International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

Analysis and Design of Detection for Liver Cancer using Particle Swarm Optimization and Decision Tree

Seema Kaloria, S.S. Saini
M. Tech Scholar, Department of CSE, Jaipur Engineering College, Kukas, Jaipur
Associate Professor, Department of CSE, Jaipur Engineering College, Kukas, Jaipur
Email: paliwalseema17@gmail.com*

Abstract

Liver cancer is taken as a major cause of death all over the world. According to WHO (World Health Organization) every year 9.6 million peoples are died due to cancer worldwide. It is one of the eighth most leading causes of death in women and fifth in men as reported by the American Cancer Society. The number of death rate due to cancer is projected to increase by45 percent in between 2008 to 2030. The most common cancers are lung, breast, and liver, colorectal. Approximately 7, 82,000 peoples are died due to liver cancer each year. The most efficient way to decrease the death rate cause of liver cancer is to treat the diseases in the initial stage. Early treatment depends upon the early diagnosis, which depends on reliable diagnosis methods. CT imaging is one of the most common and important technique and it acts as an imaging tool for evaluating the patients with intuition of liver cancer. The diagnosis of liver cancer has historically been made manually by a skilled radiologist, who relied on their expertise and personal judgement to reach a conclusion. The main objective of this paper is to develop the automatic methods based on machine learning approach for accurate detection of liver cancer in order to help radiologists in the clinical practice. The paper primary contribution to the process of liver cancer lesion classification and automatic detection for clinical diagnosis. For the purpose of detecting liver cancer lesions, the best approaches based on PSO and DPSO have been given. With the help of the C4.5 decision tree classifier, wavelet-based statistical and morphological features were retrieved and categorised.

Keywords: PSO, DPSO, C4.5 Decision Tree Clssifier, Liver Cancer**.**

I. INTRODUCTION

The human liver is one of the most complex and largest internal organs in the body. It is situated in the right topmost quadrant of the abdomen and also it is linked with two blood vessels named as Portal vein and Hepatic artery [1, 2]. The Portal vein blood vessel supplies blood to the liver from gastro intestinal tract and spleen, pancreas whereas Hepatic artery carries blood from the aorta. Generally, the liver structure has consisted of two types of lobes, i.e., right lobe and left lobe. Mostly two major kind of cells known as parenchyma and non-parenchyma cells are occupied 80% of the liver volume.

The most efficient way to decrease the death rate cause of liver cancer is to treat the diseases in the initial stage. Early treatment depends upon the early diagnosis, which depends on reliable diagnosis methods. CT imaging is one of the most common and important technique and it acts as an imaging tool for evaluating the patients with intuition of liver cancer. The diagnosis of liver cancer has historically been made manually by a skilled radiologist, who relied on their expertise and personal judgement to reach a conclusion. Even experts could have a higher incidence of inter-observer variation. Although there have been numerous studies and reviews, many researchers have not reached a consensus regarding the CT scan method and diagnostic procedure that best characterise malignant lesions. As a result, several automated procedures are needed for the readings and interpretations of the CT scans. The radiologists identify and spot the image patterns to diagnosis according to those patterns. The interpretation of expert radiologists is a natural process because it is primarily depends upon the human visual perception which is error prone. As the data set of CT scan may contain more than 600 slices, therefore the manual process is time consuming and error prone. So we have taken interest to develop automated detection systems for identifying the liver cancer. CAD based decision making systems help the radiologists in improving the diagnostic accuracy, reduce their workloads and burden.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

OBJECTIVES OF RESEARCH

The main contribution of this thesis is to develop an automated CAD methodology for the detection and classification of liver cancer from CT scans. The classification of cancerous tissue of liver using CT scan is very essential procedure for diagnosing liver cancer because it helps in early detection of the diseases in order to save lives. The automated decision system also reduces the human intervention. To accomplish automated approach, following important process has been adopted: image acquisition, preprocessing, image segmentation, feature extraction, feature reduction and classification of liver cancer type.

The following specific objectives are implemented that combines different methodology for finding the solutions of the detection problems.

- Implementation of level set method for segmentation and classification of liver cancer using random forest in CT images.
- To build an adaptive fuzzy clustering based texture analysis for classifying liver cancer using neural network classifiers in CT images.
- To introduced a technique for the detection and characterization of liver cancer using modified fuzzy clustering and decision tree classifier using CT images.

II. LITERTAURE REVIEW

An innovative technique for diagnosing liver illness is presented by Hassoon, Kouhi, Moghadam, and Abdar (2017) [1]. This method uses the Genetic algorithm to increase the efficiency and accuracy of the rules obtained from the Boosted C5.0 classification algorithm. Boosted C5.0 generates 92 rules, but GA only generates 24 rules with a 12% increase in accuracy.

