect breast cancer using the extensive thermal dataset. Identifying breast cancer at an early stage is crucial. Thermographic approaches are used in the treatment of breast cancer. The temperature of a breast cancer patient's body varies in different sections due to the disease. Methods of pre-treatment and detection are employed to recognize breast cancer in its early stages. Optimization results in improved classification effectiveness. Breast thermographic images were categorized as either normal or malignant. The detection approach identifies the portion of the image with the most malignant characteristics with minimal pre-processing. The pre-processing and detection methods have formed the foundation of the research produced.

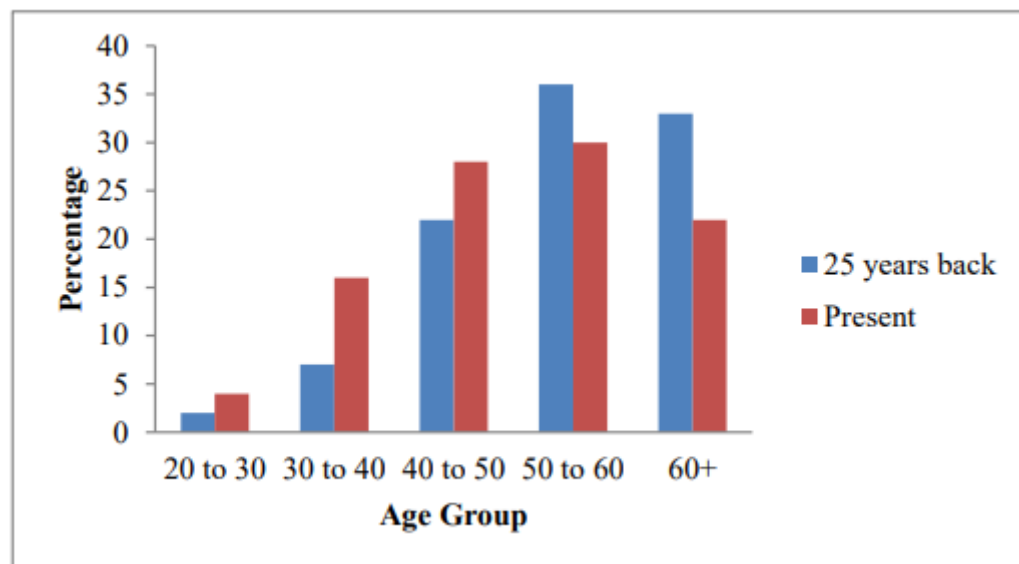
Current approaches do not improve PSNR, false negative rate, false positive rate, or detection time.

The method of image analysis was used to increase the overall lifespan of the person with disease via the procedure of early diagnosis. Therapy in the correct manner extends life expectancy. Digital mammography is one of the newer detection technologies that developed recently. An identification of the cancer in its earlier stages may be made using this approach. Yet, there were several downsides to using this strategy. Biopsy, M-R-I, and IR-imaging are some of the several alternative methods that may be used. Because of the shifts in frequency and intensity, it's possible that this approach won't provide accurate findings (Yasmin et al. 2013) [24].

Image processing served as the foundation for the development of a method known as Sequential Quadratic Programming (SQP). The observational source used to take the measurement, and the significance of the human being cannot be overstated. The approach was enhanced so that it could assess the parameters based on the highest temperature and the data obtained via the surface of the breast infrared images. In landing at a more correct estimation for the specified parameter, the unnecessary information is omitted. The addition of supplementary data based on the malignant data aides in the rapid diagnosis of illness. This approach does not make use of all of the data on malignant measurements acquired by thermogram (Bezerra, L. A et al. 2013) [25].

In order to accomplish the goal of recognising unregistered multi-view and multi-modal input mammograms and segmentation maps, a deep learning approach was made accessible as one of the possible options. In order to improve the accuracy of breast cancer diagnoses, this was done. This strategy exhibits the deep learning techniques that are used for feature selection methods and gives aid for classification via the utilisation of models that had a pre-trained set making use of computer vision datasets. Additionally, this strategy displays the deep learning approaches that are utilised for feature extraction methods. In addition, this methodology demonstrates the methods that are utilized in the process of picking features. The databases included millions of medical breast images that proved to be quite helpful when compared to one another. This technique is used as a standard for locating datasets and breast thermographic images that are available to the public. Due to the fact that it included

the interpretation of multimodal and unregistered Multiview medical images, this topic was considered to be of universal scientific interest in the area of healthcare. The FPR is not brought down by using this procedure, and the picture quality are worse (Gustavo Carneiro et al. 2017) [33]. This assisted the clinical person in obtaining data that was suited for their needs. This procedure does not have agreement since the information is not presented to the doctor in a clear manner, and these processes need to enhance both patients' and health professionals' engagement in order to be successful (Melissa Kool et al. 2018) [35].



**Fig. 1.3:** Statistics on Breast Cancer in India, Broken Down by Age Group  
(<http://www.breastcancerindia.net>)

In a nutshell, the traditional pre-treatment and optimization techniques proved to be effective in the premature finding of the cancer. Nevertheless, it has a number of drawbacks, the most notable of which are a poor PSNR, a lengthy clinical diagnostic time, a low cancer detection rate, and an inaccurate classification. Because of this, there is a pressing need for an effective strategy to enhance the process of early breast cancer identification.

The statistics facts on breast cancer in India is shown in Figure 1.3. Yet, once it relates to the breast area, the majority of people have a concept that women acquire breast cancer, therefore the debate mostly centred on breast cancer in women. Since both male and female have breast tissue, it is possible for it to occur in either gender. During the years 1993 and 1997, there was an alarmingly high incidence of breast



cancer. In India, there are around 356,256 women who have been tested positive for breast cancer. Dissimilarities in hormone intensities that arise during a lady's menstrual rotation may cause her breasts to feel misshapen, and studies have shown that about nine out of ten women develop breast lumps. Women in positions of age bracket of 45 to 49 years old are affected by breast cancer. In addition, there is a diverse population of younger women in their mid-40s that appears to have cancer. It is estimated that around 270 Chinese, Malays, and Indian women will succumb to breast cancer each and every year. This vast demographic group also includes other ethnicities, such as Indians. India has the highest rate of breast cancer incidence in Asia. This is mostly attributable to the impacts of urbanisation as well as changes in individual lifestyles. The process of urbanisation has directly led to this effect. Mammography is the usual method that is now being used for the typical approach that is being employed today for the identification of cancer by clinical examination. This method is utilised for the detection of breast cancer. In spite of the fact that cancer is sometimes found in persons with thick breasts, which is a source of tension for folks, we want an additional system that may assist the radiologist in avoiding making incorrect predictions so that we can alleviate this stress. The patterns at which these systems are applied are not only implausible for the many different kinds of information, but they also cannot be replicated. This technique does not work for women since it produces inaccurate results when used to women with enormous breasts.

## **1.6 Research Objectives**

The primary purpose of the study that was carried out for this thesis was to build a framework that was based on the characteristics of breast cells to enhance the earliest diagnosis of the cancer and categorise any anomalies that were found. The framework is meant to detect signs of malignancy by extracting characteristics from mammograms and thermal imaging. These images may be used. The study is carried out with the purpose of identifying cancerous tumours in breast thermographic images acquired through DITI.

1. To test and analyse alternative entropy definitions and recent robust divergences for classification of breast lumps or tumours as benign, pathological or suspicious using thermographic images.

2. To apply and analyze deep learning approaches for the classification of breast lumps or tumours as benign, pathological or suspicious using thermographic images.
3. To employ a novel heuristic optimization technique for hyperparameter tuning in deep learning algorithm
4. To evaluate the performance of the proposed solution for the classification of breast lumps or tumours as benign, pathological or suspicious using thermographic images.

### **1.7 Research Methodology:**

The basic framework for BC detection through thermographic images has four stages: image pre-processing, segmentation, region of interest (ROI) extraction, feature extraction and classification with three classes identified as normal denoting without any tumour, benign denoting with non-cancerous tumour and malignant denoting cancerous tumour. Various approaches differ based on the approaches used at various stages.

The thermogram raw images utilized for the analysis of breast cancer are acquired from standard resources, and they are directly provided as the input to pre-processing stage. The raw images are processed in the pre-processing phase using grayscale conversion, adaptive mean filtering and contrast enhancement. Then, breast segmentation is performed with the help of GVF (Gradient Vector Flow) to attain segmented breast images. To exploit the robustness against outliers' characteristics of Beta divergence we have uses them as features. The classifier outcomes in DNN highly depends upon the used features. Motivated from the fact that Beta divergence has a proven characteristic of robustness, we propose use of it as features, specifically in medical image, for classification and other machine learning tasks. HHO algorithm imitates the hunting mechanism of Harris Hawks. They follow various steps and mechanisms to hunt in a group, instead individually. Repeated mechanism of exploration and exploitation boost HHO performance. On the contrary, JA is simple and requires no additional parameter for initialization. This motivates to combine the community-based hunting mechanism of HHO and the simplicity of JA to achieve the best optimization algorithm which would be more efficient at finding the optimal. The

novel algorithm is identified as HHHJO algorithm used for tuning the hidden neuron count of DNN, the learning rate of DNN and DNN epochs count.

The thesis work proposes a solution balancing both accuracy and computational requirements by providing the accuracy equivalent to those of the deep learning and computations equivalent to the ML algorithm.

## **1.8 Research Contribution**

The work has the following novel contributions:

1. Beta divergence has a proven property of robustness and is been explored successfully in various applications in Robust Statistics, as well, Machine Learning. The thesis work successfully employs entropy definition induced by Beta divergence as features for the first time in a medical diagnosis and classification task. It gave almost 5% better accuracy than other features, including conventional entropy definitions.
2. A novel hybrid meta hubristic algorithm, identified as HHHJO, has been developed using Jaya Optimization to improve the local search and HHO algorithm to improve the global search based on community-based hunting approach.
3. The thermographs were segmented using gradient vector flow snakes (GVF) algorithm followed by an Optimized FCM (OFCM) algorithm, where the optimal hyperparameters were derived using the newly designed HHHJO algorithm identified effective.
4. A novel Optimized DNN (ODNN) with hyperparameters tuned using HHJO has achieved better detection for thermographic BC detection.
5. A novel solution combining ODNN with Beta entropy and other features, GVF (Gradient Vector Flow) and OFCM for tumour segmentation and HHHJO for hyperparameter tuning in DNN and FCM has been developed for BC detection using thermograms. The solution achieves almost 99% accuracies, comparable to that of CNN, just with two hidden layers.

## 1.9 Thesis Data Flow

The remaining chapters of this thesis are

**Chapter 2**, a detailed assessment of several current methods and techniques for breast cancer early detection is presented. These methods and techniques include deep learning strategies.

**Chapter 3** Explains fundamentals of Information theory entropy and divergence and techniques useful in machine learning and data science.

**Chapter 4** offers the HHHJO method. The suggested HHHJO aims to obtain better results in terms of increasing entropy while simultaneously reducing variance as much as possible. After that, the abnormality segmented breast thermographic images, where the features were extracted with the assistance of GLCM, first-order and second-order textural descriptors, LBP, entropy feature, and beta entropy. Some parameters, such as the learning rate of DNN, the epoch's count of DNN, and the concealed neuron sum of DNN, were tuned by utilising developed HHHJO in order to maximise accuracy and provide an effective classification rate.

**Chapter 5** presents the overall solution. The suggested deep neural network based proposed novel solution is described for the detection of breast cancer to detect whether it is normal or cancerous.

**Chapter 6** presents the observations of the proposed methods HHHJO and compares with existing methods. The performance metrics are evaluated with respect to CA, FPR, FNR, Precision, Sensitivity, NPV and MCC. This thesis's most important findings and conclusions are outlined in Chapter 6, along with prospective directions for further research.

**Chapter 7** Conclusion, discussion and future scope

## 1.10 Summary

This Chapter included an explanation of the research's context, although in a condensed form. In addition to this, the reasons for doing this study, a description of the issue, research contributions, and the arrangement of the thesis are discussed.

## CHAPTER 2

### THE PROBLEM STATEMENT AND LITERATURE SURVEY

---

To improve the survival rate early detection is necessary. The problem is majority of women are diagnosed with later stages and survival at these stages is difficult. So, there is a need for early diagnosis which is possible by making the people aware of early signs and symptoms. Also, there is a need for accurate diagnosis, easily available and cost-effective treatment. The target of the research is to focus on reducing the cancer death by creating the awareness among the people so that it will help them to detect the symptoms at early stage. Though mammography is best for breast cancer uncovering technique among all available techniques, there is not available any single technique which is capable for detecting all the abnormalities of breast. Hence there is a need of more reliable technique. In many medical applications, thermography as a scanning tool has attracted attention. It is used to analyse the physiological function related to the body temperature. It has capability of providing very important information of the breast that is used to detect early-stage tumours in the breast. As a scanning tool, thermography is safe as it is radiation, non-invasive, non-contact, as well, pain less as there is no compression of breast in between plates. The principle of thermography depends on black body radiation law which states that all objects above absolute zero temperature emits infrared radiations. Infrared thermography scanning for breasts is a screening method which searches for changes in temperature of the breast and captures them as an image. Due to the blood vessel activity and chemical in precancerous tissue, the heat signature in the area neighbouring the bad tissue is always higher than common body-tissue. Hence, its prerequisites a large quantity of nutrition for their growth that result in increasing surface temperature of the breast. This indicates the presence of tumour which can only see through the thermography.

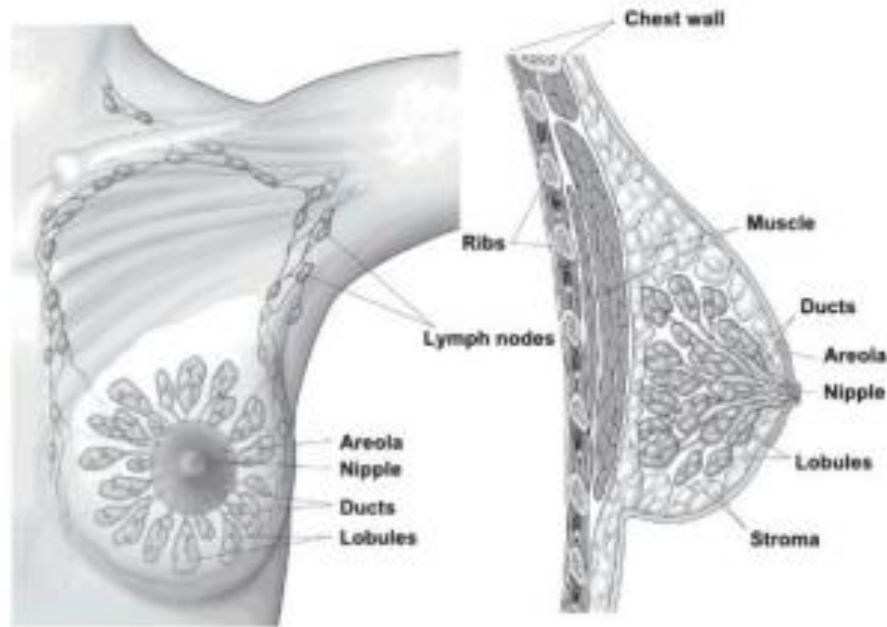
#### 2.1 Review on Available Literature

The most awful illness, cancer, is really a collection of disorders that are tied to one another. With every kind of cancer, some tissues in the body begin to divide in an unrestrained manner and blow-out the disease to the tissues in the immediate area. A diverse illness is what we mean when we talk about cancer. In order to assist the

subsequent clinical evaluation of individuals, the early detection and prevention of a cancer type are essential. The premature finding of cancer may be aided by determining the protein sequences and structural motifs that are present in a cancerous cell. The discovery of cancer is dependent on investigation at the molecular level, and a systematic, accurate, and accurate diagnosis is recommended for the many different forms of cancer.

The human body is responsible for the development of a variety of cancers. BC is one of the greatest common causes of significant concern for health in females. Because of the way their bodies are structured, females possess a greater risk of catching cancer around chest than males do. Aging, having a history of the developing cancer around breast in one's family, having dense breasts, being overweight, and drinking alcohol are all potential causes of breast cancer. In order to identify cancer from breast thermographic images from a variety of angles and vantage points, researchers are using a variety of detection and classification strategies. These strategies are differentiated by the stage of the illness and the quality of the images. On the other hand, there is still an alarmingly high percentage of incorrect breast cancer detection. Inaccurate diagnoses of cancer may lead to overtreatment, while failure to recognise malignant tumours might result in inadequate treatment. Both of these outcomes are undesirable. As a result, a handful of study articles are being studied and investigated for the purpose of cancer identification at an earlier period.

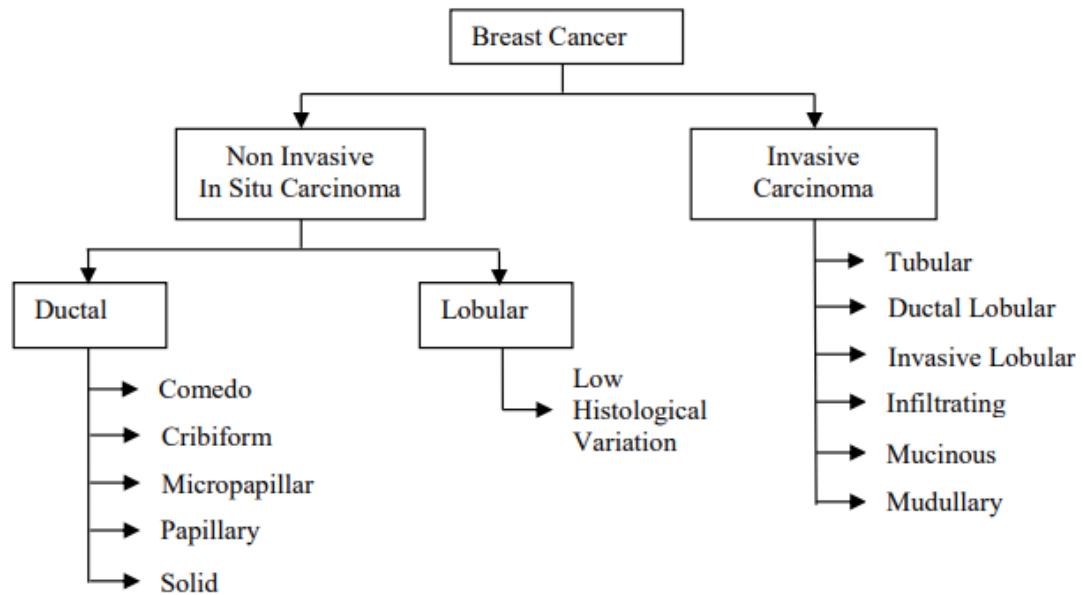
The lump that is already present in the cell might become cancerous if it invades surrounding breast tissue or spreads across it. Breast cancer may be more frequent in women, but males may still be affected by the disease. Cancer of the breast will often begin in the ducts that are responsible for milk production and then spread to the nipple area. This kind of cancer is referred to as ductal carcinoma, while cancer that begins in the glands that are responsible for milk production is known as lobular carcinoma. Sarcomas and lymphomas, both of which may be seen in various breast tissues, are considered to be benign tumours. Lumps in the breast are not usually a sign of cancer; sometimes they are benign. Benign tumours are abnormal growths that occur in the breast but do not feast to other parts of the human body.



**Fig. 2.1:** Cells in blood flow or the lymph system spread breast cancer.

Small and bean-shaped, lymph nodes are part of the immune system and are responsible for storing tissue fluid and debris. Cancer tissues move through the lymph-node network and begin to develop in the lymph-network of human body. Axillary nodes, supraclavicular and infra-clavicular nodes, and chest nodes around the breast constitute these lymph nodes (internal mammary lymph nodes). Normal cells develop cancerous owing to DNA abnormalities, genetics, or lifestyle factors, but the specific cause is unknown.

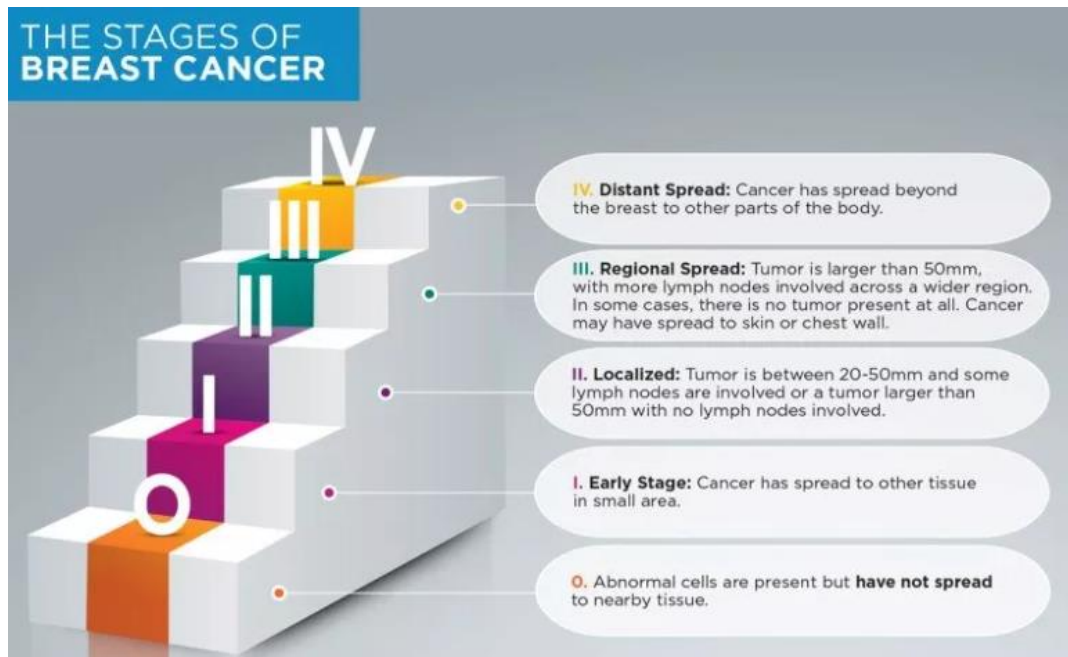
As shown in Figure 2.2 both invasive and non-invasive carcinomas can be further subdivided into subcategories, one of which is adenoid cystic carcinoma. Apocrine carcinoma, Cancer that has a stromal giant that looks like osteoclasts, Invasive lobular carcinoma, Cancer showing features of the medullary glands, Metaplastic carcinoma, Invasive micro-papillary carcinoma, Mucinous carcinoma, Neuroendocrine carcinoma, Tubular carcinoma (Horlings et al., 2013) [37].



**Fig. 2.2:** Classification of various B-Cancers reported in various countries time to time.

Around 30% of malignancies are detected early. Breast cancer has four stages. Stage I is just DNA changes, Stage II is blow-out to the axillary lymph nodes under the arms, Stage III is creeping to the lymph-network near the chest bone, or on the wall of the chest and/or the membrane of the breast with or without evidence, lymph nodes above or below the collarbone, and Stage IV spreads to other organs, including the brain, liver, lungs, and bones [38].





**Fig. 2.3:** Breast Cancer Imaging

There was discussion of monitoring for early breast cancer diagnosis in conjunction with classification in order to identify breast cancer. This study led to the development of a wide variety of other medical research projects as well as medical procedures. The detection rate of abnormal cancer portions was low during categorization. It has a very poor sensitivity for screening. Since the operation was employing a bigger database, which makes it more difficult to identify a suitable one from the database, the development of an early detection sign has shown to be highly beneficial. During the period of the study, there was a consistent detection rate for either the interval of cancer or the subsequent screening round. This technique exhibited an unacceptably high FPR and an excessive amount of invasive medical testing (Mordang JJ et al. 2017) [87].

An accelerator that is used in microwave imaging for the purpose of diagnosing breast cancer was provided as a means of fulfilling the conditions for the unique design in the field of space exploration. The two microwave imaging algorithms that are used for the early detection of breast cancer are put through their paces here. This is one of the ways that breast cancer may be detected. The design recommendation arrives at the conclusion that beam forming is still an attractive outcome designed for low-cost installations. The work that was available provided dependable adaptability and a high degree of adaptability to changes in the parameters of the system, but the high-level

synthesis provided productivity improvements in design. While there was an improvement in terms of power usage, no analysis was done on accuracy since we were considering two algorithms and synthesis at the same time (Daniele Jahier Pagliari et al. 2017) [3].

It was decided to use an algorithm based on machine learning in order to rise both the exactness and sensitiveness of the cancer decision process. In improving the accuracy of the forecast, this approach takes into account six accelerators rather than the more traditional two accelerators. The findings of each of the machine learning algorithms are shown to have a high level of performance. The results of these were obtained using binary classification algorithms. The approach was helpful for forecasting what breast thermographic images would be normal and what would be aberrant. This also helps a great deal in reducing the categorization difficulty, but at the expense of the findings' accuracy (Abien Fred M et al. 2018) [12].

## **2.2 Feature Extraction for Breast Cancer Detection**

In Acharya, U R et al. (2012) [6] attempted to verify the rate of the thermal imaging apparatus for the identification of the cancer. Their goal was to decide whether or not the instrument was effective. The infrared breast thermographic images were acquired from Singapore General Hospital and utilised in validation process. The run distance and the co-incidence matrix were used in order to achieve the features of the image's texture. After that, these characteristics were sent to the SVM classifier so that it could automatically differentiate between cancerous and normal breast states. As a result, the newly constructed model achieved a higher rate of accuracy in comparison to both its specificity and its sensitivity.

A supervised model for identifying the mitotic signature in breast histopathology Full Sliding breast thermographic images was reported by Monjoy Saha et al. (2018) [47] (WSI). A deep learning framework was built, and characteristics that were handmade and obtained through prior medical difficulties were employed in the design process. The convolution layers, the 4 max-pooling sheets, the 4 rectified linear units (ReLU), and the 2 fully linked sheets were the components that made up the deep learning framework. The morphological, textural, and intensity aspects that were handcrafted made up the handcrafted features. Recognition and evaluation of the stage of breast cancer have both been given a greater degree of precision.

Subrata Kumar Mandal et al. (2018) [48] addressed several methods for identifying breast cancer, including information extraction, feature collection, feature taking out, data discretization, and categorization. In order to get a cancer classification that is very accurate, the lowest group of characteristics possible was identified. A number of different techniques to cancer classification were investigated, and the amount of time required by each classifier was calculated and examined. Based on the findings of the study, it was determined that the logistic-regression-classifier is the one that achieves the greatest results and has the highest level of accuracy. On the other hand, there was no attempt made to determine the precise stage of the patient's breast cancer. In Al Fayez et al. (2020) [50] performed preprocessing with the assistance of adaptive histogram equalisation, top-hat transform, and homomorphism filtering; achieved segmentation of ROI with K-mean clustering; used binary masking, which was utilised to extract the features presented in signature boundary; and classified the data using Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM). Al Fayez et al. (2020) [50] also used binary masking, which was utilised to extract the features presented in signature boundary. The newly developed model was calculated making use of a dataset that is used often, and the ELM-based procedures achieved improved outcomes in comparison to those of the models that are already in use. Ekici and Jawzal et al. (2020) [51] applied CNN in order to successfully accomplish the task, and as a result, they were able to achieve an accuracy rate of 66%.

Automatic segmentation based on Otsu's approach and automated network recognition were the goals of Sánchez et al. (2020) [56]. A strategy to classification that makes use of bispectral invariant was proposed by Gomathi et al. (2020) [43]. These classifications were carried out in breast thermal images utilising an unsupervised anisotropic-feature transformation method to efficiently categorise characteristics such as benign, standard types, participants, and malignant. The created model was successful in accurately classifying both abnormal and normal women. The proposed framework reduces the amount of noise that is visible in the picture, which results in an improvement in image quality.

Schaefer et al. (2014) [44] suggested a breast thermogram analysing model to collect the set of bilateral symmetry image characteristics among right and left breast areas. These features were then employed for the classification phase of the research.

Throughout the classification procedure, the Ant Colony Optimization (ACO) based pattern identification algorithms were applied. This particular classifier was applied in order to examine the produced picture characteristics. The traditional Ant-Miner classifier was able to attain a higher classification performance rate based on a briefer rule. The results of the simulation have shown that the model that was constructed is capable of accurately extracting picture characteristics and also providing an efficient classification rate.

Automatic anomaly identification in a breast thermogram was first attempted by Francis et al. (2014) [53]. Their method included performing a curvelet transform according to features extraction methodologies. In order to carry out automated classification, the texture as well as the statistical characteristics were extracted from the thermogram in the curvelet basis and presented in SVM. Both the suggested classifiers and the textural characteristics are capable of recognising abnormalities in breast thermograms. The proposed classifiers were able to efficiently recognise aberrant thermograms with a high accuracy rate. Yadav and Jadhav et al. (2020) [52] have developed a software application that uses machine learning as a statistical approach to learn the characteristics without the need for direct coding to be performed. The primary purpose of the computational intelligence algorithms was to learn how to read the thermal scans and identify the potentially suspicious region. Thermal photography was a somewhat superior method to those that were used before, and it was simple enough to be used in medical facilities such as clinics and hospitals. An improved performance rate has been achieved by the CNN-based CAD model that was built in terms of enhancing data augmentation, network intricacy, and fine-tuning in the breast cancer dataset.

Images obtained by logical thermography were shown in order to diagnose breast cancer. This technique investigates the prospect of a rational thermograph and contributes to the development of automatic identification. Both the first-order statistical feature and the Haralick texture feature were derived from the data in the spatial domain. These characteristics were gleaned through a sequence of reasonable thermograms taken from the stage that came before it as well as the present standing of the application challenge. This technique brings about a considerable reduction in the internal body temperature while having no impression on the breast normalcy detecting

system. In order to determine how the procedure stacks up against other methodologies already in use, a comparison was carried out. Ultrasound is the auxiliary tool that is employed here. This approach proved beneficial in determining which areas of the body had cancer and which did not have cancer. This method does not provide any mechanisms that are dependable or beneficial for the automated detection algorithm (Sheeja Francis et al. 2014) [33].

### **2.3 Region of Interest (ROI) Based Breast Cancer Diagnosis**

The procedure of classification is carried out through optimization. The breast thermographic images are allocated into two groups: healthy and sick, grounded on the categorization method. The scaling of the picture and the production of an image of good quality are both important functions that need pre-processing. The breast thermographic images that are created as a consequence of the optimization approach yield accurate results. The optimization procedure results in an increase in PSNR as well as a decrease in FPR (False positive rate).

In order to improve the overall efficacy of breast cancer discovery for the goal of early illness diagnosis, a number of pre-processing and detection procedures are used. The development of effective improvement and identification strategies is now the subject of a significant amount of study. The discussion will focus on some of the more recent relevant efforts that have been done on optimization and detections. For the purpose of early cancer identification using thermal imaging, a bio spectral invariant feature selection approach has been developed. The characteristics that have been chosen are insufficient for making an accurate finding of the cancer (Mahnaz Etihad Tavakol et al. 2013) [17]. In order to gain superior performance and analyse breast thermograms based on picture characteristics while maintaining the same level of prediction performance, hybrid multiple classifier systems were developed and implemented (Bartosz Krawczyk et al. 2016) [18]. A failed pre-processing attempt is made via a new weighted algorithm that is provided in the Naive Bayes Classifier (Shweta Kharya et al. 2016) [64]. The Mega-Trend Diffusion (MTD) method was developed for the purpose of spotting the cancer in its premature stages. According to Abdul Majid et al. (2014) [19] the accuracy of cancer prediction was not enhanced by integrating a greater number of protein sequences or feature spaces.

Random Under-Sampling Boosting or RUSBoost, introduced to deliver an effective taxonomy of breast imageries by conjoining weak classifiers into solid ones with less dependability (Ehsan Kozegar et al. 2017) [21]. RUSBoost was developed to achieve this goal by integrating weak classifiers into superior using random under sampling. The Stacked Sparse Auto Encoder, or SSAE, was developed so that DNA identification on high-resolution images could be done more effectively. There is a decrease in accuracy and dependability (Jun Xu et al. 2016) [21]. With the purpose of reducing the error rate as much as possible in the breast cancer detection system, a three-dimensional convolutional neural network and prioritised candidate aggregation were provided. Dynamic neural network from the thermal pattern was developed, which helps a great deal for lowering the mortality rate of women. The system's degree of precision was not investigated (Tsung-Chen Chiang et al, 2019) [22]. These currently available techniques take into account a number of preprocessing and optimization procedures in order to detect the cancer at an earlier time frame. The performance of these approaches is negatively impacted by a variety of problems, including a lengthy period of time required to diagnose breast cancer, a low PSNR, and a high FPR.

Wenqing Sun et al. (2016) [58] presented a technique for the uncovering of cancer that was established on graph-based Semi Supervised Learning (SSL) and made use of deep Convolutional Neural Networks (CNN). For training purposes as well as for fine-tuning the variables, CNN needs a substantial quantity of labelled data. The labelled data in the training set were only required to make up a tiny portion of the total, and the diagnostic system was expanded to include more modules. Many regions of interest (ROIs), each of which was comprised of mass mined from mammography images, were used for the purpose of analysis. The results obtained from some of the ROIs were categorised as labelled data, whereas the results obtained from the other ROIs were termed unlabelled. Yet, there was little thought given to improving the SSL protocol's performance.

Isikli Esener et al. (2017) [59] proposed a useful feature ensemble with multistage organisation in a CAD methodology for the cancer finding in breast cancer. The mammography images were gathered into a dataset that was made publicly available, and a multistage classification was carried out. The mammography underwent the process of having its features extracted. The histograms of the ROI

images are then equalised using nonlocal means filtering as part of the pre-processing. The ensemble of features was produced by combining local configuration pattern-based features with quantitative and frequency domain characteristics. After that came the phase of classifying these traits according to their respective categories. The multistage classification procedure was carried out with the purpose of performing a better breast cancer diagnosis while simultaneously consuming a significant amount of time.

## **2.4 Optimization Techniques for Breast Cancer Detection**

Particle Swarm Optimizer (PSO) is analysed by Mei-Ling Huang et al. (2012) [60]. PSO is dependent on Artificial Neural Network Adaptive Neuro-Fuzzy Inference System and a Case-Based Reasoning classifier that makes use of both a decision tree model and a logistic regression model. Classification approaches were utilized on the mammographic mass data set, which resulted in an improvement to the classification as well as the identification of additional mistakes. A constructed model of ANN known as ANFIS was used. It investigates the features of resilient and truthful learning with the assistance of both language and data sets, and it has a greater capacity for generalisation than other methods. In order to chart formerly extracted words and events from files, the rule-based ANFIS approach might be helpful. Also, it was used to aid in the decision-making process for oncologists with less expertise.

A support vector machine classifier that is built on swarm intelligence techniques (PSO-SVM) was introduced by Hui-Ling Chen et al. (2012) [61] as a structure for the revealing of the cancer. Under the framework of the particle swarm architecture, both the issue of model collection and the issue of feature collection were handled. Next, in order to build the objective of PSO, a weighted function was utilised. This function takes into consideration the specified features, the number of support vectors (SVs), and the average accuracy rates of the support vector machine (SVM). In addition to that, the time-varying acceleration coefficients and the inertia weight were included into the PSO algorithm so that it could successfully maintain both the local and the global search. This was done in order to maximise efficiency. The right model parameters and a discriminative feature subset were readily retrieved with the aid of a decreased collection of SVs for training, which resulted in higher prediction accuracy. This was achieved by reducing the number of SVs used for training. Categorization was completed in a manner that was both accurate and effective thanks to the contribution

of five informative qualities. In consequence, this assisted to reach better diagnostic judgements in clinical breast cancer diagnosis despite the increased amount of time spent classifying patients.

In order to accomplish efficient gene selection throughout the process of identifying cancers Kun-Huang Chen. et al. (2014) [62] integrated PSO with a decision tree and produced a new algorithm. A comparison was made between the SVM self-organizing map, the backpropagation neural network, the C4.5 decision tree, the Naive Bayes algorithm, the CART decision tree, and the artificial immune recognition system in order to assess the effectiveness of each.

Krawczyk, B., & Schaefer, G. et al. (2014) [82] article presents the Ant Colony Optimization (ACO) classification for used while evaluating breast thermograms. The thermographs illustrate the functional information of cancer, which may be used to spot the cancer at an earlier phase. This approach employs the ACO method in order to get the desired categorization outcomes. On the basis of the characteristics that were chosen in the earlier sections, the breast thermograms were sorted into two categories: normal and malignant. The colony ant miner method was used as the foundation for the feature selection operations. A heuristic based on ant colonies, combined with an on-the-fly entropy standard, is used in this method. The discretization is carried out. ACO suffered from various inadequacies in its probability distribution, and accurate diagnostic findings could not be discovered (Krawczyk, B., & Schaefer, G. et al., 2014) [82].

## **2.5 Classifiers for Thermography Breast Cancer Identification**

M. M. Mehdy et al. (2017) [63] conducted a review that focused on the use of neural networks (NN) to medical imaging procedures. A look was taken at how neural networks may be used in a variety of medical imaging techniques. After that step, the categories of NN with the various categories of feeding data were examined. In furthermore, the use of hybrid NN adaptation in the identification of the cancer was investigated.

In study of Kharya, S. & Soni, S. et al. (2016) [64] compared the superiority requirements of a machine learning techniques known as the Naive Bayes Classifier to a novel weighted technique for classifying BC. Most fruitful methodologies to classification are called the Naive Bayes system. Instead of focusing on categorization,



the ranking performance was deemed to be the key notion. The traditional implementation of the Naive Bayes algorithm saw improved results with the incorporation of the weighted notion. Despite this, the medical prognosis was not carried out in an effective manner.

Nicandro et al. (2013) [65] introduced Bayesian network classifiers with the purpose of evaluating the investigative potential of thermography in the cancer. The information that obtained from the thermal picture was used to portray people who had cancer. The data that was obtained from the participants and used to distinguish unwell people from otherwise healthy individuals ultimately determined a score. In addition to this, Bayesian network classifiers were used as an efficient supplemental tool for the diagnosing process.

An approach for an aggregation that is based on support vectors was described. This technique diagnoses cancer by exploiting the support vector machine (SVM) algorithm as a learning prototypical. This approach is the norm in the composite algorithm for the area beneath the characteristics curve. The variety of the basic model set was improved with the help of the C-SVM and v-SVM prototypes, each having six different support vectors. The Area That Will Be Weighted comparison is made between the five different fusion procedures and the receiver operating Characteristic Curve Ensemble (WAUCE). They were conceived with the purpose of deriving the choice from the foundational models. The outcome is assessed using two standard datasets in addition to one substantial actual dataset. This was determined by looking at the model's efficiency as well as its reliability in terms of its performance. Conferring to the discoveries of Haifeng Wang et al. (2017) [34], whereas WAUCE significantly boosts cancer detection, the variation involved in these techniques was shown to be reduced. The identification of breast cancer was accomplished by the use of multiple logistic regression analyses. The accuracy of the data that was self-reported by the patient about the sentinel node biopsy and the auxiliary lymphatic nodal segmentation is much less and only exhibits modest conformity.

In an article Ebrahim Edriss Ebrahim Ali et al. (2016) [66] provided a description of the tests that were carried out with the assistance of the hospital registry in order to classify the cancer. For the determination of doing classification on breast cancer datasets, the Supervised Learning Algorithm known as SVM was used in

conjunction with kernels such as Linear and Neural Network (NN). In contrast to the SVM method, the Neural Network methodology provides more "accurate" and "precision."

Gustavo Carneiro et al. (2017) [33] automated the investigation of unregistered Cranio-Caudal (CC) and Medio-Lateral Oblique (MLO) mammography images. This determined the patient's breast cancer risk. Deep Learning models classified unregistered mammography images and breast lesion segmentation maps. This comprehensive technique categorised mammographic exams and segmentation maps. The automated technique accurately used segmentation maps from automated mass and micro-classification detection systems. Automatic detection did not lower FPR.

Abien Fred M. Agarap et al. (2018) [68] evaluated ordering test accuracy, sensitivity, and specificity using ML algorithms on the (WDBC) dataset. Digitized images of breast mass FNA tests form the dataset. All classifiers were hyper-parameterized and classified. ML algorithms boost classification test accuracy.

Maryam Mahsal Khan et al. (2018) [69] described Evolving Wavelet Neural Network ensembles. The fitness function embedded classifier accuracy alone and as part of the ensemble to find the best ensemble. A higher CA island-based model reduced training time. This model was created by studying migration topologies.

Cuong et al. (2013) [70] used accidental forest classifier and piece collection to diagnose and prognosticate breast cancer. Weighting condition solved diagnostic problems and kept important characteristics by removing extraneous dataset elements. Random forest classifiers improved classification accuracy and addressed various breast cancer diagnoses.

Higher-order spectrum properties were derived from the thermograms and shown in order to facilitate the categorization of thermograms into normal and abnormal categories. The higher-order spectral features were retrieved and put into a Feed-Forward Artificial Neural Network (FFANN) classifier in addition to a Support Vector Machine (SVM) for the purpose of increasing the accuracy and precision of the analysis. The ANN classifier that follows it also provides stronger values of sensitivity and specificity than the one that came before it did. A more accurate automated classification of normal and abnormal breast thermograms was carried out without

resorting to any form of subjective inspection. This allowed for a higher level of diagnostic precision. Using this technique, thermograms as a diagnostic tool for automatically diagnosing breast cancer cannot be demonstrated (U. R. Acharya et al. 2014) [32].

## **2.6 Deep Learning Approach for Classification:**

It was shown that deep learning based on histopathological image categorization might be used as a technique for premature identification of this cancer utilising the deep learning technology. The clinical histopathological tissue slides are of abundant use in the classification of the various subtypes of this cancer. Both the microarray tissue dataset and the ation of the various subtypes of this cancer. Both the microarray tissue dataset and the Break His dataset were used in the collection of these data. They were helpful in distinguishing normal cancer cells from aberrant cancer tissues. In order a relatively tiny window. This approach lands into account the sparseness of the datasets, cuts down on the total number of measures, and speeds up the training process. In addition to that, it prevents overfitting. The dataset also considered more extensive and deeper sparse networks, which were able to do analysis on the extensive and complicated dataset. The subsequent layers were able to obtain the performance about the information that was compared to the subsequent high-level levels. According to Mehdi Habibzadeh and his colleagues' research Carneiro et al. (2018) [33], despite the presence of picture complexity in the dataset, both techniques performed better for digital image processing.

Mohammad A. et al. (2019) [42] propose a novel method to model the changes on temperatures in normal and abnormal breasts using a representation learning technique called learning-to-rank and texture analysis methods. The proposed method generates a compact representation for the infrared images of each sequence, which is then exploited to differentiate between normal and cancerous cases. They introduced a model the changes in breast temperatures during the dynamic thermography procedures using the LTR and texture analysis methods. This method generates a compact yet descriptive representation for the whole sequence of thermograms of each case, then input the representation extracted from normal and cancerous cases into the MLP classifier to build a classification model. The proposed method obtains outstanding classification results in terms of AUC, accuracy, recall, precision and  $F$ -score.

Ekici and Jawzal et al. (2020) [51] developed software to automate breast thermograms' BC detection. Their method consists of a CNN optimized by a Bayes Algorithm. The dataset used had a cohort of 140, of which 95 were benign and the rest malignant. Due to the class disparity, the data augmentation technique was used to balance the two. Also, object-oriented image segmentation was used for the ROI extraction. The model's performance was then compared with those from the literature on the same dataset from which some used statistical features but used various classifiers, namely, SVM, ANN, RBF, etc. Dataset used was very small. More number of images should be used for classification algorithm.

Soner C. et al. (2023) [42] propose to utilize Mask R-CNN technique on images by first assigning bounding boxes and then creating a border for each tumour volume to differentiate it from adjacent tissues and structures. The results indicate that the classification and segmentation performances of Mask R-CNN method on ResNet-50 architecture are better than the data reported in the literature for thermal breast image studies. Two network architectures (ResNet-50 and ResNet-101) were trained for 60 epochs and were then evaluated based on their classification and segmentation performances. Detection and segmentation performances of this model were also higher with mAP of 0.921 and overlap score of 0.868. In this manner, thermal images handled by a single DL model to successfully perform detection, classification, and segmentation of normal and abnormal breast tissues. There is problem of resemblance arises that causes overfitting. More number of images to be used to avoid the problem.

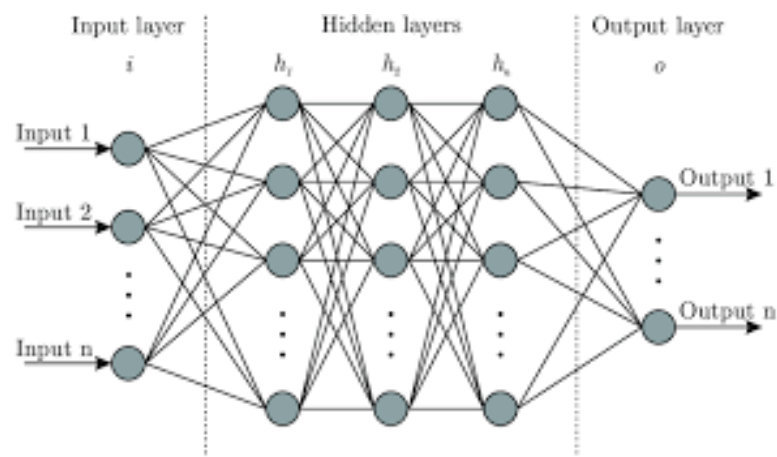
**Table 2.1:** Landscapes and challenges of conventional thermogram-based BC recognition prototypes

[citation]	Methodology	Features	Experiments
Francis <i>et al.</i> (2009) [7]	SVM	It attains a high classification rate regarding accuracy.	This model is not appropriate for large-scale datasets.
Acharya <i>et al.</i> (2012) [6]	SVM	It enhances the performance concerning different metrics like precision, accuracy and sensitivity.	Although it attains 80% accuracy, it has to be improved by extracting superior texture features.
Gomathi <i>et al.</i> (2016) [4]	Logistic regression	It has attained a low false classification ratio.	Conversely, it suffers from an incorrect classification ratio.
Schaefer (2017) [5]	ACO	It provides better detection performance on large-scale datasets.	It suffers from convergence speed and takes more time to get the optimal solutions.
Ekici and Jawzal (2020) [51]	CNN	It improves the early diagnosis rate. It gets a better accuracy rate.	It cannot identify the level of cancer.
Sánchez <i>et al.</i> (2020) [3]	ANN	It offers a better diagnosis rate using low-cost methodologies.	This model does not apply to complex features.
Yadav and Jadhav (2020) [52]	CNN	It gets better classification performance for small-scale datasets.	However, it is not suitable for larger datasets and the use of feature aggregation approaches.
Al <i>et al.</i> (2023) [1]	ELM, MLP	It provides promising detection results with superior accuracy.	The performance is affected due to the quality of protocols utilized for capturing the thermograms.

**Table 2.2:** Structures and encounters of conventional thermogram-based cancer recognition prototypes

Author [citation]	Pre-processing	Segmentation Method	Classification algorithm	Dataset used, Accuracy achieved, other remarks
Acharya et al. (2012) [6]	Pre-processing	-	SVM	Singapore General Hospital, Singapore and 88.1% of accuracy.
Francis et al. (2014) [53]	Image enhancement	K-Means clustering	SVM	90.9% accuracy.
Schaefer (2014) [44]	ACO classification of thermogram	-	ACO	dataset of 146 images (29 cancerous and 117 early detection cases) 79.52% accuracy
Santiago Tello-Mijares et al. (2019) [45]	Image denoising and curvature function	Gradient Vector flow	CNN	DMR-IR database and 97% accuracy achieved
Mohamed Abdel-Nasser, et al. (2019) [42]	Anisotropic diffusion filter	-	CNN	DMR-IR database dynamic images(37 cancerous and 19 healthy)
Yadav and Jadhav (2020) [52]	Contrast enhancement	-	CNN	project database PROENG and 98.5% accuracy
Gomathi et al. (2020) [43]	Adaptive median filter	Multi-Model Subspace Clustering	Logistic regression	DMR-IR database 97.774
Sánchez et al. (2020) [3]	RGB image	Otsu's method	ANN	Database for Mastology Research (DMR) and accuracy attains between 90.17% and 98.33%
Ekici and Jawzal (2020) [51]	Image filtering	(VPP) and (HPP)	CNN	Static dataset, 98.95% accuracy achieved
Al Fayeze (2020) [50]	A homomorphic filtering technique	Binary masking and K-mean clustering	ELM and MLP	DMR-IR database and 99.1% accuracy achieved
Esraa A Mohammad et al. (2022) [46]	Not mentioned	U-net network	Two -class CNN based deep learning	DMR_IR database accuracy = 99.33%, sensitivity = 100% and specificity = 98.67%
Sonar Civilibar [42 b] (2023)	Not mentioned	-	Mask R-CNN	DMR IR dataset, 97.1% testing accuracy

The research has been shown that the proper combination of type of segmentation technique followed by feature extraction and then good classifiers may improve the results. Other than Shannon entropy, there are many alternative definitions of entropy and divergences based on them, like, Renyi entropy, sample entropy, beta-divergence and others. Shannon entropy has been specifically designed for communication problems and works there better. But for machine learning task the alternative definitions of entropy are expected to do better. These new entropy definitions are more robust and have proven to give better results for regression tasks. Similar results are expected for classification task and need be tested. Deep learning approaches are showing extraordinary results for classification problem in other tasks, including medical diagnosis. Till now for breast cancer detection using thermography only CNN has been used and the approach was overfitting there (Soner C. et al. 2023) [42]. So, there is a large scope of applying other DNN approaches, specifically those requiring low resources and are robust. So, we have tried to use simplest structure that needs less computations. Following figure shows basic block diagram of the deep neural network. We stick to the concept that one or more than one hidden layer in neural network called as the Deep Network. We have employed this concept for the implementation of our deep neural network. Following figure shows basic deep neural network structure which consist of input layer, output layer, and number of hidden layers.



**Figure 2.4:** Basic block diagram of Deep Neural Network

## 2.7 Summary

The data set that support the findings of this study are openly available in Visual Lab at [http:// visual. ic. uff. br/ dmi/](http://visual.ic.uff.br/dmi/). Mammogram-based breast cancer detection models have a high false classification ratio and poor accuracy, and as a result, thermography is essential to grow a new automated diagnostic prototype. In spite of the fact that thermography-based breast cancer diagnosis provides greater performance, it is still lacking in accuracy owing to a variety of factors including menstruation, physiological status, and stable temperature. Table 1 contains a number of different breast cancer detection algorithms that are based on thermograms. Both ELM and MLP (Al Fayez et al.2020) [50] provide very accurate and promising detection findings. Nevertheless, the performance is negatively impacted because of the poor standard of the methods that were used to record the thermograms. CNN (Abdel-Nasser et al.2019) [42] both increases the pace at which early diagnoses are made and achieves a higher accuracy rate. Yet, it is unable to determine the stage of the malignancy. ANN (Sánchez-Ruiz et al. 2020) [3] provides a higher percentage of accurate diagnosis with methods that are more cost-effective. On the other hand, this paradigm does not apply to characteristics that are complicated. Logistic regression (Gomathi et al. 2020) [43] has achieved a low false classification ratio thanks to its improved accuracy. On the other hand, it has a problem with the classification ratio being inaccurate. ACO (Schaefer et al. 2014) [44] enables superior detection performance on large-scale datasets. Nevertheless, it has issues with convergence speed and takes a longer amount of time to arrive at ideal answers. The performance of an SVM (Acharya et al. 2012) [6] is improved in terms of a variety of various criteria including precision, accuracy, and sensitivity. Even though it has an accuracy of 80%, it still has to be enhanced by the extraction of more sophisticated texture information. SVM (Francis, *et al.* 2014) [53] is capable of achieving great classification performance in terms of accuracy. Nonetheless, this approach is not appropriate for use with datasets of a large size. When used to small-scale datasets, CNN (Sánchez-Ruiz, D. et al. 2020) [56] achieves superior classification performance. On the other hand, it is not appropriate for use with bigger datasets or methodologies that aggregate features. These issues enable the researchers in presenting a novel CAD structure for the purpose of assisting radiologists and for decreasing the human factor associated via the use of infrared breast thermographic images by applying deep learning methodologies.



## CHAPTER 3

### **BETA ENTROPY FROM BETA DIVERGENCE**

---

In image processing, entropy is a statistical measure used to quantify the amount of information or randomness in an image. It provides insights into the complexity or uncertainty of the pixel intensities within an image. Entropy is computed by considering the probability distribution of pixel intensities. The entropy value is highest when the pixel intensities are uniformly distributed across all possible intensity levels, indicating a high degree of randomness or complexity in the image. Conversely, if the image has a very narrow or concentrated intensity distribution, the entropy value will be low. In image processing, entropy is used for various purposes, including image segmentation, thresholding, and feature extraction. For example, in thresholding, entropy can be utilized to automatically determine an optimal threshold value for separating foreground and background regions in an image. By maximizing the entropy, we can find a threshold that effectively captures the significant information in the image. Overall, entropy serves as a useful measure in image processing to characterize the information content and complexity of an image, enabling various analysis and processing techniques to be applied based on its value.

#### **3.1 Introduction:**

Information theory (Thomas M. Cover and Joy A. Thomas. (n.d.). Book: *Entropy and Information Theory*) [20] and probability theory are two closely related fields of mathematics that deal with the quantification, storage, and communication of data. Probability theory is the study of randomness and uncertainty, while information theory is the study of how to measure and transmit information. The fields of exchange, interpreting, the extraction procedure, and exploitation of information are all topics that are investigated by information theory. In a more general sense, one can consider data to be the solution to the problem of ambiguity. The aforementioned abstract notion was further developed in 1948 by Claude-Shannon in an article titled A Mathematical Theory of Communication. In this framework, information is conceived of as a set of feasible correspondence, and the aim is to transmit such communications over an unstable route, as well as to have the receiver recreate the contents of the message with a low likelihood of error, in regardless of the medium's noise.

Entropy, which can be thought of as a measurement of a system's degree of uncertainty or unpredictability, is one of the most vital ideas in the field of evidence theory. The greater a system's entropy, the greater the degree of uncertainty that exists inside that system. Measurement of the degree of ambiguity associated with a random parameter is another application of entropy's use in probability theory.

There are various points of connection between information theory and probability theory. The idea of entropy, which is central to probability theory, is the inspiration behind, for instance, the Shannon entropy, which is a metric for determining the amount of information contained inside a message. In addition, a significant number of the procedures that are utilized in information theory, such as data compression and error correction, are founded on probabilistic approaches.

Here mentioned how information theory and probability theory are used in **Machine learning**: Information theory is used in machine learning algorithms, such as those used for classification and prediction.

### **3.2 Information Definition – Entropy Definition – Shannon Entropy:**

Information theory deals with extracting information from data or signals. Information is a measure of the amount of uncertainty that is reduced when a new piece of data is received. The more uncertain the data was before it was received, the more information it contains.

#### **3.2.1 Entropy**

Entropy is a measure of statistics contained in data and signal. An organism's entropy can be thought of as an indicator of how disordered or unpredictable it is. The higher a system's entropy value is, the greater the degree to which it is disorganized. Information theory is built upon the disciplines of probability theory and statistics, which serve as its foundation. In these fields, quantifiable information is often referred to in the sense of bits. Units of information that are connected with random variable dispersion are often the focus of information theory. One of the most essential metrics is known as entropy, and it serves as the fundamental constituent of a great deal of other measurement terminology. Entropy is a useful tool for quantifying the amount of evidence contained in a single random parameter. It is used to portray the quality of the input.

The following equation can be used to define the entropy of  $X$  that has a probability mass function of the  $p(y)$ :

$$H(Y) = -\sum p(y) \log_2 p(y). \quad (1)$$

### 3.2.2 Shannon Entropy:

Shannon entropy is a degree of the info content of a communication. It is demarcated as the average volume of info that is contained in each symbol of a message. The formula for Shannon entropy is:

$$H = -\sum p_i \log_2(p_i) \quad (2)$$

where:  $H$  is the Shannon-entropy

$p_i$  is the probability of symbol  $i$  occurring

The choice of logarithmic base determines the unit of measurement for the entropy (e.g. bits, nats or decimals). Using  $\log_2$  gives the entropy in bits, which is most common choice in information theory.

**In Image processing domain**, Entropy is a degree of ambiguity or randomness in a random parameter. Information-entropy or Shannon entropy  $S$ , proposed by Shannon et al. (1949), is defined as:

$$ShEn(M_t^{as}) = -\sum_{w_j}^{M-1} pb(w_j) \log_2(pb(w_j)) \quad (3)$$

Here,  $M = 2^r$  shows the distinct grey level count, the original image is shown by  $M_t^{as}$ , and the probability of occurrence of the  $w_j$  value in the  $M_t^{as}$  image is shown by  $pb(w_j)$  respectively.

The fact that the Shannon entropy has led to a extensive variety of extensions and presentations is one of the best ways to comprehend the importance of this concept.

### 3.2.3 Properties of Shannon Entropy:

1. Information gain: Shannon entropy is used to quantify the information gain or reduction in uncertainty when a new attribute is considered in decision trees or feature selection algorithms. It helps assess the relevance and importance of different features for classification and regression task.

2. **Image Compression:** Shannon entropy provides a theoretical basis for data compression. It indicates the minimum quantity of bits mandatory to encode the outcomes of a random variable. Efficient data compression algorithms aim to exploit the redundancy in the data and reduce the average code length close to the entropy.
3. **Decision Making and prediction:** Shannon entropy is used in decision theory and machine learning to measure the uncertainty or impurity in classification problems. It helps determine the optimal features or attributes to split data, build decision trees, or train models such as random forests and gradient boosting algorithms.
4. **Channel Capacity:** In communication theory, Shannon entropy represents the maximum data transmission rate or channel capacity of a noisy channel. It provides an upper bound on the info rate that can be reliably conveyed through the network.
5. **Invariance:** Shannon entropy is invariant under permutations of the outcomes. That is, the entropy remains the same regardless of the order or labeling of the outcomes. It depends only on the probabilities assigned to the outcomes
6. **Maximum Entropy:** For a discrete  $Y$  with  $n$  conceivable outcomes, the maximum entropy is achieved when all outcomes are alike, i.e.,  $P(Y) = 1/n$  for all  $y$ . In this case,  $H(Y)$  is maximized and equals  $\log_2(n)$ . It represents the highest possible uncertainty or randomness.
7. **Non-negativity:** Shannon entropy is always non-negative. It is zero when the random variable has a single outcome with probability 1, indicating no uncertainty or perfect predictability. Higher entropy values indicate more uncertainty or randomness.

Overall, Shannon entropy plays a fundamental role in quantifying uncertainty, information content, and randomness. Its properties and applications have made it a central concept in information theory and related fields.

### 3.3 KL Divergence from Entropy

The term "Kullback-Leibler divergence" is another name for this notion. It is defined as the subtraction among the actual entropy of a distribution and the value that would be anticipated for the entropy of that distribution if it were subjected to another distribution. This difference may be calculated by subtracting the actual entropy from the value that would be predicted.

Cross entropy is one of the most commonly used loss function in classification.

**KL divergence = Cross Entropy – Entropy**

$$D_{KL}(P||Q) = H(P, Q) - H(P) \quad (4)$$

$$DKL(P||Q) = -\sum P_i \log(Q_i) - [-\sum P_i \log(P_i)] \quad (5)$$

$$DKL(P||Q) = E[\log P - \log Q] \quad (6)$$

$$DKL(P||Q) = \sum_{i=1}^N P_i (\log P_i - \log Q_i) \quad (7)$$

$$DKL(P||Q) = \sum P_i * \log \frac{P_i}{Q_i} \quad (8)$$

The KL divergence cannot take on a negative value and will always be equal to 0 in the event that P and Q are identical to one another. In light of this, we may understand that the KL divergence is a measurement of the amount of information that is lost when Q is represented by P.

The KL divergence distributions P and Q is defined as:

$$DKL(P||Q) = \sum P * \log \frac{P}{Q} \quad (9)$$

Where, P and Q probabilities distributions and summation is taken over all possible outcomes.

In simple terms, KL divergence is a measure of how different two probability distributions are from each other. It is used in machine learning and information theory to compare models and to help select the best model for a given problem. In machine learning, one of the very commonly used techniques is the Maximum Likelihood estimation, which basically tries to learn parameters which maximize the Likelihood or the Conditional Likelihood. KL divergence is a “distance” between probability distributions, and measures the how different two probability distributions are.

If the KL divergence  $D_{KL}(P \parallel Q)$  is small, it indicates that the two distributions  $P$  and  $Q$  are similar, with a small additional information needed to encode outcomes from  $P$  using a code optimized for  $Q$ . Conversely, if the KL divergence is large, it suggests a significant difference between the distributions, implying a larger amount of additional information required.

### 3.3.1 Applications of KL Divergence:

**Feature selection:** KL divergence can be utilized in feature selection algorithms to measure the relevance and redundancy of features. By computing the divergence between the feature distribution and the target variable, one can identify the most informative features for a given task.

**Model selection and comparison:** KL divergence can be used to compare and select between different statistical models or hypothesis. It can measure the discrepancy between the observed data and the predicted data from different models, allowing for model selection based on goodness-of-fit.

**Optimization:** KL deviation can be used as a loss function or a regularization term in optimization problems. For instance, in variational inference, it is minimized to find the closest approximation to the true posterior distribution. It is also used in maximum likelihood estimation and maximum a posteriori estimation.

Here is an illustration of how KL deviation and entropy can be utilized. Suppose we have  $P$  and  $Q$ , that represent the probability of rain tomorrow.  $P$  is the actual distribution, and  $Q$  is a prediction of the distribution. The KL divergence between  $P$  and  $Q$  can be used to measure the accuracy of the prediction. If the KL divergence is low, then the prediction is accurate. If the KL divergence is high, then the prediction is inaccurate.

## 3.4 Alternative Entropy Definitions

There are Alternative entropy definitions in the literature [20]:

### 3.4.1 Minimum entropy, Maximum entropy:

The quantity of entropy that is considered to be "minimum" for a particular probability distribution is the value that is the lowest feasible for that distribution. When all of the probability are exactly the same, it has been accomplished. The value of

entropy that is considered to be at its maximum for a particular probability distribution is referred to as maximum entropy. The minimum entropy and maximum entropy are concepts related to the info content or ambiguity of a chance distribution.

The minimum entropy refers to the lowest possible level of uncertainty or information content in a probability distribution. It occurs when all events in the distribution have equal probabilities. In other words, the distribution is maximally uniform. The minimum entropy is achieved when the distribution has maximum entropy but with a uniform or equally likely assignment of probabilities to all events. The minimum entropy is achieved when all events in a discrete probability distribution have equal probabilities, resulting in a maximally uniform distribution. For a discrete probability distribution with  $N$  possible events, the minimum entropy  $H_{min}$  is given by:

$$H_{min} = -\log_2 \left( \frac{1}{N} \right) = \log_2(N) \quad (10)$$

where:

$H_{min}$  is the minimum entropy.

$N$ : possible trials in the distribution.

Mathematically, if a discrete probability distribution has  $N$  possible events, the minimum entropy occurs when each event has a probability of  $\frac{1}{N}$ . In this case, the entropy value is at its minimum, indicating the least amount of uncertainty.

### **Maximum Entropy:**

The maximum entropy, on the other hand, refers to the highest possible level of uncertainty or information content in a probability distribution. It occurs when all events in the distribution are equally unpredictable or have the maximum possible entropy. In this case, the distribution represents the most diverse or least biased set of events.

For a discrete probability distribution with  $N$  possible events, the maximum entropy is achieved when all events are equally unpredictable, meaning the probabilities are distributed uniformly. The maximum entropy distribution is often considered as the least informative or most uncertain distribution, as it provides no additional knowledge about the underlying events.

The idea of maximum entropy is commonly used in evidence theory and statistical modeling, where it serves as a benchmark or reference for deriving probability distributions based on limited information or constraints.

The maximum entropy occurs when all events in a discrete probability distribution are equally unpredictable or have the maximum possible entropy. For a discrete probability distribution with  $N$  possible events, the max-entropy  $H_{max}$  is given by:

$$H_{max} = -\sum_x \log_2 p(x) \quad (11)$$

where:

$p(x)$  is the likelihood assigned to event  $x$  in the distribution.

The max-entropy distribution is considered by having the highest entropy value among all possible distributions with the same number of events. It represents the least biased or most diverse set of events, providing the least amount of information or certainty about the underlying events. It's important to note that the logarithm used in the formulas can be of any base, such as 2 for binary entropy or natural logarithm ( $\ln$ ) for natural entropy, as long as the same base is used consistently throughout the calculations.

### 3.4.2 Sample entropy

Sample entropy is a measure used to quantify the amount of symmetry or unpredictability in a time series data. It is often used as a feature in various applications, including signal processing, time series analysis, and biomedical signal analysis. Unlike Shannon entropy, which processes the average amount of uncertainty in a probability distribution, sample entropy focuses on the patterns within a data sequence. The steps to compute sample entropy are as follows:

Sample entropy is calculated based on the concept of approximate entropy (ApEn). Given a time series data sequence of distance  $N$ , symbolised as  $y(1), y(2), \dots, y(N)$ , the sample entropy  $SEn(m, r)$  is well-defined as:

$$SEn(m, r) = -\log(Cm(r + 1) / Cm(r)) \quad (12)$$



Here,  $m$  is the distance of the template pattern,  $r$  is the maximum lenience or radius of the similarity criterion,  $C_m(r)$  represents the count of  $m$ -distance forms that are alike (within the tolerance  $r$ ), and  $C_m(r + 1)$  represents the total of  $(m + 1)$  distance outlines that are analogous.

**Similarity criterion:** The similarity standard is determined by the tolerance or radius parameter,  $r$ . It specifies the maximum acceptable difference between patterns to be considered similar. If the difference between two patterns is greater than  $r$ , they are considered dissimilar.

**Template pattern:** The template pattern, of length  $m$ , serves as a reference for comparison. It is a subsequence of the original data sequence, and all subsequent patterns of the same length are compared to this template pattern.

**Counting similar patterns:** Sample entropy counts the quantity of alike patterns of length  $m$  and  $(m + 1)$  in the data sequence. The similarity is determined by the tolerance parameter,  $r$ . Similar patterns have differences (Euclidean distance, for example) within the tolerance range.

**Calculating the sample entropy:** is calculated as the logarithm of the ratio of the counts of  $(m + 1)$ -length patterns and  $m$ -length patterns. The negative sign is used to make the entropy positive.

**Interpretation:** A lower sample entropy value indicates more regularity or predictability in the data sequence, suggesting that similar patterns occur more frequently. Conversely, a higher sample entropy value indicates more irregularity or randomness, indicating that similar patterns are less likely to occur.

### **3.4.3 Approximate entropy:**

It (ApEn) is an amount used to quantify the intricacy or abnormality of a time series data sequence. It is particularly useful in analysing biomedical signals, such as electroencephalograms (EEG) and heart rate unpredictability (HRV), to assess the regularity or predictability of physiological processes. Here's how approximate entropy works:

**Definition:** Given a time series data sequence of distance  $N$ , as  $p(1), p(2), \dots, p(N)$ , the approximate entropy  $ApEn(m, r)$  is calculated as follows:

$$ApEn(m, r) = -\log(Am(r + 1) / Am(r)) \quad (13)$$

In this case,  $m$  refers to the total length of the template structure,  $r$  is the greatest possible tolerance or radius of the similarity criteria,  $Am(r)$  is the average number of patterns of length  $m$  that are similar (within the tolerance  $r$ ), and  $Am(r + 1)$  is the average number of patterns of length  $m + 1$  that are similar.

**Calculating the approximate entropy:** ApEn is then computed as the logarithm of the ratio of the average counts of  $(m + 1)$  length patterns and  $m$ -length patterns. The negative sign is used to make the entropy positive. Interpretation: A lower approximate entropy value indicates more regularity or predictability in the data sequence, suggesting that similar patterns occur more frequently. Conversely, a higher approximate entropy value indicates more irregularity or complexity, indicating that similar patterns are less likely to occur. Approximate entropy provides a amount of the complication or indiscretion of a time series, capturing the degree of predictability or randomness in the data. It used in the examination of bodily signals, as well as in other domains where assessing the complexity or regularity of sequential data is of interest.

Approximate entropy (ApEn) can be utilized in machine learning in various ways to analyse and process time series data. Here are a few applications of approximate entropy in machine learning:

1. **Classification and prediction:** ApEn can be used as a feature or similarity measure in classification and prediction tasks involving time series data. For instance, in activity recognition from sensor data or gesture recognition, ApEn can capture the complexity or regularity of different activities or gestures, allowing for accurate classification. In addition, ApEn can be employed as a similarity metric for time series comparison, enabling tasks such as time series clustering or nearest neighbor classification.
2. **Assessing model performance:** ApEn can serve as an evaluation metric for machine learning models that deal with time series data.

3. Hyperparameter tuning: ApEn can be utilized as a criterion for hyperparameter tuning in machine learning algorithms. By tuning these hyperparameters based on the ApEn measure, it is possible to improve the model's performance.

These are some of the ways in which approximate entropy can be applied in machine learning for time series data analysis. By capturing complexity, irregularity, and temporal patterns, ApEn can enhance the understanding and modeling of time-dependent processes, contributing to improved performance in various tasks.

#### 3.4.4 Alpha entropy:

A time series' randomness can be measured using a concept called alpha entropy. The Shannon entropy is a well-known measure of randomness; the  $\alpha$ -entropy of a multivariate distribution is a generalization of this concept that assesses the randomness of a set of variables. Alfred Rényi introduced the concept of  $\alpha$ -entropy for the very first time in a study that was published in 1961. Since then, numerous fundamental aspects of the  $\alpha$ -entropy have been isolated and characterized (Heidari, A. et al. 2019) [26] (Venkata Rao, R. et al. 2016) [27]. It is based on the probability that a given symbol will be followed by another symbol.

The concept of Alpha -entropy in the context of a multivariate distribution serves as an extension to the more commonly recognized Shannon entropy. The alpha-entropy can yield another valuable measure known as a Jensen difference. This difference is an extension of the traditional Jensen difference and will play an important part in our efforts to expand entropic pattern matching approaches to accommodate high feature dimensions. This difference is an extension of the conventional Jensen difference.

Alpha or Renyi entropy is a concept in information theory and statistical physics that generalizes Shannon entropy. It is named after Alfréd Rényi, a Hungarian mathematician. Renyi entropy provides a family of entropy measures that encompass Shannon entropy as a special case:

$$R_{en} = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^n p(x_i)^\alpha \quad (14)$$

where  $\alpha$  is a parameter that determines the order of the Renyi entropy.

Some key properties and characteristics of Renyi entropy are:

1. **Parameter  $\alpha$ :** The value of  $\alpha$  determines the properties of the Renyi entropy. When  $\alpha$  approaches 1. As  $\alpha$  deviates from 1, Renyi entropy captures different aspects of the distribution. For example, when  $\alpha$  is larger than 1, Renyi entropy gives more weight to the most probable events, while smaller values of  $\alpha$  emphasize the contribution of less probable events.
2. **Generalization of Shannon:** Shannon is a different phase of Renyi when  $\alpha$  is equal to 1. It provides a degree of the average ambiguity or info content of a random parameter. Renyi entropy generalizes this notion by considering a broader range of orders.
3. **Order dependence:** Renyi entropy exhibits order dependence, meaning that different orders of Renyi entropy can lead to different interpretations and characterizations of the distribution. Higher orders emphasize the importance of high-probability events, while lower orders give more weight to the tails of the distribution.
4. **Additivity:** Unlike Shannon entropy, Renyi entropy is not necessarily additive for independent random variables. The additivity property holds for certain specific values of  $\alpha$ , such as  $\alpha = 0$  and  $\alpha = 2$ .
5. **Applications:** Renyi entropy has found applications in various fields, including information theory, statistical physics, signal processing, machine learning, and complex systems examination. It is used to measure diversity, distinguishability, and information content in probability distributions. It has also been applied in clustering, feature selection, and anomaly detection tasks.

Renyi entropy provides a flexible framework for characterizing the uncertainty and information content of random variables. By varying the parameter  $\alpha$ , it captures different aspects of the distribution and offers a range of entropy measures beyond Shannon entropy.

When  $\alpha$  is set to 1, Renyi entropy touches to Shannon's. However, for other values of  $\alpha$ , Renyi entropy captures additional information about the distribution. Alpha entropy, also known as Renyi entropy, captures different aspects of the probability distribution by varying the parameter  $\alpha$ . The value of  $\alpha$  determines the order of the

Renyi entropy and influences how information and uncertainty are measured. Here's how different values of  $\alpha$  capture distinct aspects of the probability distribution:

1.  $\alpha = 0$ : When  $\alpha$  is set to 0, Renyi entropy represents the min-entropy. It measures the uncertainty associated with the least probable event in the distribution.
2.  $\alpha = 1$ : At  $\alpha = 1$ , Renyi entropy meets to Shannon, which measures the average uncertainty or information content of the distribution. Shannon is widely used in information theory and provides a measure of the overall randomness or unpredictability of the distribution.
3.  $0 < \alpha < 1$ : For values of  $\alpha$  between 0 and 1, Renyi entropy emphasizes the contribution of less probable events in the distribution. It gives more weight to the tail of the distribution, capturing the information in the rare events. Lower values of  $\alpha$  focus more on the tails, providing insights into the extreme events or outliers.
4.  $\alpha = 2$ : At  $\alpha = 2$ , Renyi entropy is known as collision entropy. It is related to collision probabilities and collision theory. Collision entropy quantifies the degree of overlap or collision between elements of the distribution, measuring the redundancy or clustering of events.
5.  $\alpha > 1$ : For values of  $\alpha$  greater than 1, Renyi entropy gives more weight to the most probable events in the distribution. It emphasizes the concentration of probabilities and captures the information contained in the highly probable events. Larger values of  $\alpha$  concentrate the information in the most likely outcomes.

By varying the parameter  $\alpha$ , alpha entropy offers a range of entropy measures that highlight different characteristics of the probability distribution. It allows for a nuanced analysis of uncertainty, information content, and the concentration of probabilities, providing insights into different aspects of the distribution's behavior.

### **3.5 Density Power divergence or Beta divergence**





The beta divergence is a metric that may be used to determine the distance that exists between two probability distributions. It is an extension of the Kullback-Leibler divergence, and it is commonly used in the process of statistical inference that is

resilient. Due to the fact that it contains a variety of characteristics, the beta divergence is an extremely useful tool for estimating the relative distance that exists between two probability distributions. First, it is impossible for it to have a value that is less than zero, and it will always have a value of zero if and only if the equation  $p = q$  is correct. Second, it is symmetric, so

$$D\beta (p \parallel q) = D\beta (q \parallel p). \quad (15)$$

Third, it is parameterized by  $\beta$ , which allows us to control the amount of robustness to outliers.

Beta divergence applications:

-  Robust statistical inference
-  Non-negative matrix factorization
-  Image processing
-  Natural language processing

Here are some of the advantages of using beta divergence:

- a) It is a reliable method for determining the gap in probability distributions between two sets.
- b) It can be utilized in the process of determining the distance that exists between two probability distributions that have distinct supports.
- c) It is a parameterized measure of distance, which allows us to control the amount of robustness to outliers.

It has also been used for Nonnegative Matrix Factorization [80] to achieve robustness. The generalized alpha-Beta divergence has been recently used for dimensionality reduction in and to derive a robust Variational Auto-Encoder. Also, the free parameter significantly rises the liveness to handle diverse types and amount of clutter in data. But according to the author's knowledge the entropy definition induced from Beta divergence has not been used in classification task. To exploit the robustness against outliers' characteristics of Beta divergence this article uses them as features and expects them to be able to achieve class separation among the thermal variations due to cancer class and those due to noise in either a normal class or a benign tumor class thermographic image of breast.

The beta divergence is defined as follows:

$$D_{\beta}(P||Q) = \frac{1}{\beta} \int \left( P^{\beta}(x) - Q^{\beta}(x) \right) P(x) dx, \quad (16)$$

The density power divergence can be calculated between two probability distributions by comparing their power-transformed PDFs and weighting the divergence by the original PDF of the reference distribution. The parameter  $\beta$  allows you to control the emphasis on different aspects of the divergence and the overall magnitude of the divergence measure.

Where,  $P(x)$  and  $Q(x)$  These are the probability density functions (PDFs) of the two distributions, The PDF represents the probability of a random variable taking a specific value.

$P^{\beta}(x)$  and  $Q^{\beta}(x)$  These terms raise the PDFs  $P(x)$  and  $Q(x)$  to the power of  $\beta$ . This is known as the power transformation. The power transformation is applied to emphasize or de-emphasize certain aspects of the distributions, depending on the value of  $\beta$ . By adjusting  $\beta$ , you can control the sensitivity to differences between the distributions.

$1/\beta$ , this term scales the integral by dividing the entire expression by  $\beta$ . This scaling factor adjusts the magnitude of the divergence measure. Larger values of  $\beta$  tend to increase the sensitivity to differences between the distributions.

### 3.6 Deriving Beta entropy from Beta divergence

To begin deriving the entropy of the Beta distribution from the Beta divergence, we can begin with the meaning of the Beta divergence. This will allow us to deduce the entropy of the Beta distribution. The Beta divergence is an indicator of the dissimilarity between two probability density functions, whereas the entropy of a Beta is a metric of the ambiguity or unpredictability linked with the distribution. To derive beta entropy from beta divergence, we can start with the formula for beta divergence:

$$D_{\beta}(P||Q) = \frac{1}{1-\beta} [\sum x(Pi(x)^{\beta} * Qi(x)^{1-\beta}) - 1] \quad (17)$$

where:

$D_{\beta}(P||Q)$  is the beta divergence between the probability distributions  $P$  and  $Q$ .

$\beta$  is the parameter of the beta divergence.

To convert beta divergence into entropy:

$$H_{\beta}(P) = \frac{1}{\beta} \int (P^{\beta}(x) - 1)P(x)dx. \quad (18)$$

where:

$H_{\beta}(P)$  is the beta entropy of the probability distribution  $P$ .

$P(x)$ : probability density function (PDF) distribution  $P(x)$ .

$P^{\beta}(x)$ : This term raises the PDF  $P(x)$  to the power of  $\beta$ . This is known as the power transformation. The power transformation is applied to emphasize or de-emphasize certain aspects of the distribution, depending on the value of  $\beta$ .

### Advantages of Beta Divergence:

#### 1. Parameterization and Flexibility

- Adjustable Sensitivity ( $\beta$ ): Beta divergence includes a parameter  $\beta$  that can be adjusted to emphasize different aspects of the data distribution. This flexibility allows it to generalize various divergence measures:
  - $\beta = 1$  yields KL divergence, which measures the relative entropy between two distributions.
  - $\beta = 2$  corresponds to the squared Euclidean distance, focusing on the absolute differences.
  - $\beta = 0$  results in Itakura-Saito divergence, which is more sensitive to small values in the distribution.
- Tailored Feature Sensitivity: By tuning  $\beta$ , beta divergence can be made more sensitive to certain features of the data, such as tails or central tendencies, depending on the classification task.

#### 2. Adaptability to Data Characteristics

- Handling Different Distributions: Different datasets might have varying distributions, such as heavy-tailed, skewed, or multimodal distributions. Beta divergence can adapt to these characteristics by selecting an appropriate  $\beta$ , which is not possible with fixed measures like KL divergence.
- Robustness to Outliers: Higher values of  $\beta$  make beta divergence less sensitive to outliers, improving robustness in noisy datasets. This



adaptability can result in better performance in real-world scenarios where data often contains noise.

### 3. Enhanced Discrimination Power

- **Capturing Complex Structures:** Beta divergence can capture complex structures in the data more effectively. For example, in image classification, textures and patterns can be better distinguished by tuning  $\beta$  to focus on relevant features.
- **Improved Feature Extraction:** By providing a more nuanced measure of divergence, beta divergence can enhance the feature extraction process, leading to more discriminative features for classification models.

By converting beta divergence into beta entropy, can be use to quantify the uncertainty or information content in a more interpretable manner.

Beta entropy can offer greater robustness compared to Shannon entropy and other entropies for classification tasks due to several factors related to its formulation and flexibility.

### 3.7 Summary

Various entropy measures discussed in this chapter provide valuable tools for quantifying uncertainty, similarity, dissimilarity, and information content within probability distributions and data. Entropy measures play a crucial role in statistical inference and model selection. They provide a way to assess the goodness-of-fit between observed data and a model, allowing researchers to compare and choose the most appropriate distribution or model. Divergence metrics, such as Kullback-Leibler divergence, quantify the dissimilarity between probability distributions, enabling tasks like density estimation, model fitting, and hypothesis testing. Entropy and divergence metrics are extensively used in machine learning algorithms. They serve as objective functions or regularization terms in training models, aiding in tasks such as classification, clustering, and feature selection. These measures provide a basis for quantifying uncertainty, model complexity, and similarity between data samples. They offer quantitative insights into the nature of data, facilitate model development and evaluation, and enable the extraction of meaningful information from complex systems. Beta divergence's key advantage lies in its parameter  $\beta$ , which allows it to adapt to different data distributions and focus on various aspects of the data, such as tails or

central values. This adaptability makes it more robust to noise and outliers, enhances feature extraction, and improves the discrimination power of the features used for classification. By carefully tuning  $\beta$  and integrating beta divergence into the feature extraction process, you can achieve more accurate and robust classification outcomes compared to using fixed divergence measures. Similarly, Beta entropy's robustness stems from its flexibility and adaptability. The parameter  $\beta$  allows for fine-tuning the entropy measure to better capture the characteristics of the data, making it a powerful tool for improving classification accuracy, especially in tasks involving complex or noisy data. By effectively leveraging beta entropy, the discriminative power of features can be enhanced, leading to more accurate and robust classification outcomes.

## CHAPTER 4

### HYBRID HARRIS-HAWK JAYA OPTIMIZATION ALGORITHM

---

It is known that the performance of a Deep Learning algorithm highly depends upon the hyperparameter tuning. Metaheuristic algorithms are increasingly used for parameter tuning in Deep Neural Networks (DNNs) due to their ability to efficiently search large, complex, and non-convex hyperparameter spaces. DNNs have many hyperparameters, including learning rates, batch sizes, number of layers, number of neurons per layer, activation functions, and regularization parameters. The hyperparameter space is high-dimensional and difficult to explore exhaustively. These algorithms are designed to handle high-dimensional spaces efficiently, offering a practical way to find good solutions without exhaustive search. The performance landscape of DNNs is often non-convex with multiple local minima and saddle points. Traditional optimization methods may struggle to navigate this complex landscape effectively. Exhaustive search methods like grid search become impractical for large hyperparameter spaces due to their high computational cost. These algorithms use intelligent sampling and search strategies to explore the hyperparameter space more efficiently, often requiring fewer evaluations to find good solutions compared to exhaustive methods. As DNNs and datasets grow in size, the hyperparameter tuning process becomes more challenging and computationally expensive. They scale better with problem size and can be parallelized to further enhance efficiency, making them suitable for large-scale hyperparameter tuning. Metaheuristic algorithms can adapt their search strategies based on the progress of the search, dynamically adjusting to the characteristics of the problem.

We have incorporated newly developed metaheuristic algorithm Hybrid Harris Hawk JAYA optimization (HHHJO) to fine tune these DL algorithm parameters including the learning rate, number of epochs and hidden neuron count, as well, the parameters in FCM including epsilon, maximum iteration count in fuzzy and fuzziness parameter. HHHJO is a combination of two optimization algorithms HHO (Harris Hawk Optimization) and JA (Jaya Algorithm). HHO [84] holds repeated mechanisms of exploration and exploitation to enhance the system performance. It requires good

number of algorithmic parameters to be decided for better performance. Too many algorithms specific parameters not only increase the computational effort. But it may suffer in population diversity and local optima space. Compare to it, JA (Venkata Rao, R. et al. 2016) [27] is simple and do not require any additional parameters for the initialization. This motivates to combine both HHO and JA to overcome these issues, and a novel hybrid HHHJO algorithm is developed.

#### 4.1 Introduction

Hyperparameter modification strongly impacts Deep Learning algorithm performance. Trial-and-error hyperparameter tweaking requires skill and often fails. We constructed a new heuristic optimization algorithm HHHJO to fine tune DL algorithm parameters learning rate, number of epochs and hidden neuron count, as well, three FCM parameters epsilon, maximum iteration count in fuzzy, and fuzziness.

In the proposed HHHJO, the final update is attained by the below condition by using the condition if  $\text{num} \geq 0.5$  and then updating the position using JA; otherwise, updating the position using HHO.

#### 4.2 HHO Algorithm:

HHO stands for Harris Hawks Optimization. It is a metaheuristic algorithm that was developed by Heidari in 2019. HHO is motivated by the chasing conduct of Harris hawks. HHO works by iteratively probing for the best resolution to a problem. In each repetition, HHO updates the positions of the hawks in the search space. The hawks are attracted to the finest answers that have been establish so far, and they are also repelled by the worst solutions. HHO has been exposed to be active in cracking a variety of optimization complications:

**Stage-1 Exploration (Tracing):** Typically, Harris Hawks make their homes on telegraph poles or enormous trees. This allows them to maintain the ideal position to hunt their prey by looking for it, waiting for it, and locating it with their keen eyesight. The search behavior of Harris Hawks is referred to as global exploration, and it is equated in Eq. (1)) with two different strategies: perch when the other members of the family are close by, which is indicated by  $y \geq 0.5$ , and rabbit when they are far away, which is denoted by  $y < 0.5$  as in the equation.

$$A_j^{v+1} = \begin{cases} A_{rnd}^v - s_1 |A_{rnd}^v - 2 * s_2 * A_j^v|, & y \geq 0.5 \\ (A_{rbt} - A_{mn}^v) - s_3 * (mv - s_4 * (nv - mv)), & y < 0.5 \end{cases} \quad (19)$$

Where a random Harris hawk presented in  $v^{th}$  iteration of the population is indicated as  $A_{rnd}^v$ , the term  $A_j^v$  denotes the population of  $j^{th}$  individual in  $v^{th}$  iteration, the individual attained in current optimal position is termed as  $Arbt$ , the factor utilized for conversion is given as  $y$  and its random numbers are equally distributed as  $s_1, s_2, s_3$  and  $s_4$  and  $F$  in the range of  $(0, 1)$ , the upper limit is denoted as  $nv$  and the lower limit is given as  $mv$  and the position mean value in current population is given as  $A_{mn}^v$ . The expected value of the present position is formulated in Eq. (2) as under

$$A_{mn}^v = \frac{1}{p \sum_{j=1}^p A_j^v} \quad (20)$$

The term  $p$  indicates the size of the population

**Stage 2-Transition from exploration to exploitation:** The transition process achieved from global exploration to local exploration in HHO is highly organized by escaping energy factor and is showcased in Eq. (3) as under:

$$F = 2 * F_0 * (1 - v/v_{mx}) \quad (21)$$

Where,

The fleeing energy is represented by the value  $F_0$ , the simple random sample provided in the limit  $(-1, 1)$  is represented by the value, the max number of repetitions is shown by the value  $v_{mx}$ , and the current repetition is shown by the value  $v$ , respectively. When the escaping energy is less than one, the hawks search various regions to discover where the rabbits are hiding, which is recognized as the investigation phase. On the other hand, when the escaping energy is equal to one, the hawks search only within the immediate region, which is known as the exploitation stage.

### Stage 3-Exploitation (Attack):

After the Harris hawks have determined that the intended prey is nearby, they will begin to circle the prey and patiently wait for an opportunity to attack. Nevertheless, the real process of prediction is quite complicated, and as a result, the hawks need to make certain modifications in accordance with the actions of their prey.

The criteria  $s \in (0,1)$  is used to determine whether or not the prey is successful in breaking out of the encirclement. The phase of attacking, also known as exploitation, may be approached in four distinct ways, each of which is explained in more detail below.

**Soft besiege:** In this procedure when  $s \geq 0.5$  and  $|F| \geq 0.5$ , the victim is unable to run away, despite the fact that it has enough energy to get away from the ring. As a result, while pursuing their prey, Harris Hawks used a tactic known as "soft beige," which is represented by the existing component in Equation (22), which is expressed in the following format:

$$A_j^{v+1} = \Delta A_j^v - F * |K * A_{rbt} - A_j^v| \quad (22)$$

$$\Delta A_j^v = x_{rbt} - A_j^v. \quad (23)$$

The vector distance away from the prey is provided by  $A_j^v$ , and the equation for determining the prey's jumping length, denoted by the variable  $K$ , while it is trying to escape is written as follows in Eq. (24):

$$K = 2 * (1 - s_5) \quad (24)$$

The  $s_5$  is spread in the range of  $(0,1)$ , respectively.  $|F| < 0.5$

**Hard besiege:** At the time when  $|F| < 0.5$  and  $s \geq 0.5$ , i.e., The hunting strategy used by Harris hawks is known as "hard besieges," and it consists of surrounding their victim with a great number of obstacles in reducing the prey's ability to flee and the number of opportunities it has to do so. Harris hawks typically hunt in groups of two to four individuals. The mathematical representation of this is as follows:

$$A_j^v = x_{rbt} - F * |\Delta A_j^v| \quad (25)$$

In the condition of mild besiege, the energy of the prey has not been depleted to the point where it is unable to move or escape, in contrast to the state of hard besiege, in which the energy of the prey has been depleted to the point where it is unable to move or break free.

**Soft surround with advanced rapid dives:** When  $|F| \geq 0.5$  and  $s < 0.5$ , the target is able to break out of the encirclement circle, and it has the strength to do so. The Harris hawks have two different implementation techniques for their intelligent assault, which may be derived from the following equations (Eq. 26) and (Eq.27)

$$B = Arbt - F * Arbt - A_j^v \quad (26)$$

$$C = B + t * lvy(dm) \quad (27)$$

The optimization dimension is given as  $dm$  and the random vector is  $t$  and  $lvy$  is represented in Eq. (28) as

$$lvy(A) = 0.01 \times \frac{b \times \sigma}{|\mu|^{\frac{1}{\beta}}} \quad (28)$$

$$\text{Where, } \sigma = \left( \frac{\tau(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\tau\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}} \right)^{\frac{1}{\beta}} \quad (29)$$

The term  $b$  and  $\mu$  be the random variable in the range (0,1) and  $\beta$  be the default constant and it is fixed as 1.5. The depreciation of the optimization problem is given in Eq. (30) as

$$A_j^{v+1} = \begin{cases} B, & \text{if } g(B) < g(A_j^v), \\ C, & \text{if } g(C) < g(A_j^v) \end{cases} \quad (30)$$

in which case the word in question is the fitness function. As a result, the hawks evaluate the results of the prospective movement  $B$  in evaluation to the consequences of past dive, and if the results are not reasonable, the hawks engage in dives that are erratic, abrupt, and quick.

**Hard surround with advanced rapid dives:** When  $|F| < 0.5$  and  $s < 0.5$ , the prey can escape but its energy level is low. So, the Harris hawks form a hard sphere formerly performing the attack. The strategy is expressed again as Eq. (12) but now  $A_j^v$  in Eq. (25) is replaced by  $A_{mn}^v$  in Eq. (20) to obtain  $B$ .

### 4.3 JAYA Algorithm:

The JAYA (Venkata Rao, R. et al. 2016) [27] algorithm is a metaheuristic algorithm that is inspired by the behaviour of ants. It is a population-based algorithm that works by iteratively searching for the best solution to a problem. The JAYA algorithm has been used in a variety of applications, including breast cancer diagnosis. In this application, the JAYA algorithm is used to optimize the parameters of a support vector machine (SVM) classifier. The SVM classifier is a machine learning algorithm that is used to classify data into two or more categories. The JAYA algorithm has been shown to be effective in improving the accuracy of SVM classifiers for breast cancer diagnosis. In one study, the JAYA algorithm was used to optimize the parameters of an SVM classifier on the Breast Cancer Wisconsin (Diagnostic) dataset. The results showed that the JAYA optimized SVM classifier achieved an accuracy of 97.18%, which was significantly higher than the accuracy of the SVM classifier without the JAYA algorithm (96.49%).

Here are some of the advantages of using the JAYA algorithm in breast cancer diagnosis:

- The JAYA system is a relatively simple system to implement.
- The JAYA system is a robust system that is not easily trapped in local minima.
- The JAYA algorithm has been shown to be active in refining the accuracy of SVM classifiers for cancer diagnosis.

However, there are also some disadvantages to using the JAYA algorithm in breast cancer diagnosis:

- The JAYA algorithm can be computationally expensive.
- The JAYA algorithm may not be suitable for all breast cancer datasets.

The JA (Venkata Rao, R. et al. 2016) [27] optimization technique is deceptively simple despite its considerable capability. It is a population-oriented iterative method that is based on a basic concept: the idea that the best answer might be reached by advancing near the present finest resolution and fleeing gone from the poorest option. Let the term  $h(e)$  be the objective function that should be minimized or maximized, the generation be indicated as  $k$ , the design variables be given as  $f$  and the size of a population be presented as  $m$ . The finest worth attained for the function in the



$k^{th}$  iteration is given as  $h(e)_{bst,k}$  and the worst value for the same is presented as  $h(e)_{wst,k}$ . The symbol  $E_{f,c,k}$  denote the value in  $f^{th}$  design parameter value in  $c^{th}$  population and  $k^{th}$  generation are getting updated in Eq. (31) as

$$E'_{f,c,k} = E_{f,c,k} + s_{1f,c}(E_{f,bst,k} - |E_{f,c,k}|) - s_{2f,c}(E_{f,wst,k} - |E_{f,c,k}|) \quad (31).$$

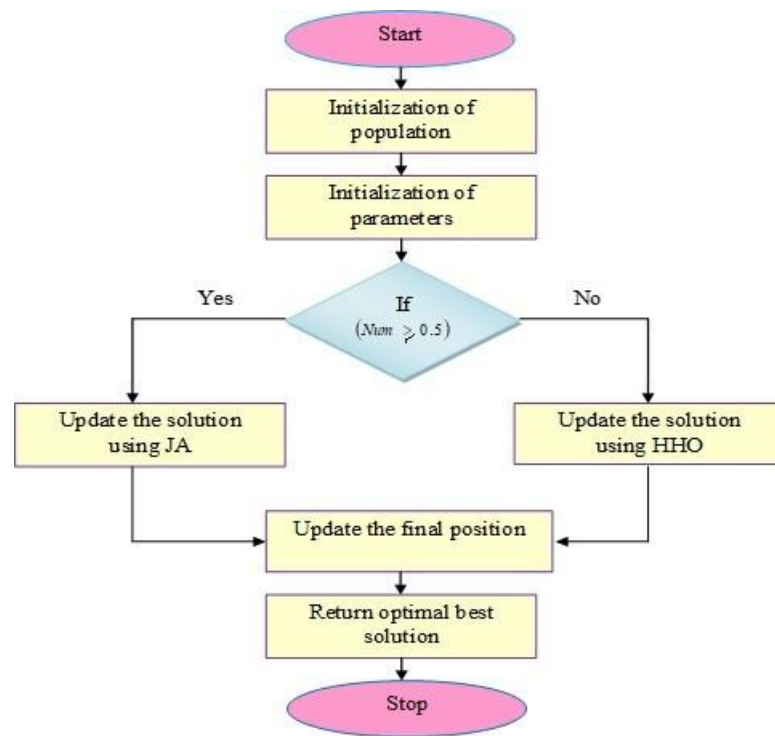
Here, the term  $E_{f,bst,k}$  indicates the best population and  $E_{f,wst,k}$  denote the worst population in  $f^{th}$  the region. The term  $E'_{f,c,k}$  is the improved value of  $E_{f,c,k}$  and the random variables are indicated as  $s_{1f,c}$  and  $s_{2f,c}$  in the range (0,1). The term  $s_{1f,c}(E_{f,bst,k} - |E_{f,c,k}|)$  represents the movement towards the best solution and  $s_{2f,c}(E_{f,wst,k} - |E_{f,c,k}|)$  indicates the movement away from the worst solution. The term  $E_{f,c,k}$  gets accepted if it offers superior functional value. In final generation, all the accepted functional values are considered as the input for upcoming generation.

#### 4.4 Novel HHHJO Algorithm Proposal

It is common knowledge that the hyperparameter tweaking of a Deep Learning algorithm has an important effect on the effectiveness of the procedure. The traditional method for tuning hyperparameters is called trial and error, which needs a significant amount of knowledge and competence, and even then, the level of performance that can be reached is not always optimal. Therefore, in order to fine tune these deep learning algorithm parameters, which include the learning rate, number of epochs, and hidden neuron count, and the additional parameters in FCM, which include epsilon, maximum number of iterations number in fuzzy, and fuzziness parameter, we have incorporated the recently designed HHHJO. This has allowed us to fine tune these Deep Learning optimization algorithms.

HHO [84] has many processes of exploration and exploitation, with the end goal of improving the system's overall performance. In order to get higher effectiveness, it is required to decide on a significant quantity of algorithmic factors. Having an excessive number of algorithm-specific parameters not only makes the calculation more difficult. Nonetheless, it may have negative effects on population diversity and the availability of local optimal space. In contrast to it, JA (Venkata Rao, R. et al. 2016) [27] is uncomplicated and does not call for the inclusion of any extra parameters during the start-up process. This provides the impetus for combining HHO and JA in order to

solve these challenges, and as a result, a novel hybrid algorithm known as HHHJO is devised.



**Fig. 4.1:** Flow diagram of developed HHHJO

**Algorithm 1:** Developed HHHJO

```

Initialize the population of HHO and JA
Parameters determination
For all solution
  validate fitness for several solutions
  If (num ≥ 0.5)
    Upgrade the position using JA
  Else
    Upgrade the position using HHO
  End if
End for
Find the optimal best solution
End
  
```

The flow chart for the developed HHHJO is given in Fig.4.1

The flow chart shown in figure has the steps population initialization and parameters initialization and it is crucial step in evolutionary and optimization algorithms. This process involves creating an initial set of potential solutions represented as individuals in populations. Here we incorporate two optimization algorithms i.e Jaya optimization and Harris Hawk optimization. The Jaya optimization is a simple optimization technique that aims to improve the objective function iteratively. Initialization of population in Jaya algorithm involves setting up the initial solution for optimization process. To initialize the population in the Jaya algorithm we first start by defining the search space, which includes the lower and upper bounds for each parameter of the problem. Then, create a population of potential solutions randomly within this search space. The size of the population is typically specified in advance and can vary depending on the problem's complexity. Each individual in the population represents a possible solution to the problem, with its set of parameters values.

1. Initialize the Population: Create an initial population of solutions. Each solution is a candidate set of hyperparameters for the deep neural network. The population is initialized with random values within the specified ranges for each hyperparameter.
2. Parameter Determination: Set the parameters for the optimization process.

#### Iterative Optimization

3. Evaluate Fitness: For each solution in the population, compute the fitness value. (since optimization algorithms generally minimize the objective function).
4. Update Solutions: For each solution in the population, perform the following steps: Generate a random number.

If the random number is greater than or equal to 0.5: Use the JAYA Optimization Algorithm (JA) to update the position of the solution. JA is known for its simplicity and effectiveness, driving the search towards better solutions by minimizing the difference between the current solution and the best solution in the population.

Else: Use the Harris Hawk Optimization (HHO) to update the position of the solution. HHO simulates the hunting behavior of Harris hawks, involving

strategies such as surprise pounce and encircling prey, which helps balance exploration and exploitation.

5. Iteration: Repeat the process of fitness evaluation and position updating for a number of iterations or until a convergence criterion is met.
6. Final Selection: Find the Optimal Solution: After completing all iterations, identify the best solution from the population. This solution should have the highest fitness value.

In the flow chart shown above there is a decision variable num with the threshold of  $\geq 0.5$  to choose between the two optimization algorithms, Jaya Optimization and Harris Hawk optimization. This threshold-based decision point serves as a key component in determining which algorithm employ based on output probabilistic mechanism. The decision variable num is crucial in guiding the flow of the process and is based on threshold of  $\geq 0.5$ . This threshold is applied to a probabilistic or scoring outcome generated by the algorithm. When the outcome is equal to or exceeds 0.5, flow chart directs the process towards the utilization of Jaya optimization algorithm. Conversely, if the outcome falls below the 0.5 threshold, the flow chart steers the process towards implementation of Harris Hawk optimization algorithm. The threshold of 0.5 is chosen because it divides the probability space into two equal parts. This means there is an equal chance (50%) of selecting either JA or HHO. The choice of Num and the threshold of 0.5 are used to introduce a random but balanced decision-making process to select between the two optimization algorithms (JA and HHO). This approach leverages the strengths of both algorithms and can help in achieving a more robust optimization by probabilistically alternating between them.

#### **4.5 Summary:**

Harris Hawk Optimization (HHO) is a powerful metaheuristic algorithm for hyperparameter tuning in deep neural networks, particularly for classification tasks. Its ability to balance exploration and exploitation, handle high-dimensional and complex search spaces, and adapt dynamically to the optimization landscape makes it an effective tool for improving classification performance. By leveraging HHO, we can achieve better-tuned models that generalize well to unseen data, leading to more

accurate and robust classification results. This hybrid approach leverages the strengths of both HHO and JA to efficiently navigate the search space and find optimal solutions for hyperparameter tuning. By balancing exploration and exploitation, the algorithm can effectively explore new regions leading to better performance in classification tasks with deep neural networks.

## CHAPTER 5

### OVERALL SOLUTION

---

The Deep Neural Network with the novel beta entropy for feature extraction and HHHJO for optimized hyperparameter selection is the solution provided by the work reported in this thesis. Hyperparameter tuning plays a critical role in determining the performance of deep learning algorithms. Achieving optimal performance requires careful consideration of several algorithm-specific parameters. While algorithmic adjustments can significantly improve computational efficiency, challenges such as reduced population diversity and the risk of becoming trapped in local optima remain. The JAYA Optimization (JA) algorithm, which requires no initialization, is integrated into the hybrid HHHJO framework to address these challenges. By combining these techniques with HHO, the proposed hybrid algorithm effectively enhances performance and resolves key optimization issues. The newly developed HHHJO algorithm is employed to fine-tune key parameters of deep learning algorithms, including the learning rate, number of epochs, and the number of hidden neurons. Additionally, it optimizes Fuzzy C-Means (FCM) parameters, such as epsilon, maximum iteration counts, and the fuzziness coefficient. The HHHJO algorithm leverages exploration and exploitation mechanisms to enhance the performance of the Harris Hawks Optimization (HHO) algorithm.

#### **5.1 Optimized Deep Neural Network for the Breast Cancer Detection in Thermographic Images.**

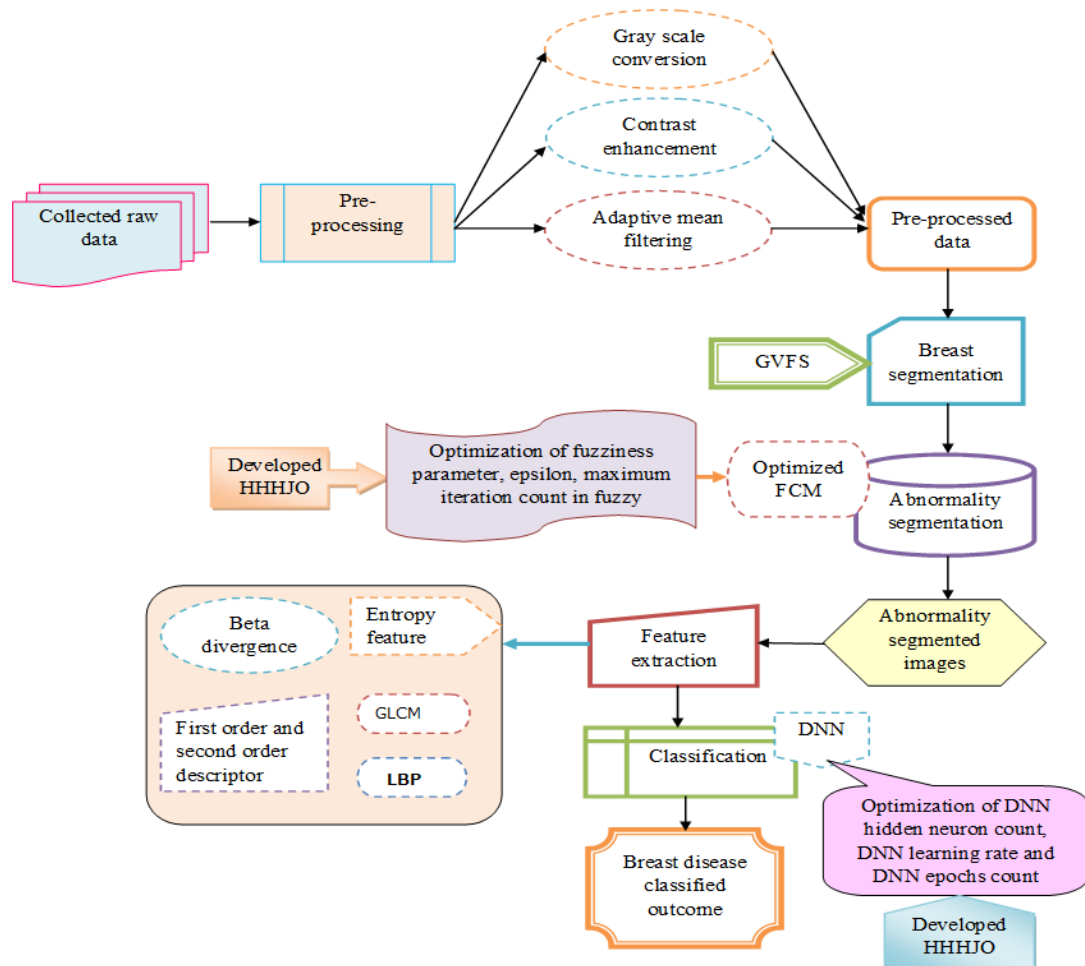
In order to diagnose breast cancer in people, a novel thermograph breast cancer detection model that uses mathematical model backed deep learning algorithm has been created [73]. The thermogram raw pictures that are utilized for the examination of breast cancer are obtained from conventional resources, and they are immediately delivered as the input to the pre-treating step. This phase analyses the images to determine whether or not breast cancer is present. During the pre-processing step, the raw pictures go through a series of transformations, including grayscale conversion, adaptive mean filtering, and contrast enhancement. After that, the collected pre-processed pictures are transferred to the breast subdivision phase, where breast image subdivision is passed out with the assistance of GVF (Gradient Vector Flow) in

order to get segmented chest images. Later on, during the abnormality segmentation phase, breast segmented images are used as the input for OFCM to segment the irregularity images. During this phase, developed HHHJO performs optimization by tuning the parameters including epsilon, maximum iteration count, and the fuzziness parameter in fuzzy in order to improve the anomalies by segmentation efficiency. The segmented abnormality pictures may be obtained by first decreasing the variance between the segmented deformity images and then growing the entropy of the segmented abnormalities imageries. In next step, the segmented abnormal pictures are put through a phase that is called feature extraction. During this phase, the features are extracted with the assistance of GLCM, first-order and second-order texture descriptors, LBP, entropy feature, and beta entropy. After this, the extracted characteristics that were gathered are brought into the categorization step. The newly developed ODNN is put to use in the process of breast cancer detection. This process involves optimization, which is carried out with the assistance of the newly designed HHHJO method. We have employed DNN with 2 hidden layers and with optimized hyperparameters, including the concealed neuron count, the learning rate and the number of epochs, for training. It is tested for the improvement in the outcome in Chapter 6. This method is used to tune the concealed neuron count of DNN, the learning rate of DNN, and the number of epochs for DNN. Increasing the accuracy rate and getting a superior breast cancer detection rate both contribute to the achievement of the detection result.

A separate feature extraction process is needed in a DNN to handle the complexity of raw data. Improved feature extraction reduces the amount of data to be processed, leading to significant savings in computational resources. Reduced data dimensionality translates to faster training and enhanced learning efficiency, improving model performance. It also leverages domain-specific insights and allows for effective processing of different data types. In this approach, we have not employed the concept of transfer learning, which is important for reducing feature inference times, especially in applications requiring real-time processing.

We have employed feature extraction separately. Our aim was not to reduce features but to combine them. When we combined all four types of features together, we observed improved accuracy and needed to verify the significance of beta entropy

as a feature. We adapted the strategy by adding features incrementally and checking the impact of the beta entropy feature. We first examined the results for statistical features, then combined statistical and texture features. Next, we combined three types of features: statistical, texture, and entropy. Finally, we combined all the features—statistical, texture, entropy, and beta entropy—to see how much the results improved with these additions. The architectural view of the thermogram-based breast cancer detection ideal is presented in Fig. 5.1.



**Fig. 5.1:** Architectural view of developed thermogram-based breast cancer detection model

An IR database, nominated by us as the **Database for Mastological Research with Infrared Images (DMR-IR)**, has been developed. The DMR-IR contains IR images, digitized mammograms, and clinical data obtained from patients at the Antônio Pedro University Hospital. These patients are from the screening department as well as the gynecologic department. Therefore, the DMR-IR includes data from both healthy patients and those with breast diseases, including cancer. The IR images are captured



using a FLIR thermal camera, model SC620, which has a sensitivity of less than 0.04°C and a capture range of −40°C to 500°C. The IR images have a resolution of 640×480 pixels. The acquisition of the images and their use in research have been approved by the Ethical Committee of the hospital and registered with the Brazilian Ministry of Health under number [insert number]. The DMR-IR is accessible through an online, user-friendly interface (<http://visual.ic.uff.br/dmi>) for managing and retrieving information from breast exams and clinical data of voluntary patients.

## 5.2 Input Image Pre-processing

The collected breast thermograph raw images  $IMG_y^{inp}$  are offered as the input for pre-processing phase. Here, it utilized grayscale conversion, adaptive mean filtering and contrast enhancement approach to perform pre-processing in the raw thermograph breast images.

**Conversion to grey scale:** While capturing the picture, it distributes the illumination restrictions in a random manner and at various times or places, which results in the creation of non-uniform grayscale photographs. The routine procedure of transforming a color picture into a grayscale image is known as "grayscale conversion." After it is done, the value that is utilized to represent one hue out of an image's 3-color channels is given its average value. The pre-fabricated data that was obtained from the grayscale conversion are referred to as and then supplied to the breast image segmentation step of the procedure.

**Adaptive mean filtering:** The adaptive Wiener filter is what gets rid of the high frequency noise while keeping the edges intact, and this is how it is able to accomplish this goal. The data that was obtained and processed by adaptive mean filtering are denoted as  $IMG_y^{pre}$  and then passed on to the breast segmentation step afterwards.

**Enhancement of contrast of the digital image:** Several different techniques for enhancing contrast are used in order to improve the quality of a parameter that has a low contrast. The primary goal of this method is to maintain the average brightness of the picture that is being read in, and it also adjusts the contrast of thermogram breast images in the specific places where they are being shown. In the outset, the RGB picture input channel is converted into the HSI color channel. After that, the value of the

intensity is split into two sub-parameters, which are referred to as the high group and the low group. These sub-parameters are calculated with the assistance of the golden section search model, and their results are described in Eq. (32,33) as

$$\gamma_{hh} = \{\gamma(n) | x > \gamma_q\}, \quad (32)$$

$$\gamma_{ll} = \{\gamma(n) | y \leq \gamma_q\}, \quad (33)$$

The higher intensity is given as  $\gamma_{hh}$  and the lower intensity is indicated as  $\gamma_{ll}$ , the threshold intensity value is presented as  $\gamma_q$  and it can separate the input image into two sub-images. Once the image intensity estimation for two sub-parameters is received, a new combination is performed among them to achieve an enhanced intensity rate and it is represented in Eq. (34).

$$\gamma_{ehnc}(n) = \gamma_{ll} + (\gamma_{hh} - \gamma_{ll}) \times \zeta(n) \quad (34)$$

The cumulative intensity is represented as  $\zeta(n)$  for  $n$ th pixel. The enhanced basic value of saturation as well as hue and intensity level are integrated and converted into an RGB channel to offer the final output image. The pre-processed data attained from contrast enhancement is given  $IMG_y^{pre}$  and they are further provided to the breast segmentation stage.

### 5.3 GVF (Gradient Vector Flow) -based breast segmentation with optimized abnormality segmentation for breast cancer detection: GVF-based Breast Segmentation

The picture  $IMG_y^{pre}$  that has been pre-processed is used as the input for the breast segmentation step of the GVF [45] approach, which is used to segment the breast area. The fundamental presumption is that both of the breast areas are substantially round and similar in shape. The GVF is built with two ellipses serving as the starting point each time. The traditional snakes are shown as curves  $(a(u) = [d(u), I(u)], u \in [0,1])$ , and both the outside and the inner image domains are taken into account. It is able to navigate the internal forces that were formed by the external resources, and the curve that it follows is certified by the picture data. The ranges of contour capture in binary image counters may be improved with the use of the GVF. The preliminary

results are presented here based on a number of previous studies and our current body of information for both breasts.

The elasticity of the snake is shown as and merged with the second derivative term in this presentation. In this case, the step size is denoted by the symbol  $\gamma$ . The energy-related component of the sailing factor is denoted by the symbol  $k$ . The intensity-oriented potential term  $wEline$  makes use of the weighting component. “For the edge-oriented potential term, the weighting factor is what's going to be utilised. The weighting factor is used for the prospective period of termination. An illustration of the iteration count is provided so that the location of the counter may be determined. The gradient  $wEline$  is then validated using the edge map that was provided before. A significant amount of reliance is placed on the GVF in order to convey the deformation of snakes at the boundary edges. The abnormality-segmented phase receives its input from the breast-segmented pictures  $IMG_y^{seg}$  that have been obtained before.

#### 5.4 Optimized FCM-based Abnormality Segmentation

The abnormality-segmented phase receives its input  $IMG_y^{seg}$  in the form of the pictures that were obtained from the breast segmentation process. The FCM [74] is a well-organized technique to segmentation that enables clustering via the use of a membership value function. The objective function is the FCM clustering objective, which measures the quality of the clustering. Lower values indicate better clustering. The HHHJO algorithm will be used to optimize the solution vector by iteratively updating it to minimize the FCM objective function. We utilize FCM for abnormality segmentation using a variety of algorithmic parameters including epsilon, fuzziness parameter, and maximum iteration, all of which are ideally optimized with the assistance of constructed HHHJO. The fuzziness parameters are tuned in the range of [2,5], epsilon is tuned by the range of [0.01,0.99] and the maximum iteration count of fuzzy is tuned in the range of [5,50].

The name of the newly developed technique is the Optimized FCM (OFCM). The following expression is the minimization objective function that must be used in order to assign a pixel to a certain cluster:

$$U_{cf}(G, H, I) = \sum_{r=1}^D \sum_{p=1}^E g_{r,p}^w e^2(i_r, h_p) \quad 1 < w < \infty \quad (35)$$

where the vector  $I = \{i_1, i_2, \dots, i_D\}$  consists of  $D$  pixels, multidimensional cluster center with  $E$  clusters is given as  $H = \{h_1, h_2, \dots, h_E\}$ , the Euclidean distance of  $i_r$  from the center of  $p^{\text{th}}$  cluster is presented as  $e^2(i_r, i_p)$ , the fuzzy membership matrix is given as  $G$  and the blur exponent is given as  $w$ ,  $w > 1$ . The membership degrees denoted as  $g_{r,p}$  fulfills the following conditions presented in Eq. (36) as

$$\begin{cases} 0 \leq g_{r,p} \leq 1 \\ \sum_{p=1}^E g_{r,p} = 1 \\ \sum_{r=1}^D g_{r,p} \leq D, r = 1, 2, \dots, D, p = 1, 2, \dots, E \end{cases} \quad (36)$$

The cluster center, as well, the membership degrees are updated iteratively. The formulation of membership matrix and the cluster center is given in Eq. (37) and Eq. (38) as under:

$$g_{r,p} = \frac{1}{\sum_{v=1}^E \left( \frac{e_{r,p}}{e_{r,v}} \right)^{\frac{2}{w-1}}} \quad (37)$$

and

$$h_p = \frac{\sum_{r=1}^D g_{r,p}^w i_r}{\sum_{r=1}^D g_{r,p}^w} \quad (38).$$

The fitness function to decide the parameters in OFCM is based on minimization of the variance and maximization of the entropy and given below in Eq. (39):

$$Ft_1 = \arg \min_{Fp_n^{FCM}, Ep_e^{FCM}, IC_w^{FCM}} \left( var + \left( \frac{1}{etpy} \right) \right) \quad (39)$$

Where  $var \rightarrow$  the variance,  $etpy \rightarrow$  the entropy between the segmented images,  $Fp_n^{FCM}$  is the fuzziness parameters,  $Ep_e^{FCM}$  is the epsilon parameters and  $IC_w^{FCM}$  is the maximum iteration.

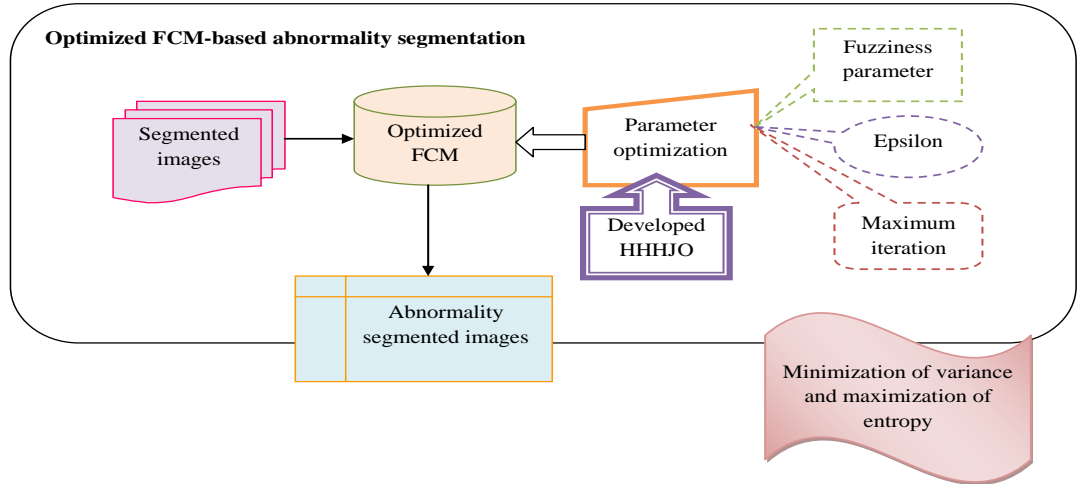
Variance is presented in Eq. (40) as

$$K^2 = \frac{\sum (vg_i - \bar{vg})^2}{L-1} \quad (40)$$

where  $K^2$  is the sample variance,  $\bar{u}g$  is the whole means of observation value shown as, the value of one observation is provided by  $vg_i$ , and the number of observations is presented as correspondingly. The expression for entropy, which can be found in Eq. (41), reads as "the degree of randomness achieved in a picture."

$$etpy = -\sum_h \sum_i T_d(C_i) \ln T_d(C_i) \quad (41)$$

where  $T_d$  represents the probability, and  $\ln T_d$  (natural log) denotes the logarithmic magnitude of the probability ( $C_i$ ). The abnormally segmented pictures that were able to be obtained afterwards serve as the input for the phase that is dedicated  $IMG_y^{Asg}$  to the process of feature extraction. On display in Figure 5.2 is a diagrammatic representation of the improved FCM-based BC segmentation.



**Fig. 5.2:** Diagrammatic representation of optimized FCM-aided breast cancer segmentation

## 5.5 Enhanced thermogram breast cancer detection using weight-optimized deep neural network: Feature Extraction

The pictures  $Ep_e^{FCM}$  that have been segmented based on abnormalities are given as the input for the phase of feature extraction. During this phase, properties such as GLCM, first-order & second-order texture descriptors, LBP, entropy feature, and beta entropy are applied to obtain the extracted features. These features, along with others, will be detailed in the following paragraphs.

**GLCM features [37]:**

The abnormalities segmentation pictures  $Ep_e^{FCM}$  are used as the input for this procedure, and the breast textural characteristics are then extracted. To obtain the object texture, which represents the pixel display in spatial connection, the GLCM is put to extensive use and is widely employed. The co-variance matrix is validated with the pixel value in order to get the object textures. The feature vectors of the input pictures are verified with respect to contrast, difference, equality, vitality, correlation, and angular 2<sup>nd</sup> moment (ASM), all of which are expounded on as described in the following:

**Contrast:** It estimates the local variation in images. When the nearby pixel is changed with a superior value, then the contrast is validated by Eq. (42) as

$$cntrst = \sum_{d,e=0}^{q-1} Q_{de} (d - e)^2 \quad (42)$$

Where  $Q_{de}$  is the pixel of the given copy for the position & variance  $d$  and  $e$ , respectively.

Attained inconsistencies are used to verify the weights, and they are provided in Eq. (43), as follows: Similarity: It is estimated that "by computing the weights in the contrast quantity as pixel travels away from the diagonal the weights rises substantially."

$$Dissmlrty = \sum_{d,e=0}^{q-1} Q_{de} |d - e| \quad (43)$$

**Homogeneity:** is calculated "when numerous pixels shown in the picture have the same colours," and its results are confirmed in the equation (44).

$$Homogeneity = \sum_{d,e=0}^{q-1} \frac{Q_{de}}{1 + |d - e|^2} \quad (44)$$

**Energy:** It is "the uniformness the picture with square elements summing in the GLCM," and it is calculated by using the equation (45).

$$engy = \sum_{d,e=0}^{q-1} (Q_{de})^2 \quad (45)$$

**Correlation:** It shows "how to attach a pixel to its neighbor pixels in the complete picture," and the estimates are given in the form of an equation (46).

$$correlation = \sum_{d,e=0}^{q-1} \frac{Q_{de}(d-\mu)(e-\mu)}{\sigma^2} \quad (46)$$

**ASM:** It highlights "the homogeneity qualities of a digital image size or the size of the vicinity of each element of the occurrence matrix," and those features are stated in the equation below (47).

$$ASM = \sum_{d,e=0}^{q-1} \frac{Q_{de}}{1+|d-e|} \quad (47)$$

In the GLCM matrix, the quantities of grey values are expressed as, and the average value of the whole pixel is displayed as  $\mu$ . The characteristics that were obtained from the GLCM  $FT_c^{glcm}$  are then sent to the concatenation step.

**1st-order and 2<sup>nd</sup>-order textural descriptors [75]:** In the interest of obtaining the textural aspects of first-order and second-order descriptors, it takes the abnormality segmented features  $IMG_y^{Asg}$  as its input. The following is an explanation of the characteristics that are included inside the first-order descriptor. These features are as follows: minimum, maximum, mean, median, and standard deviation.

**Minimum:** it is defined as "the minimum value in a set of values" and it is provided in Eq. (48),

$$Minimum = \sum_{d,e=0}^{q-1} Q_{de} (d - e) \quad (48)$$

The word in this context both d refers to the variance of the picture and e denotes the location of the image.

**Maximum:** It is termed as "a highest number for a set of data" and given in Eq. (49)

$$Maximum = \sum_{d,e=0}^{q-1} Q_{de} (d + e) \quad (49)$$

**Mean:** it is refer "the size of the dispersion of an image" and given in Eq. (50)

$$\mu = \sum_{d,e=0}^{q-1} Q_{de} \quad (50)$$

**Median:** It is "the mid-value of the set of data, when organised in order and provided in Eq. (51).

$$Median = \sum_{d,e=0}^{q-1} Q_{de} \left( \frac{\mu}{2} \right) + 1 \quad (51)$$

**Standard deviation:** It is “a measure of how dispersed the data is with the mean” and is offered in Eq. (52).

$$\sigma = \sum_{d,e=0}^{q-1} \sqrt{Q_{de}(Q_{de} - \mu)} \quad (52)$$

The second-order descriptors hold features like variance, kurtosis and skewness, which are explained below.

**Variance:** It is already discussed in Eq. (53).

**Kurtosis:** It is defined as “the level of sharpness relatively curves on the histogram of an image” and it is presented in Eq. (53)

$$\alpha_4 = \sum_{d,e=0}^{q-1} \frac{1}{\sigma^4} (Q_{de} - 3)(Q_{de} - \mu)^4 \quad (53)$$

**Skewness:** It is referred as “the relative level of slope of the curve on the histogram of an image” and it is represented in Eq. (54)

$$\alpha_3 = \sum_{d,e=0}^{q-1} \frac{1}{\sigma^3} (Q_{de})(Q_{de} - \mu)^3 \quad (54)$$

The acquired features from first-order and second-order descriptors  $FT_v^{fs}$  are further provided to concatenation.

**LBP [39]:** It is a very productive technique for extracting the features, and it incorporated the local intensity variation of the abnormality segmented picture  $IMG_y^{Asg}$  that was supplied. This resulted in improved discriminating characters being extracted. The intensity of the segmented picture J is shown as a range of pixels  $(t_z, u_z)$  and does not take into account the centre-pixel. The value of the LBP for the picture pixel  $(t_z, u_z)$  may be found in the equation (55).

In this case, the features based on entropy are obtained from the pictures  $IMG_y^{Asg}$  that have been segmented based on abnormalities. The following is an explanation of the many types of acquired entropy-based attributes: minimum entropy,



sample entropy, maximum entropy, Shannon entropy, and approximation entropy. Renyi entropy, maximum entropy, and sample entropy are also included.

**Renyi entropy:** It is otherwise said to be collision entropy. They are satisfied with the condition  $\alpha = 2$  and equated in Eq. (56).,

$$RE_p(V) = -\log \sum_{v_k=1}^N q^3(qc(v_k))^2 \quad (56)$$

Where  $c(v_k)$  defines the probability distribution function.

**Minimum entropy:** In minimum entropy, the order  $\alpha$  is denoted as  $\infty$  i.e.  $\alpha = \infty$ . The order  $\alpha$  move towards  $\infty$  then, the attained probability for Renyi entropy for  $V$  probable result is showcased in Eq. (57)

$$JEp_1 \lim_{\alpha \rightarrow 1} JEp_{\infty}(V) = -\log_2 \max qc(v_k) \quad (57)$$

**Sample entropy:** It indicates the complexity measures and the sample entropy are presented as  $SmEp$  and given in Eq. (58). The template vector with length is given as  $\alpha$  for  $V_a(k) = \{v_k, v_{k+1}, v_{k+2}, \dots, v_{k+\alpha-1}\}$  and the distance is indicated as  $Dis[V_a(k), V_a(p)] (k \neq p)$  later considered for Chebyshev distance.

$$SmEp = -\log \frac{x}{y} \quad (58)$$

Here, the term  $x$  indicates the template vector pair count  $Dis[V_{a+1}(k), V_{a+1}(p)] < n$  and the term  $y$  presents the template vector pair count  $Dis[V_a(k), V_a(p)] < n$ , respectively.

**Maximum entropy:** Here, the order  $\alpha$  is denoted as 0 and so,  $\alpha = 0$ . Here, the overall probability is given as  $qc$  and max entropy or Hartley is represented by  $He_0$ . The log count is given  $\log_2 m$  and the positive probability is presented  $v$ , respectively.

**Shannon entropy:** It is utilized to validate the uncertainty presented in the information source and also it is represented  $SHEp$  in Eq. (59).

$$SHEp(F_t^{as}) = -\sum_{v_j=0}^{N-1} qc(v_k) \log_2(qc(v_k)) \quad (59)$$

Here, the distant gray-level count is indicated by  $F = 2^r$ , the original image presented as  $IMG_y^{Asg}$  and the term  $v_k$  the probability of occurrence.

**Approximate Entropy:** The approximation entropy is denoted by  $ApEn$ . The number sequence is given as  $b = \{b(1), b(2), \dots, b(N)\}$  for the length  $N$ , and the non-negative integer is denoted as  $\alpha$  and uses the condition  $\alpha \leq N$ . The positive real numbers are represented as  $s$  then the blocks are illustrated as  $h(k) = \{b(k), b(k+1), \dots, b(k+\alpha-1)\}$  and  $h(p) = \{b(p), b(p+1), \dots, b(p+\alpha-1)\}$  also the distances presented within them are validated as  $Dis[h(k), h(p)] = \max_{s=1,2,\dots,\alpha} (|b(k+s-1) - b(p+s-1)|)$ .

Later, the value  $G_k^\alpha = (\text{number of } p \leq N - \alpha + 1 \text{ and } Dis[h(k), h(p)] \leq n/(-\alpha + 1))$  is used to validate. The numerator is given as  $G_k^\alpha$  and it is counted between the solutions  $n$  and the block length  $\alpha$  and the identical blocks are given in Eq. (60).

$$\varphi^\alpha(n) = \frac{1}{N-\alpha+1} \sum_{k=1}^{N-\alpha+1} \log G_k^\alpha(n) \quad (60)$$

It is useful for interpreting  $ApEn(a, n, N)(b) = \varphi^a(n) - \varphi^{a+1}(n)$  having  $a \geq 1$  and also  $ApEn(a, n, N)(b) = -\varphi^1(n)$ . The frequency of logarithmic is validated  $ApEn(a, n, N)(b)$  through the longitude blocks  $a$  which stay near the subsequent position and the negative value of  $ApEn$  is presented as  $-ApEn(a, n, N)(b) = \varphi^{a+1}(n) - \varphi^a(n)$  is the average over  $j$  of the logarithm conditional probability of  $|v(k+n) - v(j+n)| \leq s$ , and they are checked  $|b(p+s) - b(k+s)| \leq n$  for the range of  $(s = 0, 1, \dots, \alpha - 1)$ .  $ApEn(a, n, N)$  Showcase the statistical estimator for the parameter  $ApEn(a, n)$  in Eq. (61).

$$ApEn(a, n) = \lim_{N \rightarrow \infty} [\varphi^a(n) - \varphi^{a+1}(n)] \quad (61)$$

The acquired features from entropy features  $FT_k^{EpF}$  are further provided to concatenation.

**Beta entropy:** The beta entropy [38] is mainly utilized to increase the entropy measure with the consideration of probability value because the outlier trail is weighted by itself and it is considered for  $\beta > 0$  in Eq. (62).

$$D_{\beta}(P||Q) = \frac{1}{\beta} \int \left( P^{\beta}(x) - Q^{\beta}(x) \right) P(x) dx \quad (62)$$

The two-probability distribution is given as  $g(x)$  and  $w(x)$  in the above equation. The extracted features from beta entropy  $FT_z^{Bd}$  are further provided to concatenation. The different values of  $\beta$  can emphasize different aspects of beta distribution, making it more flexible tool for various application of machine learning statistics and information theory.  $\beta > 1$  has a robust similarity measure against outliers between two probability distribution. Here, outliers have low probability and exist near zero probability. For  $\beta > 1$ , the beta entropy becomes more sensitive to the tail behavior of the distribution. This means it can better handle outliers and heavy-tailed distributions, which is essential for robustness in various applications.  $\beta > 1$ , the beta entropy puts more emphasis on larger probabilities in the distribution. This can help in scenarios where the focus is on the dominant features or events in the data, leading to a more robust analysis against noise and small fluctuations. As  $\beta$  varies, beta entropy smoothly transitions between different entropy measures, allowing for a controlled and gradual adjustment. This smooth transition provides a robust mechanism to fine-tune the entropy measure to specific data characteristics and noise levels.

In this phase, the features  $\{FT_c^{glcm}, FT_v^{fs}, FT_d^{Lbp}, FT_k^{EpF}, FT_z^{Bd}\}$  are extracted from GLCM, first-order and second-order texture descriptors, LBP, entropy feature and beta entropy are provided for concatenation. The concatenated features are indicated by  $FT_y^{Cf} = \{FT_c^{glcm}, FT_v^{fs}, FT_d^{Lbp}, FT_k^{EpF}, FT_z^{Bd}\}$  and they are further provided to classification phase.

## 5.6 Basic DNN Model

The Feed orward Neural Network (FFN) and the Multilayer Perceptron (MLP) methods are combined in the DNN [76] methodology. The DNN is made up of multiple neurons, all of which are linked to the other neurons in the network in a path that is forward and it is an acceptable representation for the DNN. The input vector of the DNN is shown as  $z = z1, z2, z2, \dots, zl$ , and their sizes are given as. In addition, the input is supplied as  $W(z)$ , and their size is. The validation of the hidden neuron is shown in Equation (63) and Equation (64).

$$u_g(d_g^{q+1}) = z(S_{ig} + I_g^{q+1}) \quad (63)$$

$$S_{ig} = j_i^q y_{ig}^{(q,q+1)} \quad (64)$$

The entire lower layers presented in the neurons are connected towards the  $g^{th}$  neuron and the term  $j_i^q$  is the neuron of  $i^{th}$  activation function in the  $q^{th}$  layer. The contribution function is given as  $S_{ig}$  in  $i^{th}$  activation region of  $q^{th}$  layer. The non-linear activation function is as  $z$ , the function weight is given as  $y_{ig}^{(q,q+1)}$  and the neuron bias is presented as  $I_g^{q+1}$ . A huge number of stacking hidden layers presented in MLP is said to be DNN and the mathematical representation of DNN with multiple hidden layers is presented in Eq. (65).

$$T_q(y) = T_q \left( T_{q-1} (T_{q-2} \left( \dots \dots (T_q(y)) \right) \right) \right) \quad (65)$$

The normal DNN design holds multiple hidden layers. The ReLU performance rate is fast and easy to train large number of hidden layers and the outcome achieved in the DNN is the trust value for node.

## 5.7 Optimized DNN-based Detection

The collected input features  $FT_y^{cf}$  from the feature concatenation phase are used in the process of creating DNN with hyperparameter tweaking produced using HHHJO for the uncovering of the cancer. This particular approach is referred to as Optimized DNN (ODNN). DNN offers a higher level of reliability and performance to the network than other providers. Because of the versatility with which it presents itself in the network input layer, it is able to readily engage itself in the querying of features, therefore assisting the user in obtaining particular features and offering a suggestion based on their requirements. Yet, an enormous quantity of data is necessary in order to obtain a high level of accuracy throughout the training phase. In order to address all of these problems, a new ODNN has been designed to classify breast cancer. This new ODNN was created by modifying the parameters of DNN using the HHHJO method, such as the number of hidden neurons, the learning rate, and the number of epochs. The classification results are obtained by using the fitness function described in Equation (66), which is written as

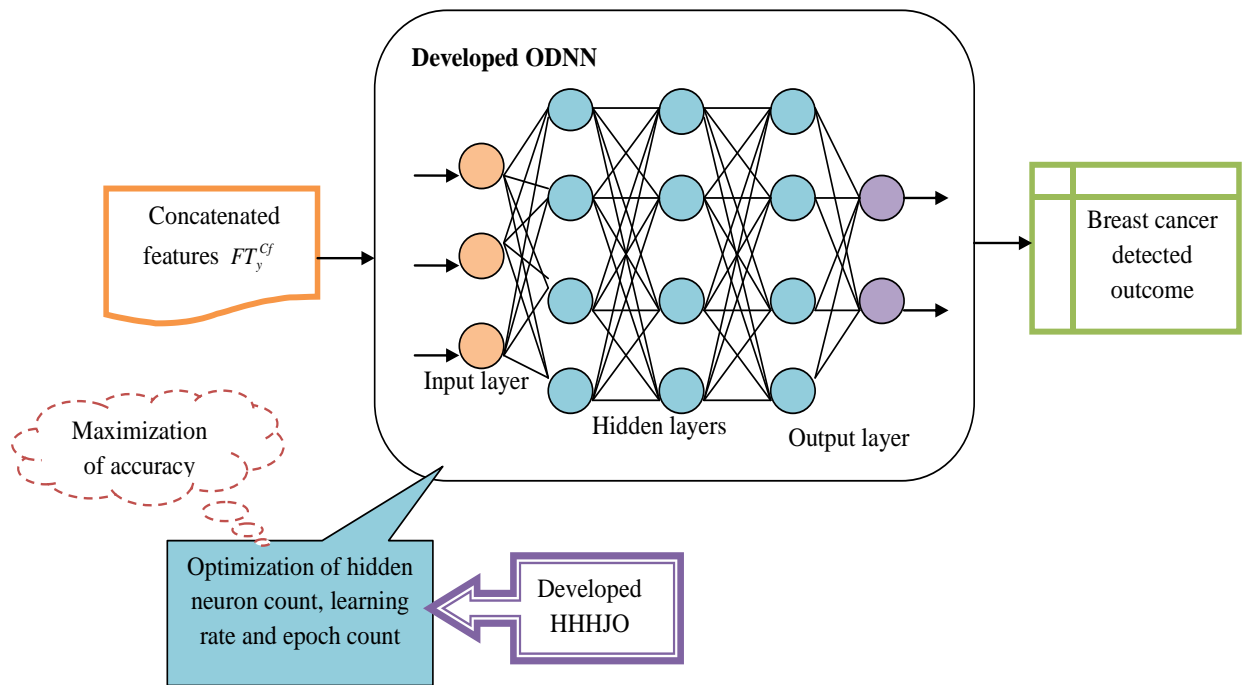
$$Ft_1 = arg \min_{HD_m^{DNN}, LR_f^{DNN}, EC_x^{DNN}} \left( \left( \frac{1}{ACCY} \right) \right) \quad (66)$$

where the  $HD_m^{DNN} \rightarrow$  the hidden neuron count of DNN,  $LR_f^{DNN} \rightarrow$  the learning rate of DNN and  $EC_x^{DNN}$  referred the epochs count of DNN. Optimization of hyperparameters such as the number of hidden neurons, learning rate, and the number of epochs in a Deep Neural Network (DNN) involves selecting the best set of values from the given ranges that minimize the loss function or maximize the performance metric on the validation set. An initial population with random hyperparameter combinations is selected, they are tuned in the range of [200,255], [0.01,0.99] and [50,100] respectively. The training and evaluation of DNN is carried out for each combination of population with the help of developed HHHJO. Select the best-performing combinations to form a new population. Repeat the process for a set number of generations or until convergence. We called it as an optimized DNN (ODNN).

Accuracy  $ACCY$  is dignified as the nearness of the measurements to a specific value and given by the Eq. (67) as

$$ACCY = \frac{(d_p + f_p)}{(d_p + f_p + g_n + h_n)} \quad (67)$$

where the true  $+ve$  and true  $-ve$  values are revealed as  $d_p$  and  $f_p$  respectively & false  $+ve$  and false  $-ve$  are given as  $g_n$  &  $h_n$  respectively. The developed ODNN-based cancer revealing model is represented in Fig. 5.3.



**Fig. 5.3:** Developed ODNN-based breast cancer detection model

## **5.8 Conclusion**

The problems that arise in Computer Aided Diagnosis (CAD) systems that use computer technology for abnormality level prediction in breast cancer images were covered in this chapter's discussion. The usefulness of the suggested approach in medical diagnosis was validated by an investigation that compared it to other segmentation and feature extraction techniques already in use. This investigation focused on the metrics of accuracy, sensitivity, specificity, and probability (both positive and negative).

## CHAPTER 6

### RESULT AND ANALYSIS

---

In order to diagnose breast cancer in people, a novel thermograph breast cancer detection model has been built by the use of heuristic methods in addition to deep learning techniques. The thermogram raw pictures that are used for the analysis of breast cancer are obtained from standard resources, and they are immediately delivered as the input to the pre-processing step. This stage analyses the images for breast cancer. During the phase known as "pre-processing," the breast thermographic pictures that were entered are subjected to various processing techniques such as grayscale conversion, adaptive mean filtering, and contrast enhancement. After that, the collected pre-processed pictures are transferred to the breast segmentation phase, where breast segmentation is carried out with the assistance of GVF (Gradient Vector Flow) in order to get breast segmented images. Later, in the abnormality segmentation phase, breast segmented images are offered as the input for optimized FCM to generate abnormality segmented images, where optimization is performed by developed HHHJO by tuning the parameters such as epsilon, maximum iteration count and fuzziness parameter in fuzzy to enhance the abnormality segmentation efficiency. The developed WO-DNN is utilized to detect breast cancer, where the optimization is performed with the help of the developed HHHJO approach for tuning the hidden neuron count of DNN, the learning rate of DNN and DNN epochs count. The detection outcome is attained by maximizing the accuracy rate and achieving a better breast cancer detection rate.

We have employed the DNN structure keeping the concept that number of hidden layers more than one can be called as Deep Neural Network. A simple DNNs can be more appropriate for non-spatial data, offering benefits in terms of computational efficiency, lower risk of overfitting, ease of implementation, and versatility. For the given problem CNN has been implemented with HHHJO optimization algorithm. The HHHJO is used to iteratively search the hyperparameter space that identifies the combination learning rate, epochs and performance matrix.

#### 6.1 Experimental Set-up

The experiment was done with total 1315 breast thermograms images of 640\*480 which has 870 healthy and 445 sick patients and 5 images for each patient.



The camera used to capture images was FLIR SC-620. The model analysis done on the proposed thermogram breast cancer identification model was developed in python. The efficacy analysis was executed in the developed model with the help of existing approaches. The maximum iteration count was kept 20 and the number of population count was kept 10 for experimental analysis. Distinct algorithms have been considered like (HHO) (Heidari, A. A. et al 2019) [26] and JAYA (Venkata Rao, R. et al. 2016) [27]. Various baseline classifiers like Decision Tree (DT) (Hu, Q., Che, X et al. 2012) [28], KNN (Zhang, S., Li, X et al. 2018) [29], SVM (Acharya, U R et al.2012) [6], NN (Liu, Y., Wang et al. 2011) [30] and DNN (Narayana Rao, K et al. 2021) [31] were considered for performance comparison.

## 6.2 Efficiency Metrics

The recommended thermogram breast cancer recognition model is computed with diverse qualitative metrics.

(a) Precision,  $PS$ , is referred as the fraction of true positive instances over the total instances and presented below in Eq. (68) as

$$PS = \frac{d_p}{d_p + g_n} \quad (68)$$

(b) F1-score,  $JN$ , is presented as the dimension of the accurateness in the showed test and showcased in Eq. (69) as

$$JN = 2 \times \frac{2d_p}{2d_p + g_n + h_n} \quad (69)$$

(c) False Negative Rate (FNR),  $RF$ , is the proportion of positives which yield negative test outcomes with the examination as explained in Eq. (70) below:

$$RF = \frac{h_n}{h_n + d_p} \quad (70)$$

(d) False Positive Rate (FPR),  $Rn$ , is referred as the ratio between the numbers of negative events wrongly characterized as positive (false positives) and the aggregate number of real negative events and it is represented as in Eq. (71) below:

$$Rn = \frac{g_n}{g_n + f_p} \quad (71)$$

(e) Specificity,  $Sf$ , is the proportion of negatives that are correctly identified and given in Eq. (72) as

$$Sf = \frac{p}{f_p + g_n} \quad (72)$$

(f) Negative Predictive Value (NPV),  $GS$ , is known as the totality of all peoples without illness in analysis and it is mentioned in Eq. (73) as

$$GS = \frac{f_p}{f_p + h_n} \quad (73)$$

(g) Sensitivity  $BA$  is the proportion of positives that are properly recognized and given in Eq. (74) as

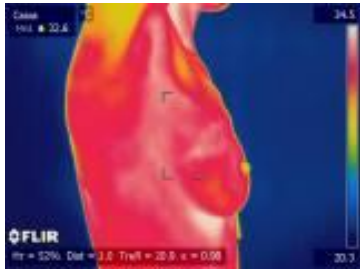
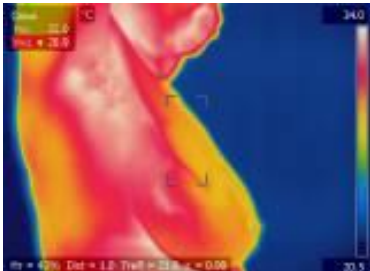
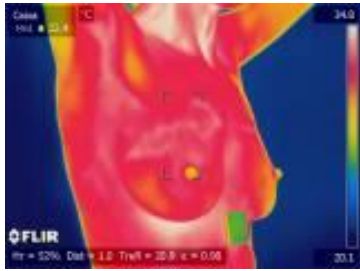
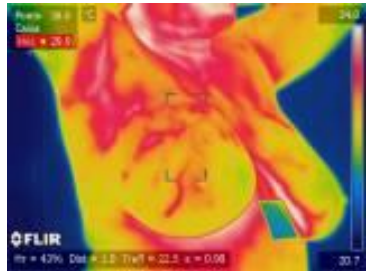
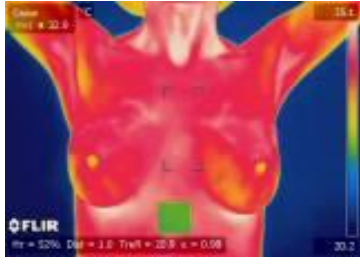
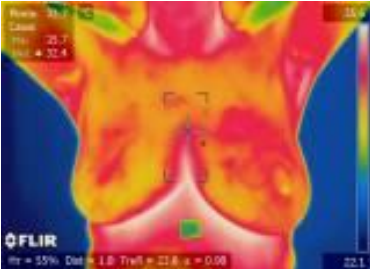
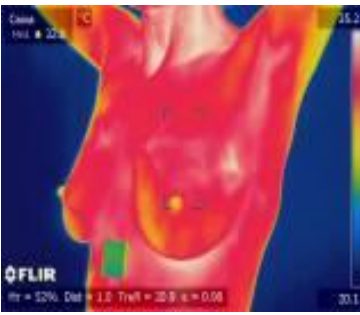
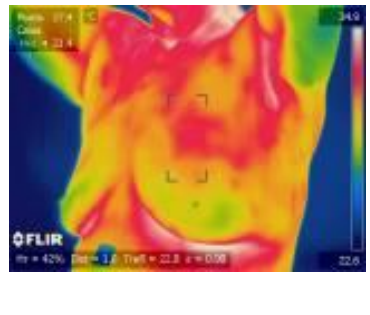
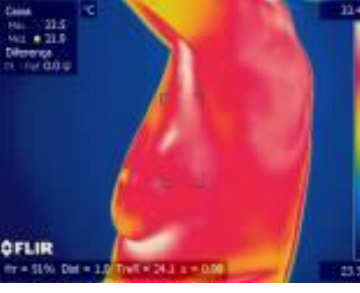
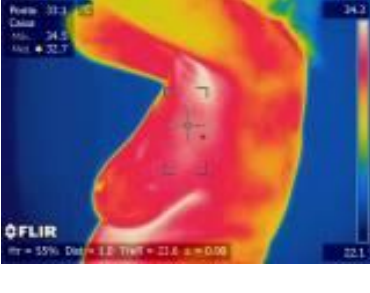
$$BA = \frac{d_p}{d_p + h_n} \quad (74)$$

(h) Matthew's correlation coefficient MCC,  $DN$ , is a portion of the quality of binary categorizations of analysis and it is represented in Eq. (75).

$$DN = \frac{d_p \times f_p - g_n \times h_n}{\sqrt{(d_p + g_n)(d_p + h_n)(f_p + d_p)(f_p + h_p)}} \quad (75)$$

### 6.3 Thermogram Dataset Description

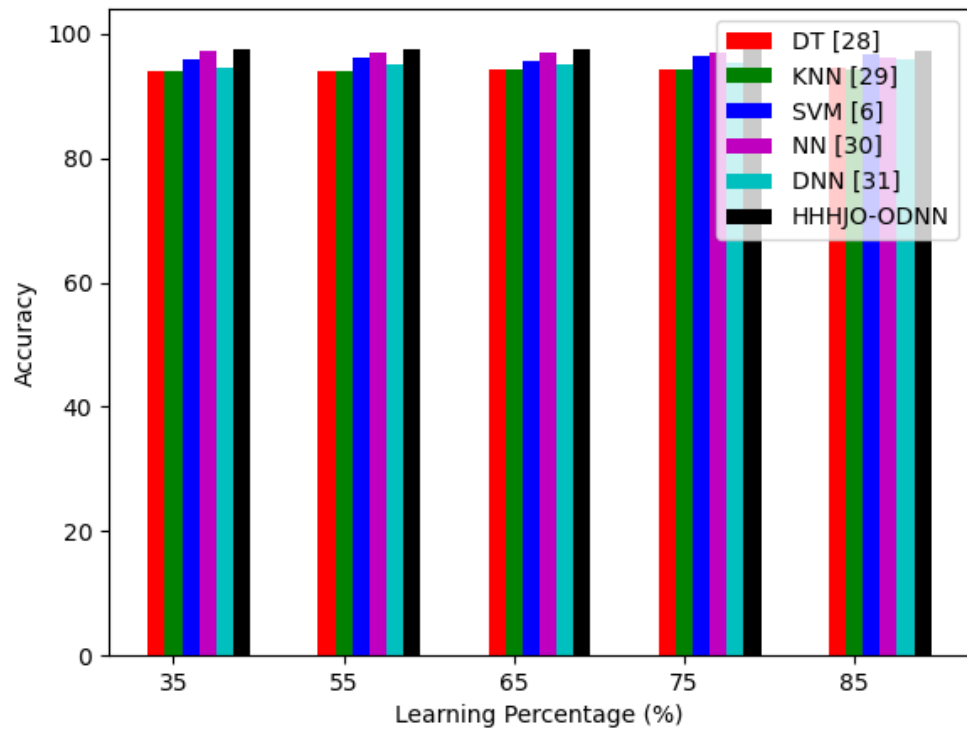
The dataset named 'Database for Mastology Research' (DMR) [71] is utilized to acquire the thermographic image dataset for the analysis in this section. The link preserves and handles mastological images through various imaging techniques to perform early breast cancer detection in individuals. Fig.6.1 shows a healthy and cancerous breast static thermograms from the dataset available from five different viewpoints. The five different thermographic images showing different orientations front image, left  $90^0$ , left  $45^0$ , right  $90^0$  and right  $45^0$ .

Sample images	Healthy	Breast cancer affected
1		
2		
3		
4		
5		
<b>Fig. 6.1:</b> Samples of breast cancer-affected and healthy images from thermogram dataset		

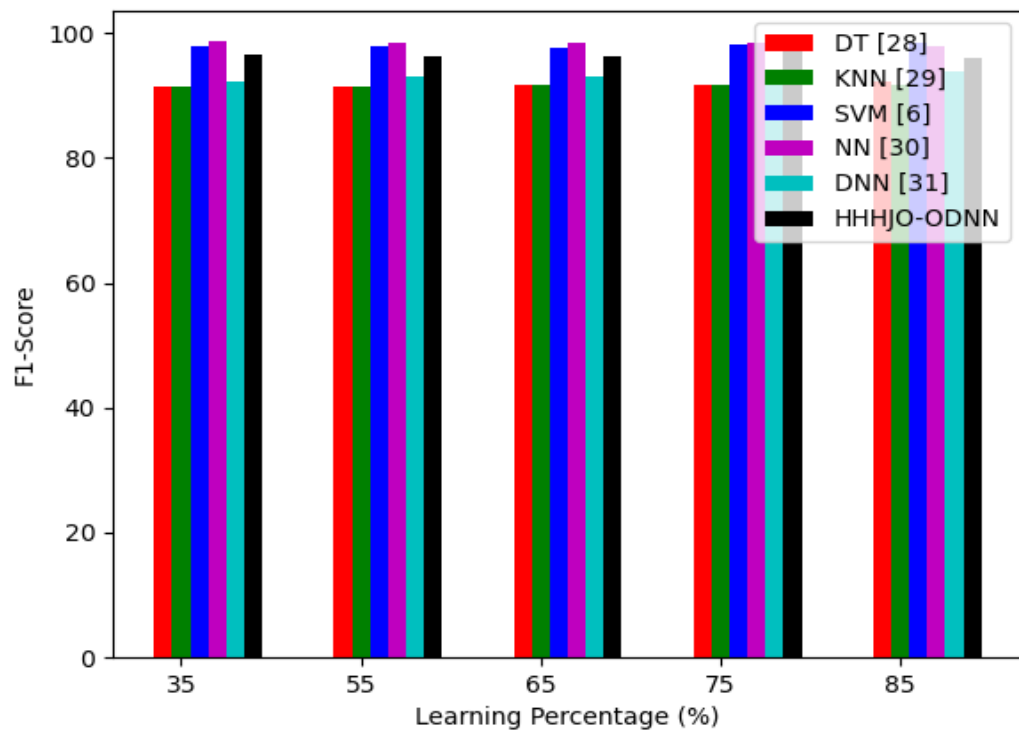
The acquired breast cancer thermograph raw images are denoted as  $IMG_y^{inp}$  and they are offered as the input for the pre-processing stage. Here,  $y = 1, 2, \dots, Y$ , where  $Y$  is the total number of breast thermograph images acquired for breast cancer detection. The acquired segmented images from the OFCM for the developed breast cancer detection model are displayed in Fig.6.2 This figure represents the original image, pre-processed image and abnormality segmented images in three consecutive rows. The columns are obtained by three different image samples.

Image Description ↓		Image Description 1	Image Description 2	Image Description 3
Original image				
Pre-processed image				
GVF based segmented				
Abnormality segmented				

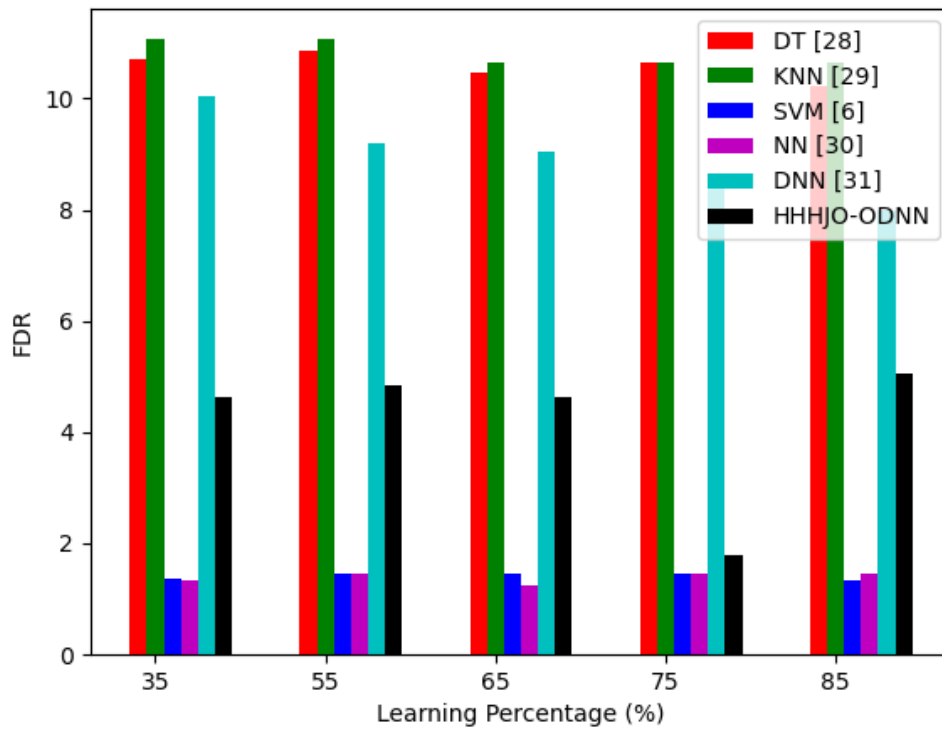
**Fig. 6.2:** Resultant abnormality segmented images from the Optimized FCM technique



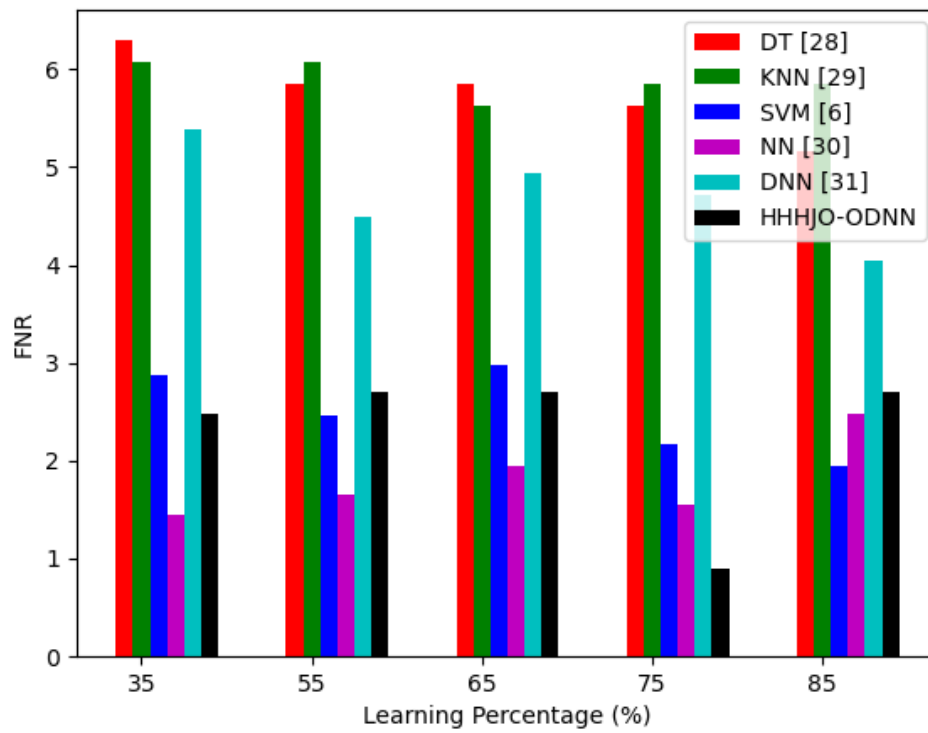
(a)



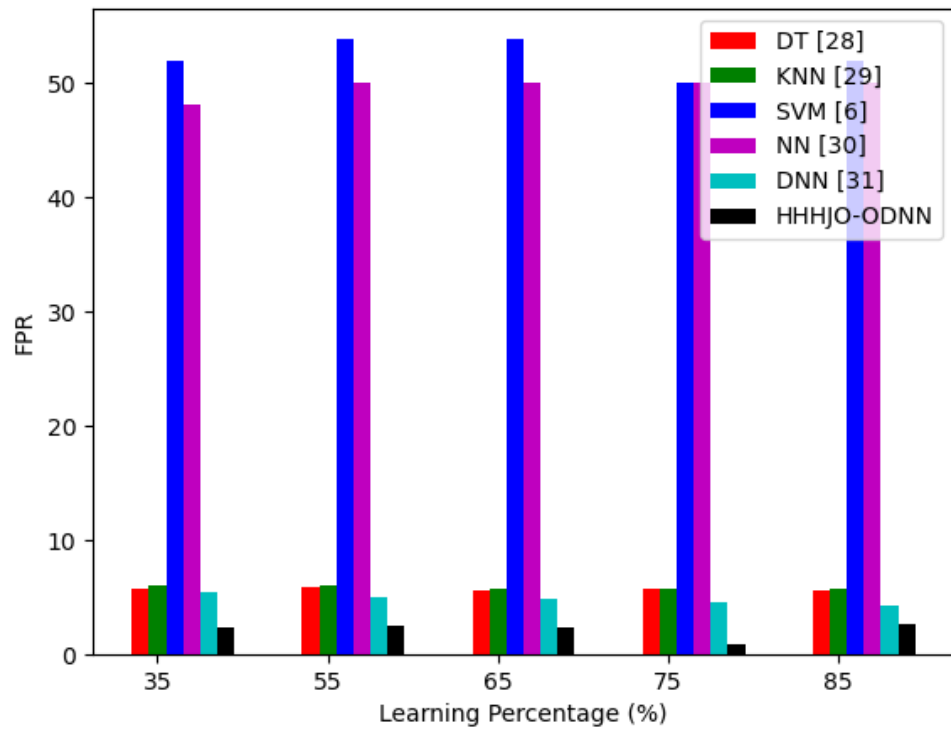
(b)



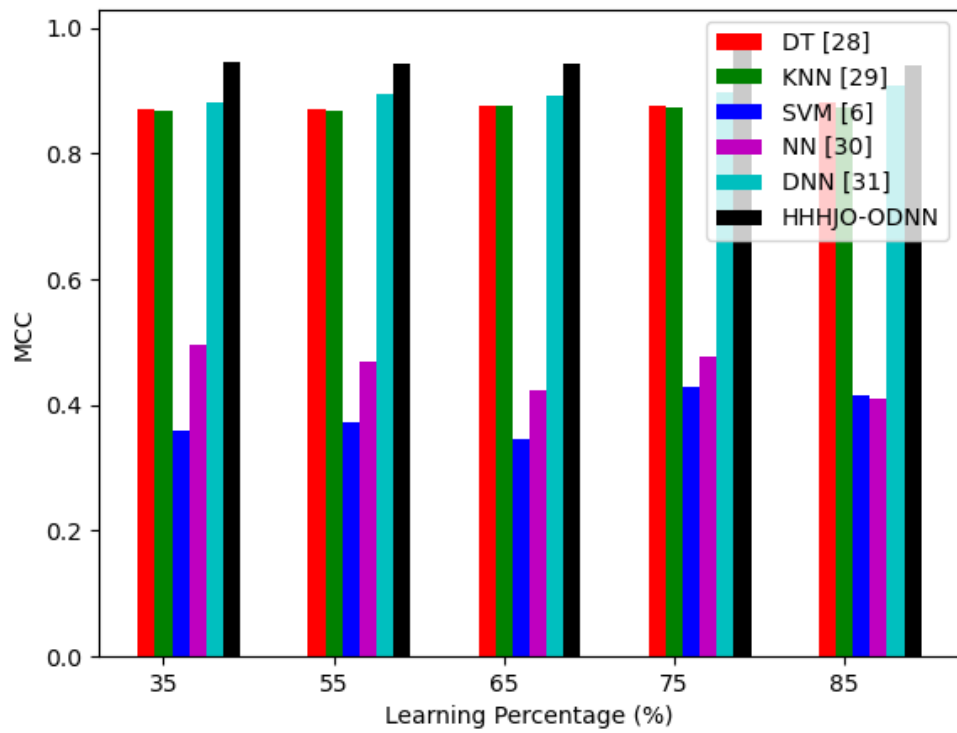
(c)



(d)

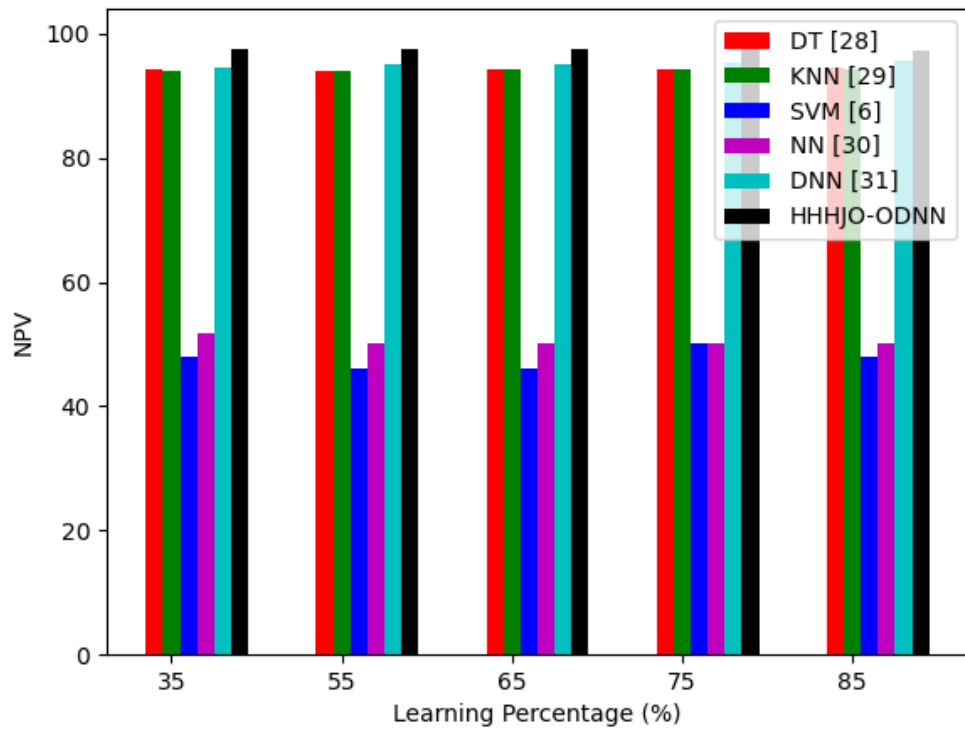


(e)

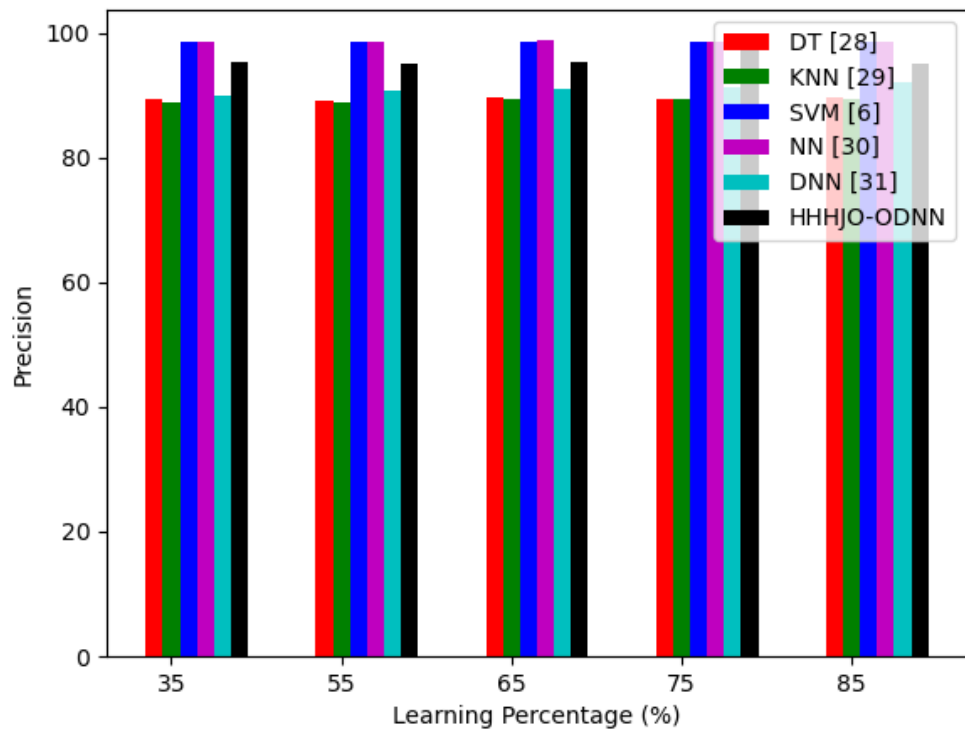


(f)

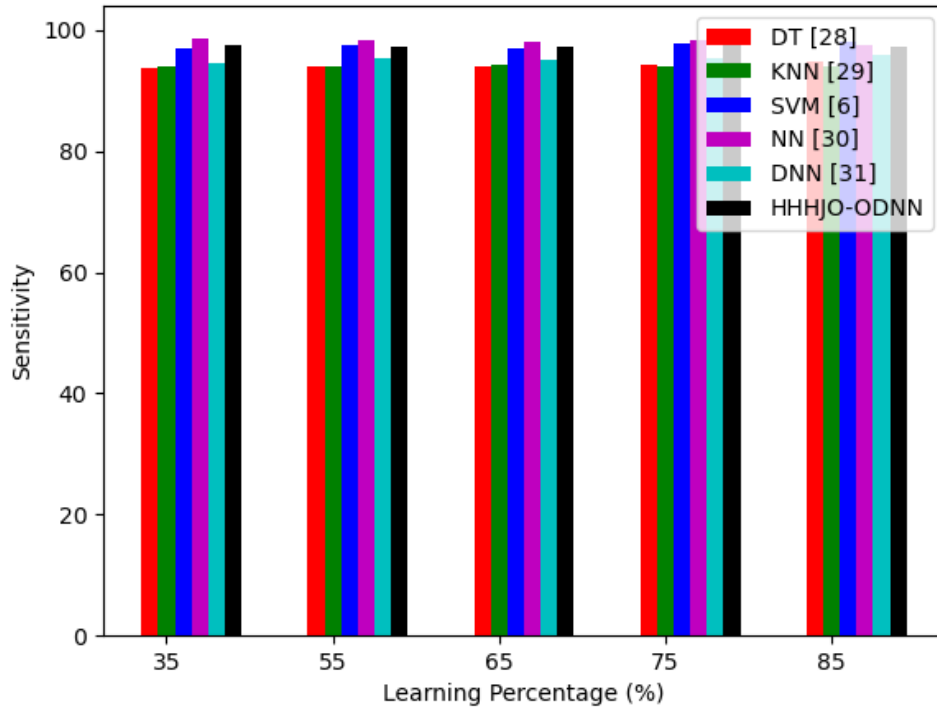




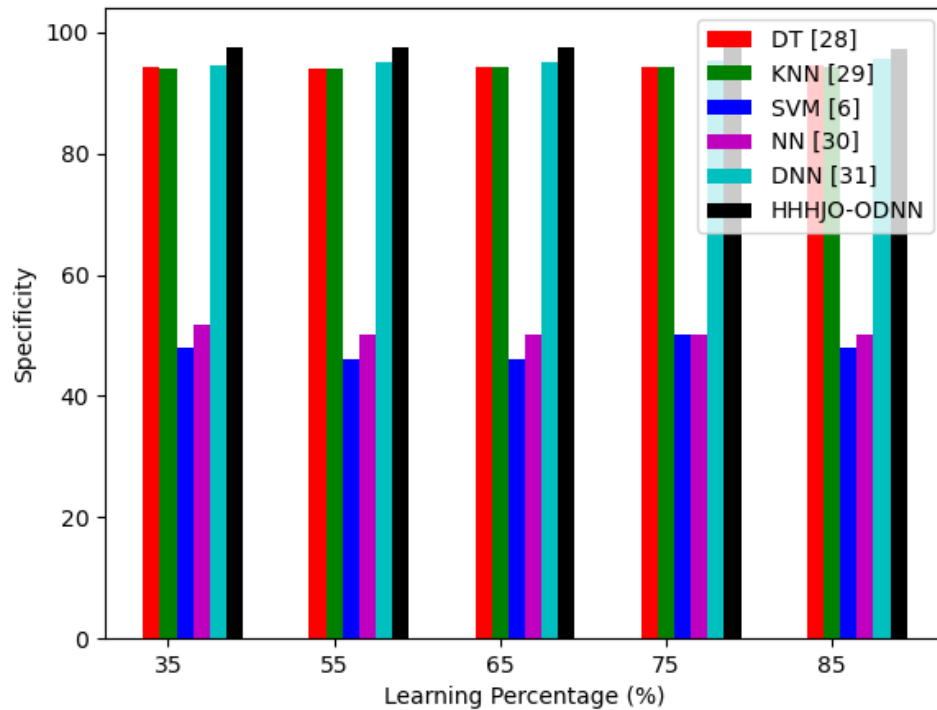
(g)



(h)



(i)



(j)

**Fig. 6.3:** Evaluation on proposed thermogram-based breast cancer detection model with multiple classifier over “(a) accuracy , (b) F1-score, (c ) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NVP, (h) precision, (i) sensitivity & (j) specificity”

## 6.4 Convergence Examination of Established Model with Diverse Optimization Algorithms

The convergence analysis of the suggested detection model is shown in Fig. 6.4 and compared against other heuristic approaches. It is evaluated for four different feature-sets and plotted as the value of the loss function against varying number of iterations. The convergence rate of recommended model attains as 2% more than TSA-ODNN 1% more than ITSA-ODNN and HHO-ODNN as well as 1.5% more than JAYA-ODNN, accordingly. Thus, the higher convergence rate ensures to enhance the detection performance for early diagnosing the breast cancer disorder. Higher convergence rates enhance detection rates by enabling faster, more accurate, and more efficient learning processes. This leads to better model performance, timely detection, and improved adaptability to changing data patterns and larger datasets. As a result, models with high convergence rates are more effective in identifying true positives while minimizing false positives, thereby ensuring robust and reliable detection in various applications.

Combining JAYA and HHO into a hybrid metaheuristic algorithm introduces additional computational overhead due to initialization, evaluation, iteration, update, and synchronization complexities. Whether it is worth using depends on the specific optimization problem, the performance gains achieved, and the available computational resources. Empirical testing on the target problem is crucial to determine the effectiveness and efficiency of the hybrid approach compared to individual algorithms.

Combining JAYA and Harris Hawk Optimization (HHO) into a hybrid metaheuristic can be worthwhile due to several potential benefits. These benefits can outweigh the computational overhead, especially in complex optimization problems where finding a high-quality solution is more important than computational speed alone. Here is detailed explanation why hybrid approach is valuable.

### 1. Complementary Strength:

**Jaya Algorithm** focuses on exploitation by iteratively moving solutions towards the best and away from the worst solutions in the population. This helps in fine-tuning the solutions and converging towards an optimum. **Harris Hawk** optimization balances exploration and exploitation dynamically through different phases (e.g., surprise pounce, soft besiege, hard besiege). HHO's adaptive behavior can escape local optima and explore the search space more thoroughly.

## **2. Improved Solution Quality:**

The hybrid algorithm leverages JAYA's strong local search capabilities and HHO's effective global search mechanisms. This combination increases the chances of finding a better global optimum by avoiding premature convergence and exploring more diverse areas of the search space.

## **3. Dynamic Adjustment:**

The dynamic adjustment between JAYA and HHO strategies allows the algorithm to switch focus based on the current state of the search process. For example, it can start with a broad exploration using HHO and gradually shift towards exploitation using JAYA as it converges towards potential solutions.

## **4. Increased Convergence Speed and Solution Accuracy:**

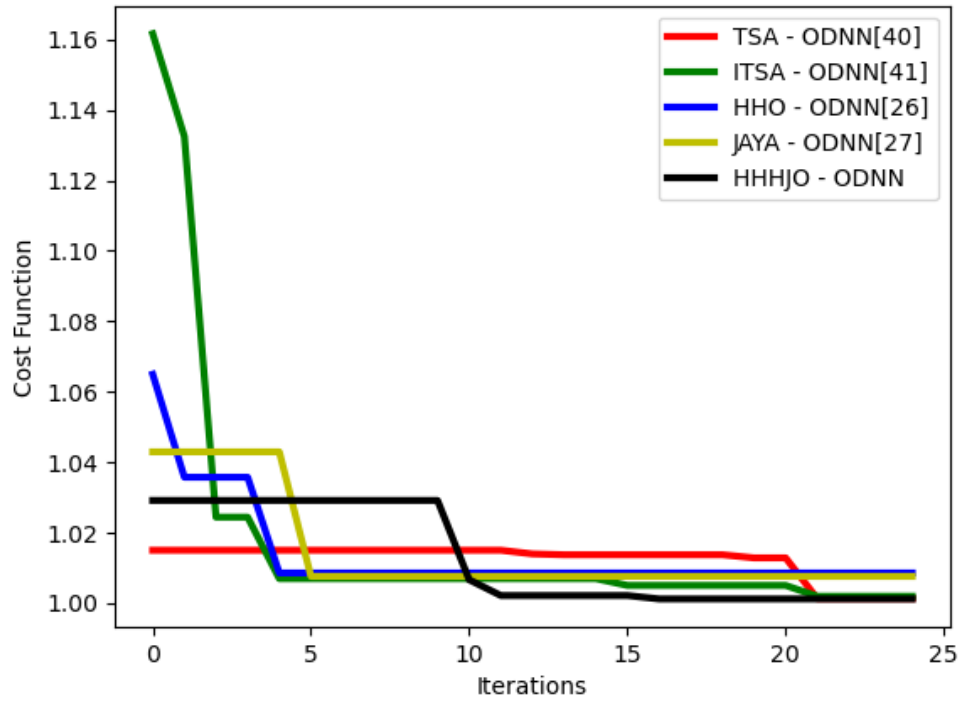
The complementary nature of JAYA and HHO can lead to faster convergence rates by effectively combining their strengths. This synergy can result in more efficient search processes, reducing the total number of iterations required to reach an optimal or near-optimal solution. Higher convergence speed does not only mean fewer iterations but also better utilization of each iteration, as both algorithms contribute to refining the solutions.

The hybrid approach can improve the precision and accuracy of the solutions. JAYA's focus on the best and worst solutions helps in fine-tuning, while HHO's adaptive strategies ensure a broad and effective search.

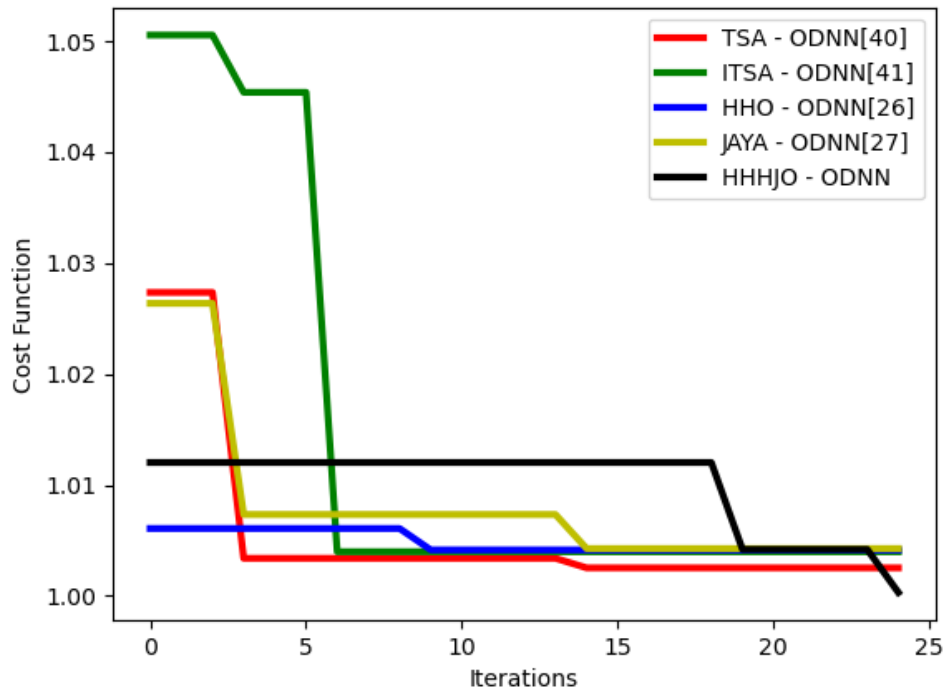
## **5. Adaptability:**

Complex problems often require adaptive methods to handle changing landscapes and dynamic constraints. The hybrid algorithm's ability to adjust its search strategy dynamically ensures better adaptability to such complexities.

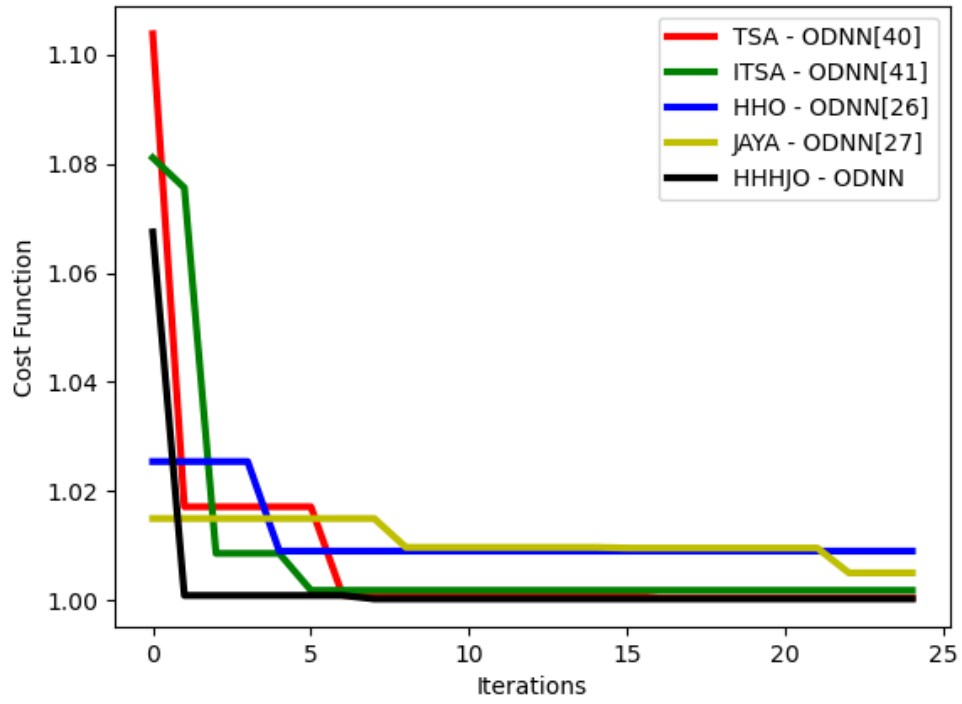
Using a hybrid JAYA and Harris Hawk Optimization algorithm can be highly beneficial due to enhanced exploration and exploitation, robustness across different problem types, increased convergence speed, and improved solution quality. While the computational overhead is higher, the potential gains in efficiency, accuracy, and adaptability can make this approach worthwhile for complex and large-scale optimization problems. Empirical testing and practical application evidence are essential to quantify these benefits and justify the additional computational costs.



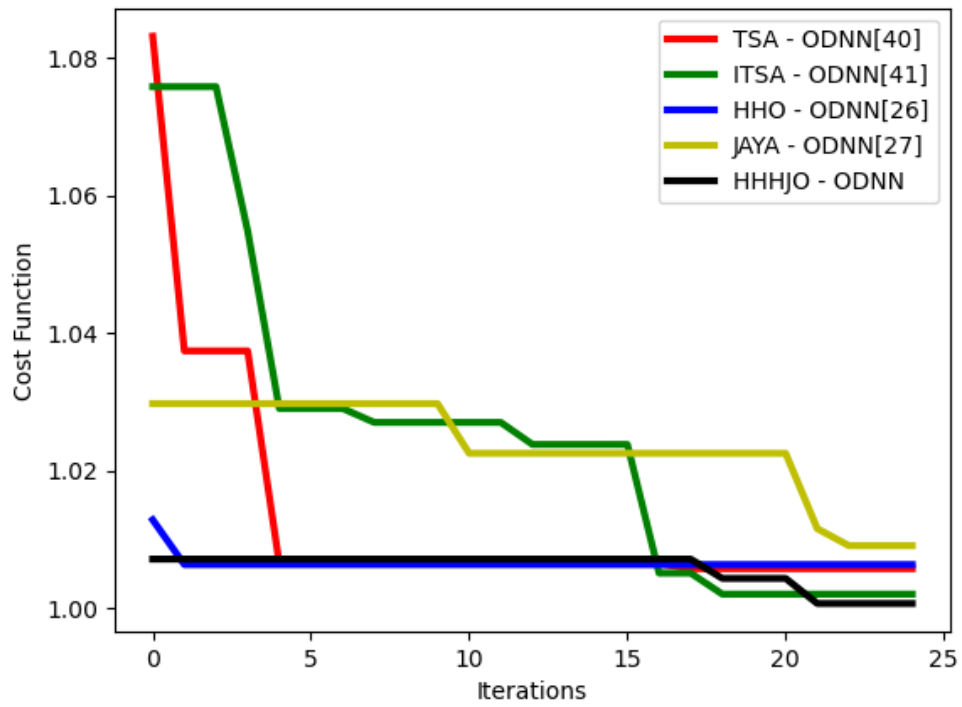
(A)



(B)



(C)



(D)

**Fig.6.4:** Convergence examination on thermogram-based breast cancer detection using proposed model with multiple classifier using (a) GLCM featuresF, (b) 1st and 2nd order features, (c ) LBP features and (d) Entropy features

The convergence plot in figure 6.4 have been achieved using DNN acted as a binary classifier with sigmoid as activation function and with the hyperparameters number of hidden layers, number of neurons in these hidden layers and learning rate as optimized using the specified optimization algorithm

### **6.5 Overall performance analysis of developed model with diverse classifier:**

All the results mentioned below are with the 75% of training and 25% of testing breast thermogram images. Similarly, the statistical and textural feature-based analysis on the developed thermogram-based breast cancer recognition model over the classifiers is presented in Table 6.2 The developed thermogram based breast cancer detection model statistical, texture and entropy feature analysis performed over the classifiers are represented in Table 6.3 Then, the analysis of statistical, texture, entropy and beta-entropy are displayed in Table 6.4. The overall efficiency analysis achieved over the developed thermogram breast cancer identification model with conventional approaches is given in Table 6.5.

**Table 6.1:** Statistical feature analysis with developed thermogram-based breast cancer detection model over classifier

Measures	SVM [6]	NN [30]	DNN [31]	HHHJO-ODNN
Accuracy	93.51%	93.82%	94.89%	95.19%
Sensitivity	93.48%	93.71%	95.06%	95.28%
Specificity	93.53%	93.87%	94.80%	95.14%
Precision	88.14%	88.72%	90.38%	90.99%
FPR	6.47	6.13	5.20	4.86
FNR	6.52	6.29	4.94	4.72
NPV	93.53%	93.87%	94.80%	95.14%
FDR	11.86	11.28	9.62	9.01
F1-Score	90.73%	91.15%	92.66%	93.08%
MCC	85.83	86.48	88.81	89.46

The statistical features are only considered for this analysis over the developed HHHJO-WO-DNN-based thermogram breast cancer detection model with conventional classifiers that are presented in Table 6.1 The F1-score analysis of the developed model achieved 2.5% better than SVM, 2.1% superior to NN and 4.5 enhanced than DNN. Similarly, the recommended model achieved enhanced accuracy rate than other classifiers. Thus, the developed thermogram-based breast cancer detection model HHHJO - ODNN attain effective detection rate than existing breast cancer detection model by considering the analysis only with the statistical features.



**Table 6.2:** Statistical and textural feature analysis executed on the established breast cancer detection model with already existing classifiers

Measures	SVM [6]	NN [30]	DNN [31]	HHHJO-ODNN
Accuracy	93.59%	93.97%	95.11%	95.73%
Sensitivity	93.48%	93.71%	95.28%	95.96%
Specificity	93.64%	94.10%	95.03%	95.61%
Precision	88.32%	89.10%	90.79%	91.83%
FPR	6.36	5.90	4.97	4.39
FNR	6.52	6.29	4.72	4.04
NPV	93.64%	94.10%	95.03%	95.61%
FDR	11.6	10.90	9.21	8.17
F1-Score	90.83%	91.35%	92.98%	93.85%
MCC	85.99	86.79	89.30	90.63

The statistical and textural feature-based analysis on the developed thermogram-based breast cancer detection model over the classifiers are presented in Table 6.2. The developed HHHJO-ODNN-based thermogram breast cancer recognition model achieved 2.6%, 2.3% and 0.2% superior than SVM, NN and DNN, respectively. Thus, the developed HHHJO-ODNN achieved effective breast cancer detection rate in individuals than the existing classifiers by considering the statistical and textural features.

**Table 6.3:** Statistical, texture and entropy feature analysis on developed thermogram breast cancer detection model over classifiers

Measures	SVM [6]	NN [30]	DNN [31]	HHHJO-ODNN
Accuracy	93.89%	94.12%	95.19%	96.03%
Sensitivity	93.93%	94.38%	95.06%	95.96%
Specificity	93.87%	93.99%	95.26%	96.07%
Precision	88.75%	88.98%	91.16%	92.62%
FPR	6.13	6.01	4.74	3.93
FNR	6.07	5.62	4.94	4.04
NPV	93.87%	93.99%	95.26%	96.07%
FDR	11.25	11.02	8.8	7.38
F1-Score	91.27%	91.60%	93.07%	94.26%
MCC	86.66	87.18	89.44	91.26

The developed thermogram-based breast cancer detection model statistical, texture and entropy feature analysis performed over the classifiers are represented in Table 6.3. The developed HHHJO-ODNN-based thermogram breast cancer detection model achieved efficient breast cancer detection rate than the existing approaches. The accuracy analysis performed over the developed HHHJO-WO-DNN model attained 2.14%, 1.91% and 0.84% enhanced performance rate than SVM, NN and DNN. So, the developed breast cancer detection model achieved effective detection rate than existing approaches by considering the statistical, textural and entropy features.

**Table 6.4:** Statistical, texture, entropy and beta-entropy analysis over the breast cancer identification model with several classifiers

Measures	SVM [6]	NN [30]	DNN [31]	HHHJO-ODNN
Accuracy	96.50%	97.10%	95.34%	99.08%
Sensitivity	97.84%	98.46%	95.28%	99.10%
Specificity	97.42%	97.24%	95.38%	99.08%
Precision	98.55%	98.56%	91.38%	98.22%
FPR	50.00	50.00	4.62	0.92
FNR	2.16	1.54	4.72	0.90
NPV	95.06%	95.18%	95.38%	99.08%
FDR	1.45	1.44	8.62	1.78
F1-Score	98.19%	98.51%	93.29%	98.66%
MCC	42.94	47.64	89.77	97.96

The developed thermogram breast cancer detection model is weighed up with classifiers for the analysis of statistical, texture, entropy and beta-entropy features are displayed in Table 6.4. The sensitivity analysis achieved over the developed HHHJO WO-DNN acquired enhanced performance rate 1.26%, 0.45% and 3.82% than SVM, NN and DNN, respectively. Thus, the developed breast cancer identification model achieved efficient recognition rate than existing approaches by considering the statistical, textural, entropy and beta entropy features.

**Table 6.5:** Overall performance analysis on developed breast cancer detection model with classifier approaches

Measures	DT [28]	KNN 29]	SVM [6]	NN [30]	DNN [31]	HHHJO-ODNN
Accuracy	94.27%	94.20%	96.50%	97.10%	95.34%	<b>99.08%</b>
Sensitivity	94.38%	94.16%	97.84%	98.46%	95.28%	<b>99.10%</b>
Specificity	94.22%	94.22%	50.00%	50.00%	95.38%	<b>99.08%</b>
Precision	89.36%	89.34%	98.55%	98.56	91.38%	<b>98.22%</b>
FPR	5.78	5.78	50.00%	50.00%	4.62	<b>0.92</b>
FNR	5.62	5.84	2.16	1.54	4.72	<b>0.90</b>
NPV	94.22%	94.22%	50.00%	50.00%	95.38%	<b>99.08%</b>
FDR	10.64	10.66	1.45	1.44	8.62	<b>1.78</b>
F1-Score	91.80%	91.68%	98.19%	98.51%	93.29%	<b>98.66%</b>
MCC	87.49%	87.31%	42.94%	47.64%	89.77%	<b>97.96%</b>

The overall analysis over the developed HHHJO-ODNN-based breast cancer detection model with several classifiers is displayed in table 6.5. The BC detection model performed 4.81% better than DT, 4.88% better than KNN, 2.58% better than SVM, 1.98% better than NN, and 3.74% better than DNN. Thus, the thermogram BC detection model classified individuals better than other models. Beta entropy can increase detection rates in breast thermographic images classification and medical diagnosis applications.

**Table 6.6:** Statistical feature analysis with developed thermogram-based breast cancer detection model over classifier

Measures	CNN	CNN with optimization	HHHJO-ODNN
Accuracy	94.64%	96.1%	99.08%
Sensitivity	96.27%	97.67%	99.10%
Specificity	48.61%	47.85%	98.21%
Precision	98.15%	99.28%	9.28%
FPR	5.13	5.21	9.2
FNR	3.72	2.32	8.9
NPV	48.61%	47.85%	99.07%
FDR	11.86	11.28	9.62
F1-Score	97.20%	97.97%	98.65%
MCC	36.57%	42.50%	97.97%

The developed thermogram breast cancer detection model is weighted up with different neural network classifier like CNN, CNN with optimization and HHHJO-ODNN shown in table 6.6. The overall accuracy of developed HHHJO-ODNN is better than the all classifiers which is 99.08%.

## 6.6 Summary

This experiment was to analyse the performance of the Beta entropy as features. It used all the classifiers in the test against consecutively adding the feature-sets and thereby measuring the performance improvement to decide the importance. The experimental results are tabulated in Table 6.1 using statistical features and 6.2 using statistical and texture features combined in 6.3 together using statistical, texture and entropy features combined together and table 6.4 using statistical, texture, entropy and Beta entropy features combined.

The table entries denote that in all the cases, the proposed HHHJO-ODNN classifier is giving the best performance with respect to all other classifiers and all the metrics used. Gradually with addition of every new set of features, the performance metrics are improving, but the maximum improvement came with the addition of Beta entropy. To take an example, the improvements in classification accuracy metric with addition of texture features over statistical features for HHHJO-ODNN is 0.54, further addition of entropy features improves the metric by 0.3 and finally the addition of Beta entropy feature improves the metric by 3.06. Similarly, the improvements in F1-score metric with addition of texture features over statistical features for HHHJO-ODNN is 0.77, further addition of entropy features improves the metric by 0.41 and finally the addition of Beta entropy feature improves the metric by 4.4. The overall efficiency analysis achieved over the developed thermogram breast cancer identification model with conventional approaches is given in Table 6.5. The accuracy analysis performed over the suggested breast cancer detection model acquired 4.81% better than DT, 4.88% enhanced than KNN, 2.58% improved than SVM, 1.98% higher than NN and 3.74% superior than DNN. Also, CNN with optimization gives accuracy 96.1% which is 3% less than developed HHHJO-ODNN. The result of comparison of all classifiers is shown in table 6.5 in which the HHHJO-ODNN gives the improved result over all classifiers. Thus, the developed model achieved effective early detection rate in individuals than existing models. The observation denotes that Beta entropy as features most significantly improved the classification performance and demonstrated that it can be successfully used for medical diagnosis and detection.

## CHAPTER 7

### CONCLUSION, FUTURE WORK AND DISCUSSION

#### 7.1 Conclusion

An optimization algorithm approach known as HHHJO has been presented as a means of enhancing the cancer detection. The HHHJO approach that has been developed incorporates a variety of techniques, such as a pre-processing technique known as the methodology, an optimization strategy. The breast thermal dataset serves as the source for the breast thermographic images used as input. The newly proposed thermographic breast cancer detection model with heuristic and deep learning approaches has offered an effective early detection rate in individuals. Using different thermographic images as the input, image pre-processing was performed with the help of greyscale conversion, adaptive mean filtering and contrast improvement. Then, the acquired pre-processed imageries were offered to the segmentation phase, where the GVF (Gradient Vector Flow) technique was utilized to segment the pre-processed image and exposed it to the abnormality segmentation phase. The abnormalities were segmented with the help of optimized FCM and parameter optimization was performed to tune the maximum iteration in fuzzy, epsilon and fuzziness parameter with the help of developed HHHJO. The fitness function to decide the parameters in optimized FCM is based on minimization of the variance and maximization of the entropy. Later, the abnormality segmented images were provided to the feature extraction phase and the features were acquired with the help of GLCM, 1st-order and 2nd-order textural descriptor, LBP, entropy feature and beta entropy. The parameter  $\beta$  in beta entropy allows for fine-tuning the entropy measure to better capture the characteristics of the data, making it a powerful tool for improving classification accuracy. Further, the concatenated features were offered to the classification stage that was performed with the help of ODNN and some parameter like learning rate of DNN, epoch's count of DNN and hidden neuron count of DNN was adjusted by utilizing developed HHHJO for get the best out of the accuracy to offer effective classification rate. The accuracy analysis achieved over the developed breast cancer detection model acquired 5.1% better than DT, 5.18% enhanced than KNN, 2.6% improved than SVM, 2.03% higher

than NN than DNN and 3.9% superior and 3% higher than CNN. Thus, the developed thermogram breast cancer detection model offered enhanced classification in the individuals than other conventional models.

## **7.2 Future Work**

In future research will be devoted to locate the affected regions accurately, and analyze the possibility of predicting breast cancer development in very early stages. The design of reliable and economical non-invasive methods of diagnosis and prevention is very important to increase the survival rates of women with breast cancer and lowering the costs of treatment in public healthcare systems. Further work is needed to produce breast thermogram datasets or make them publicly accessible for research, as only a few are easily accessible. Also, since the previous work was focused primarily on frontal breast thermograms, more work on the segmentation and classification of lateral breast thermograms is required to cater to lesions that may develop on lateral breast sides. In future the study would involve experimenting with datasets of huge sizes by combining different datasets and investigating various augmentation methods. The suggested models will be assessed in further work using various biomedical imaging datasets. Another future work direction would be exploring the use of other pre-trained DL models.

## **7.3 Discussion**

The suggested approaches are developed with the goal of enabling efficient breast cancer diagnosis while maintaining a high accuracy. A number of different screening methods are used in order to identify cancer in breast thermographic images with a lower false positive rate. As a result, it helps women lower their chance of developing breast cancer and improves the overall health of their breasts. Hence, effort may be done in the future to create a variety of screening tools in order to achieve earlier breast cancer identification. It is not possible to produce improved performance when the sizes of the breast thermograms are changed. During the process of identifying thermograms that include a tumour, however, it is impossible to differentiate between images that are quite similar. In addition, there are strategies for the extraction of characteristics that may be tailored to extract the features that are more pertinent to the identification of breast cancer detection as benign and malignant masses with a better



level of accuracy and precision. In a future improvement, the constraints that were mentioned with regard to the lowest detection time would be the primary emphasis.

Thermography, being less expensive and potentially more accessible, could and middle-income countries, access to conventional breast cancer screening technologies like mammography or MRI may be limited due to cost, infrastructure, or lack of trained fill an important gap in early cancer detection in these regions. With the rise of mobile health applications and telemedicine, thermographic data could be uploaded remotely to central databases for analysis by experts, improving accessibility to high-quality diagnostics in underserved regions. There is a renewed wave of interest in thermography because of the improvement of IR cameras sensors, though the development of such cameras still does not provide a more quantitative and robust procedure to detect the breast tumour. Thermography can have a significant impact on developing countries, where there is less availability of healthcare workers. Due to the low cost involved in thermography, communities with limited resources will benefit from providing the modality for early breast malignancy detection. Early cancer identification will reduce the burden on undeveloped communities with inadequate health infrastructure.

## REFERENCES

1. Retrieved from <https://www.who.int/health-topics/cancer>, In-Text Citation: "Cancer," 2023
2. "Cancer – Signs and symptoms". NHS Choices. Archived from the original on 8 June 2014. Retrieved 10 June 2014
3. Sánchez-Ruiz, D., Olmos-Pineda, I., & Olvera-López, J. A. (2020). Automatic region of interest segmentation for breast thermogram image classification. *Pattern Recognition Letters*, 135, 72–81. <https://doi.org/10.1016/j.patrec.2020.03.025>
4. Jayasekara, H., MacInnis, R. J., Room, R., & English, D. R. (2016). Long-Term Alcohol Consumption and Breast, Upper Aero-Digestive Tract and Colorectal Cancer Risk: A Systematic Review and Meta-Analysis. *Alcohol and Alcoholism*, 51(3), 315–330. <https://doi.org/10.1093/alcalc/agv110>
5. World Cancer Report 2014. World Health Organization. 2014. pp. Chapter 1.1. ISBN 978-92-832-0429-9. Archived from the original on 12 July 2017
6. Acharya, U. R., Ng, E. Y. K., Tan, J.-H., & Sree, S. V. (2012). Thermography Based Breast Cancer Detection Using Texture Features and Support Vector Machine. *Journal of Medical Systems*, 36(3), 1503–1510. <https://doi.org/10.1007/s10916-010-9611-z>
7. Moy, L., Elias, K., Patel, V., Lee, J., Babb, J. S., Toth, H. K., & Mercado, C. L. (2009). Is Breast MRI Helpful in the Evaluation of Inconclusive Mammographic Findings? *American Journal of Roentgenology*, 193(4), 986–993. <https://doi.org/10.2214/AJR.08.1229>
8. <https://seer.cancer.gov/statfacts/html/breast.html>. (2020).
9. (PDQ®)". NCI. 23 May 2014. Retrieved 29 June 2014. (n.d.). *Breast Cancer Treatment*.
10. Ali Salem Ali Bin, S. M. S. B. (2017). Breast Cancer Classification Enhancement Based on Entropy Method. *International Journal of Engineering and Applied Computer Science*, Volume: 02(8).

11. Mambou, S. J., Maresova, P., Krejcar, O., Selamat, A., & Kuca, K. (2018). Breast cancer detection using infrared thermal imaging and a deep learning model. *Sensors (Switzerland)*, 18(9). <https://doi.org/10.3390/s18092799>
12. Agarap, A. F. M. (2018). On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset. *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*, 5–9. <https://doi.org/10.1145/3184066.3184080>
13. Boyd, N. F., Guo, H., Martin, L. J., Sun, L., Stone, J., Fishell, E., Jong, R. A., Hislop, G., Chiarelli, A., Minkin, S., & Yaffe, M. J. (2007). Mammographic Density and the Risk and Detection of Breast Cancer. *New England Journal of Medicine*, 356(3), 227–236. <https://doi.org/10.1056/NEJMoa062790>
14. American Cancer Society. Archived from the original (PDF). (2007). *Cancer Facts & Figures 2007*.
15. National Cancer Institute. Archived from the original Retrieved. (2012). *"Obesity and Cancer Risk*.
16. Gage, M., Wattendorf, D., & Henry, L. R. (2012). Translational advances regarding hereditary breast cancer syndromes. *Journal of Surgical Oncology*, 105(5), 444–451. <https://doi.org/10.1002/jso.21856>
17. EtehadTavakol, M., Chandran, V., Ng, E. Y. K., & Kafieh, R. (2013). Breast cancer detection from thermal images using bispectral invariant features. *International Journal of Thermal Sciences*, 69, 21–36. <https://doi.org/10.1016/j.ijthermalsci.2013.03.001>
18. Krawczyk, B., Galar, M., Jeleń, Ł., & Herrera, F. (2016). Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy. *Applied Soft Computing*, 38, 714–726. <https://doi.org/10.1016/j.asoc.2015.08.060>
19. Majid, A., Ali, S., Iqbal, M., & Kausar, N. (2014). Prediction of human breast and colon cancers from imbalanced data using nearest neighbor and support vector machines. *Computer Methods and Programs in Biomedicine*, 113(3), 792–808. <https://doi.org/10.1016/j.cmpb.2014.01.001>

20. Thomas M. Cover and Joy A. Thomas. (n.d.). *Entropy and Information Theory* .
21. EtehadTavakol, M., Chandran, V., Ng, E. Y. K., & Kafieh, R. (2013). Breast cancer detection from thermal images using bispectral invariant features. *International Journal of Thermal Sciences*, 69, 21–36. <https://doi.org/10.1016/j.ijthermalsci.2013.03.001>
22. Chiang, T.-C., Huang, Y.-S., Chen, R.-T., Huang, C.-S., & Chang, R.-F. (2019). Tumor Detection in Automated Breast Ultrasound Using 3-D CNN and Prioritized Candidate Aggregation. *IEEE Transactions on Medical Imaging*, 38(1), 240–249. <https://doi.org/10.1109/TMI.2018.2860257>
23. Saniei, E., Setayeshi, S., Akbari, M. E., & Navid, M. (2016). Parameter estimation of breast tumour using dynamic neural network from thermal pattern. *Journal of Advanced Research*, 7(6), 1045–1055. <https://doi.org/10.1016/j.jare.2016.05.005>
24. Yasmin M, S. & M. (2013). Survey Paper on the diagnosis of breast cancer using image processing techniques. *Research Journal of Recent Science*, 2(10), 88–98.
25. Bezerra, L. A., Oliveira, M. M., Rolim, T. L., Conci, A., Santos, F. G. S., Lyra, P. R. M., & Lima, R. C. F. (2013). Estimation of breast tumor thermal properties using infrared images. *Signal Processing*, 93(10), 2851–2863. <https://doi.org/10.1016/j.sigpro.2012.06.002>
26. Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, 97, 849–872. <https://doi.org/10.1016/j.future.2019.02.028>
27. Venkata Rao, R. (2016). Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *International Journal of Industrial Engineering Computations*, 19–34. <https://doi.org/10.5267/j.ijiec.2015.8.004>

28. Hu, Q., Che, X., Zhang, L., Zhang, D., Guo, M., & Yu, D. (2012). Rank Entropy-Based Decision Trees for Monotonic Classification. *IEEE Transactions on Knowledge and Data Engineering*, 24(11), 2052–2064. <https://doi.org/10.1109/TKDE.2011.149>
29. Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2018). Efficient kNN Classification With Different Numbers of Nearest Neighbors. *IEEE Transactions on Neural Networks and Learning Systems*, 29(5), 1774–1785. <https://doi.org/10.1109/TNNLS.2017.2673241>
30. Liu, Y., Wang, F.-L., Chang, Y.-Q., & Li, C. (2011). A SNCCDBAGG-Based NN Ensemble Approach for Quality Prediction in Injection Molding Process. *IEEE Transactions on Automation Science and Engineering*, 8(2), 424–427. <https://doi.org/10.1109/TASE.2010.2077279>
31. Narayana Rao, K., Venkata Rao, K., & P.V.G.D., P. R. (2021). A hybrid Intrusion Detection System based on Sparse autoencoder and Deep Neural Network. *Computer Communications*, 180, 77–88. <https://doi.org/10.1016/j.comcom.2021.08.026>
32. Acharya, U. R., Ng, E. Y. K., Sree, S. V., Chua, C. K., & Chattopadhyay, S. (2014). Higher order spectra analysis of breast thermograms for the automated identification of breast cancer. *Expert Systems*, 31(1), 37–47. <https://doi.org/10.1111/j.1468-0394.2012.00654.x>
33. Carneiro, G., Nascimento, J., & Bradley, A. P. (2017). Automated Analysis of Unregistered Multi-View Mammograms With Deep Learning. *IEEE Transactions on Medical Imaging*, 36(11), 2355–2365. <https://doi.org/10.1109/TMI.2017.2751523>
34. Wang, H., Zheng, B., Yoon, S. W., & Ko, H. S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, 267(2), 687–699. <https://doi.org/10.1016/j.ejor.2017.12.001>
35. Kool, M., Bastiaannet, E., van de Velde, C. J. H., & Marang-van de Mheen, P. J. (2018). Reliability of Self-reported Treatment Data by Patients With Breast

- Cancer Compared With Medical Record Data. *Clinical Breast Cancer*, 18(3), 234–238. <https://doi.org/10.1016/j.clbc.2017.08.005>
36. Aghdam, N. S., Amin, M. M., Tavakol, M. E., & Ng, E. Y. K. (2013). Designing and Comparing Different Color Map Algorithms for Pseudo-Coloring Breast Thermograms. *Journal of Medical Imaging and Health Informatics*, 3(4), 487–493. <https://doi.org/10.1166/jmihi.2013.1191>
  37. Horlings, H. M., Weigelt, B., Anderson, E. M., Lambros, M. B., Mackay, A., Natrajan, R., Ng, C. K. Y., Geyer, F. C., van de Vijver, M. J., & Reis-Filho, J. S. (2013). Genomic profiling of histological special types of breast cancer. *Breast Cancer Research and Treatment*, 142(2), 257–269. <https://doi.org/10.1007/s10549-013-2740-6>
  38. Breast Cancer Stages 0, 1, 2, 3 and 4 | Memorial Sloan Kettering Cancer Center. (n.d.). <https://www.mskcc.org/cancer-care/types/breast/diagnosis/stages-breast>.
  39. Nover, A. B., Jagtap, S., Anjum, W., Yegingil, H., Shih, W. Y., Shih, W.-H., & Brooks, A. D. (2009). Modern Breast Cancer Detection: A Technological Review. *International Journal of Biomedical Imaging*, 2009, 1–14. <https://doi.org/10.1155/2009/902326>
  40. Kaur, S., Awasthi, L. K., Sangal, A. L., & Dhiman, G. (2020). Tunicate Swarm Algorithm: A new bio-inspired based metaheuristic paradigm for global optimization. *Engineering Applications of Artificial Intelligence*, 90, 103541. <https://doi.org/10.1016/j.engappai.2020.103541>
  41. Suryawanshi, S. P., & Dharmani, B. C. (2023). COMPARATIVE STUDY OF HEURISTIC-BASED SUPPORT VECTOR MACHINE AND NEURAL NETWORK FOR THERMOGRAM BREAST CANCER DETECTION WITH ENTROPY FEATURES. *Biomedical Engineering: Applications, Basis and Communications*, 35(02). <https://doi.org/10.4015/S1016237222500478>
  42. Abdel-Nasser, M., Moreno, A., & Puig, D. (2019). Breast Cancer Detection in Thermal Infrared Images Using Representation Learning and Texture Analysis Methods. *Electronics*, 8(1), 100. <https://doi.org/10.3390/electronics8010100>

43. Gomathi, P., Muniraj, C., & Periasamy, P. S. (2020). RETRACTED: Breast thermography based unsupervised anisotropic- feature transformation method for automatic breast cancer detection. *Microprocessors and Microsystems*, 77, 103137. <https://doi.org/10.1016/j.micpro.2020.103137>
44. Schaefer, G. (2014). ACO classification of thermogram symmetry features for breast cancer diagnosis. *Memetic Computing*, 6(3), 207–212. <https://doi.org/10.1007/s12293-014-0135-9>
45. Tello-Mijares, S., Woo, F., & Flores, F. (2019). Breast Cancer Identification via Thermography Image Segmentation with a Gradient Vector Flow and a Convolutional Neural Network. *Journal of Healthcare Engineering*, 2019, 1–13. <https://doi.org/10.1155/2019/9807619>
46. Mohamed, E. A., Rashed, E. A., Gaber, T., & Karam, O. (2022). Deep learning model for fully automated breast cancer detection system from thermograms. *PLOS ONE*, 17(1), e0262349. <https://doi.org/10.1371/journal.pone.0262349>
47. Saha, M., Chakraborty, C., & Racoceanu, D. (2018). Efficient deep learning model for mitosis detection using breast histopathology images. *Computerized Medical Imaging and Graphics*, 64, 29–40. <https://doi.org/10.1016/j.compmedimag.2017.12.001>.
48. Subrata Kumar Mandal. (2017). Performance Analysis of Data Mining Algorithms for Breast Cancer Cell Detection Using Naïve Bayes, Logistic Regression and Decision Tree. *International Journal of Engineering and Computer Science*, 6(2), 20388–20391.
49. Civilibal, S., Cevik, K. K., & Bozkurt, A. (2023). A deep learning approach for automatic detection, segmentation and classification of breast lesions from thermal images. *Expert Systems with Applications*, 212, 118774. <https://doi.org/10.1016/j.eswa.2022.118774>
50. AlFayez, F., El-Soud, M. W. A., & Gaber, T. (2020). Thermogram Breast Cancer Detection: A Comparative Study of Two Machine Learning Techniques. *Applied Sciences*, 10(2), 551. <https://doi.org/10.3390/app10020551>.

51. Ekici, S., & Jawzal, H. (2020). Breast cancer diagnosis using thermography and convolutional neural networks. *Medical Hypotheses*, 137, 109542. <https://doi.org/10.1016/j.mehy.2019.109542>
52. Yadav, S. S., & Jadhav, S. M. (2022). Thermal infrared imaging-based breast cancer diagnosis using machine learning techniques. *Multimedia Tools and Applications*, 81(10), 13139–13157. <https://doi.org/10.1007/s11042-020-09600-3>
53. Francis, S. v., Sasikala, M., & Saranya, S. (2014). Detection of Breast Abnormality from Thermograms Using Curvelet Transform Based Feature Extraction. *Journal of Medical Systems*, 38(4), 23. <https://doi.org/10.1007/s10916-014-0023-3>
54. National Cancer Institute. (2007). *Defining Cancer*.
55. Lee, H., & Chen, Y.-P. P. (2015). Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications*, 42(12), 5356–5365. <https://doi.org/10.1016/j.eswa.2015.02.005>
56. Sánchez-Ruiz, D., Olmos-Pineda, I., & Olvera-López, J. A. (2020). Automatic region of interest segmentation for breast thermogram image classification. *Pattern Recognition Letters*, 135, 72–81. <https://doi.org/10.1016/j.patrec.2020.03.025>
57. Katiyar, S. & V. A. (2013). A review over the applicability of image entropy in analyses of remote sensing datasets. *Photogrammetric Engineering and Remote Sensing*, 79.
58. Sun, W., Tseng, T.-L. (Bill), Zhang, J., & Qian, W. (2017). Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. *Computerized Medical Imaging and Graphics*, 57, 4–9. <https://doi.org/10.1016/j.compmedimag.2016.07.004>
59. Isikli Esener, I., Ergin, S., & Yuksel, T. (2017). A New Feature Ensemble with a Multistage Classification Scheme for Breast Cancer Diagnosis. *Journal of Healthcare Engineering*, 2017, 1–15. <https://doi.org/10.1155/2017/3895164>



60. Huang, M.-L., Hung, Y.-H., Lee, W.-M., Li, R. K., & Wang, T.-H. (2012). Usage of Case-Based Reasoning, Neural Network and Adaptive Neuro-Fuzzy Inference System Classification Techniques in Breast Cancer Dataset Classification Diagnosis. *Journal of Medical Systems*, 36(2), 407–414. <https://doi.org/10.1007/s10916-010-9485-0>
61. Chen, H.-L., Yang, B., Wang, G., Wang, S.-J., Liu, J., & Liu, D.-Y. (2012). Support Vector Machine Based Diagnostic System for Breast Cancer Using Swarm Intelligence. *Journal of Medical Systems*, 36(4), 2505–2519. <https://doi.org/10.1007/s10916-011-9723-0>
62. Chen, K.-H., Wang, K.-J., Tsai, M.-L., Wang, K.-M., Adrian, A. M., Cheng, W.-C., Yang, T.-S., Teng, N.-C., Tan, K.-P., & Chang, K.-S. (2014). Gene selection for cancer identification: a decision tree model empowered by particle swarm optimization algorithm. *BMC Bioinformatics*, 15(1), 49. <https://doi.org/10.1186/1471-2105-15-49>
63. Mehdy, M. M., Ng, P. Y., Shair, E. F., Saleh, N. I. M., & Gomes, C. (2017). Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer. *Computational and Mathematical Methods in Medicine*, 2017, 1–15. <https://doi.org/10.1155/2017/2610628>
64. Kharya, S., & Soni, S. (2016). Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection. *International Journal of Computer Applications*, 133(9), 32–37. <https://doi.org/10.5120/ijca2016908023>
65. Nicandro, C.-R., Efrén, M.-M., María Yaneli, A.-A., Enrique, M.-D.-C.-M., Héctor Gabriel, A.-M., Nancy, P.-C., Alejandro, G.-H., Guillermo de Jesús, H.-R., & Rocío Erandi, B.-M. (2013). Evaluation of the Diagnostic Power of Thermography in Breast Cancer Using Bayesian Network Classifiers. *Computational and Mathematical Methods in Medicine*, 2013, 1–10. <https://doi.org/10.1155/2013/264246>
66. Ebrahim Edriss Ebrahim Ali & Wu Zhi Feng. (2016). Breast Cancer Classification using Support Vector Machine and Neural Network. *International Journal of Science and Research (IJSR)*, 5(3), 1–6. <https://doi.org/10.21275/v5i3.NOV161719>

67. Carneiro, G., Nascimento, J., & Bradley, A. P. (2017). Automated Analysis of Unregistered Multi-View Mammograms With Deep Learning. *IEEE Transactions on Medical Imaging*, 36(11), 2355–2365. <https://doi.org/10.1109/TMI.2017.2751523>
68. Agarap, A. F. M. (2018). On breast cancer detection. *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*, 5–9. <https://doi.org/10.1145/3184066.3184080>
69. Khan, M. M., Mendes, A., & Chalup, S. K. (2018). Evolutionary Wavelet Neural Network ensembles for breast cancer and Parkinson's disease prediction. *PLOS ONE*, 13(2), e0192192. <https://doi.org/10.1371/journal.pone.0192192>
70. Nguyen, C., Wang, Y., & Nguyen, H. N. (2013). Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. *Journal of Biomedical Science and Engineering*, 06(05), 551–560. <https://doi.org/10.4236/jbise.2013.65070>
71. (<https://visual.ic.uff.br/dmi/prontuario/home>).
72. Shofwatul 'Uyun, S. H. A. H. and S. (2013). Selection Mammogram Texture Descriptors Based on Statistics Properties Backpropagation Structure. *Computer Vision and Pattern Recognition*.
73. Mammoottil, M. J., Kulangara, L. J., Cherian, A. S., Mohandas, P., Hasikin, K., & Mahmud, M. (2022). Detection of Breast Cancer from Five-View Thermal Images Using Convolutional Neural Networks. *Journal of Healthcare Engineering*, 2022, 1–15. <https://doi.org/10.1155/2022/4295221>
74. Chen, J., Yang, C., Xu, G., & Ning, L. (2018). Image Segmentation Method Using Fuzzy C Mean Clustering Based on Multi-Objective Optimization. *Journal of Physics: Conference Series*, 1004, 012035. <https://doi.org/10.1088/1742-6596/1004/1/012035>
75. Shofwatul 'Uyun, S. H. A. H. and S. ". (2013). Selection Mammogram Texture Descriptors Based on Statistics Properties Backpropagation Structure. *Computer Vision and Pattern Recognition*.

76. Narayana Rao, K., Venkata Rao, K., & P.V.G.D., P. R. (2021). A hybrid Intrusion Detection System based on Sparse autoencoder and Deep Neural Network. *Computer Communications*, 180, 77–88. <https://doi.org/10.1016/j.comcom.2021.08.026>
77. Nagalakshmi, T. (2022). Breast Cancer Semantic Segmentation for Accurate Breast Cancer Detection with an Ensemble Deep Neural Network. *Neural Processing Letters*, 54(6), 5185–5198. <https://doi.org/10.1007/s11063-022-10856-z>
78. TAN, T., QUEK, C., NG, G., & NG, E. (2007). A novel cognitive interpretation of breast cancer thermography with complementary learning fuzzy neural memory structure. *Expert Systems with Applications*, 33(3), 652–666. <https://doi.org/10.1016/j.eswa.2006.06.012>
79. Conforte, A. J., Tuszynski, J. A., Silva, F. A. B. da, & Carels, N. (2019). Signaling Complexity Measured by Shannon Entropy and Its Application in Personalized Medicine. *Frontiers in Genetics*, 10. <https://doi.org/10.3389/fgene.2019.00930>
80. Cichocki, A., Cruces, S., & Amari, S. (2011). Generalized Alpha-Beta Divergences and Their Application to Robust Nonnegative Matrix Factorization. *Entropy*, 13(1), 134–170. <https://doi.org/10.3390/e13010134>
81. Shofwatul 'Uyun, S. H. A. H. and S. (2013). Selection Mammogram Texture Descriptors Based on Statistics Properties Backpropagation Structure. *Computer Vision and Pattern Recognition*.
82. Krawczyk, B., & Schaefer, G. (2014). A hybrid classifier committee for analysing asymmetry features in breast thermograms. *Applied Soft Computing*, 20, 112–118. <https://doi.org/10.1016/j.asoc.2013.11.011>
83. Francis, S. v, Sasikala, M., Bhavani Bharathi, G., & Jaipurkar, S. D. (2014). Breast cancer detection in rotational thermography images using texture features. *Infrared Physics & Technology*, 67, 490–496. <https://doi.org/10.1016/j.infrared.2014.08.019>

84. A. Rényi, On measures of entropy and information. (1961). *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 547-561.
85. Jiang, X., Marti, C., Irniger, C., & Bunke, H. (2006). Distance Measures for Image Segmentation Evaluation. *EURASIP Journal on Advances in Signal Processing*, 2006(1), 035909. <https://doi.org/10.1155/ASP/2006/35909>
86. Neemuchwala, H., Hero, A., & Carson, P. (2005). Image matching using alpha-entropy measures and entropic graphs. *Signal Processing*, 85(2), 277–296. <https://doi.org/10.1016/j.sigpro>.
87. Mordang, J. J., Gubern-Mérida, A., Bria, A., Tortorella, F., Mann, R. M., Broeders, M. J. M., den Heeten, G. J., & Karssemeijer, N. (2018). The importance of early detection of calcifications associated with breast cancer in screening. *Breast Cancer Research and Treatment*, 167(2), 451–458. <https://doi.org/10.1007/s10549-017-4527-7>

## ANNEUXRE

### LIST OF PUBLICATIONS

Sr. No	Name of Journal / Conference	Name of the Paper
1	Biomedical Engineering - Applications, Basis and Communications	Comparative Study of Heuristic-based Support Vector Machine and Neural Network for Thermogram Breast Cancer Detection with Entropy Features
2	Virtual Symposium on applications of machine learning data science in inter disciplinary areas	Gradient Vector flow for automatic segmentation: A step Ahead Towards early breast cancer detection through thermography
3	International conference on Intelligent System and Application (2023)	Heuristic neural network for the thermography breast cancer detection
4	Review paper published online in Think India Journal	“Breast cancer detection through thermography: A review”