Moloud Abdar et al. (2017) [2] For the ILPD dataset, contrast the results of two novel classification algorithms, Boosted C5.0 and CHAID. The outcome reveals that Boosted C5.0's accuracy is 93.75%, which is higher than CHAID's accuracy of 65.00%. The consequence of including gender in the boosted C5.0 algorithm's prediction of liver disease is that females are more susceptible than males.

Some of the noteworthy contributions have been summarized in table 2.1.

Table 2.1

Brief summary of literature based on machine learning

S.No.	Author	Year	Techniques used	Finding
1	M. Pant et al.[35]	2019	Classification	Papers reviewed from 2014 to 2019 based on Brain Tumor Segmentation and Classification
2	B. Chandra et al. [36]	2006	Decision Tree(DT)	Work for reducing the choice of the split
3	Lipo Wang et al. [37]	2019	convolutional neural network	Brain histology classification using ML
4	N.P. Gopalan et al.[38]	2014	Fuzzy rule based classifier	ANFIS system is proposed for erythemato-squamous disease analysis
5	AK. Dwivedi [39]	2018	Classification	Heart disease prediction
6	Youn et al. [40]	2018	Ensemble Learning	Nearly noise-free(NNF) model is proposed for elimination of mislabeled training data
7	Nahato et al. [41]	2016	Classification	Proposed fuzzy sets and

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

				extreme learning machine based model for disease prediction
8	Sokolova et al. [42]	2009	Measure Invariance	Done analysis of performance measures for classification tasks
9	Verma et al.[43]	2019	Ensemble Learning	Skin disease prediction
10	Liu et al.[44]	2012	Fuzzy KNN	Thyroid disease prediction

III. PROPOSED METHODOLOGY

It can be difficult to find liver cancer utilising Computed Tomography (CT) Scan images in clinical settings. For segmenting the liver lesions affected by cancer utilising CT images, we have suggested an unique method in this research that is based on optimised techniques such as particle swarm optimization (PSO) and Darwinian particle swarm optimization (DPSO) algorithms. A specific slice from a set of 112 real-time photos with malignant afflicted lesions obtained from various people was used to evaluate the system. To categorise four different types of malignant cells in the liver, the proposed method involves pre-processing, segmentation, feature extraction, and classification stages. Using optimal procedures, images were first segmented, and from there, a variety of statistical and morphological information were extracted. Using a decision tree (C4.5) classifier, the feature set was subsequently divided into four categories: cyst, hemangioma, hepatocellular carcinoma (HCC), and metastatic carcinoma (MET). With a classification accuracy of 98.26%, the DSPO procedure efficiently segregated the lesion, outperforming the PSO method's accuracy of 92.01%. Similarly, the DSPO approach was found to segment the cancer-affected lesions with improved results for all other assessed parameter values. Hence, these findings demonstrate that the method based on the DSPO technique with the C4.5 classifier is a successful means of detecting liver cancer in CT scans.

In this study, we proposed an automatic detection and classification of liver cancer in CT scan images. The flow diagram of proposed approach is represented in Fig.3.1. A total 112 number of cancer affected images were considered for evaluation that consists of hemangioma, HCC and metastatic carcinoma. In preprocessing stage, a moving average filter is used for smoothing images to reduce the intensity variation between neighboring pixels for better classification accuracy. Then adaptive thresholding was applied on the CT images to isolate the liver from other organ of the body. The separated image is further processed by PSO and DPSO approaches todelineate the affected lesions in the liver. Wavelet transform was applied in further process in the segmented image to eliminate noise. The segmented images were decomposed into different components depending on the frequency characteristics of the image. It converts the images with maximum information components into approximation, horizontal, vertical and diagonal parts of the image. The LL sub-band is the taken as approximation coefficient of the original image which is consider for extraction of image information. Various statistical and morphological features were extracted from the approximation coefficients at level three of wavelet decompositions. The feature set is then classified using C4.5 classifier to classify different types of liver cancers.

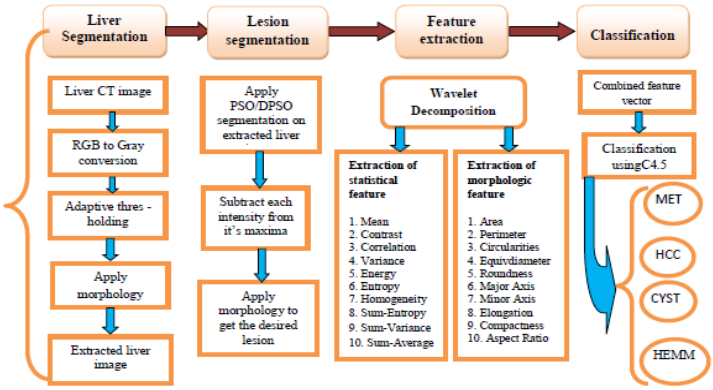

Figure 3.1: Flow diagram of the proposed approach in delineating cancer lesions in CT scan of liver

3.2.1 Adaptive Thresholding

In adaptive thresholding the threshold value at each pixel is determined using the neighboring pisel intensity unlike a fixed threshold. To compute the threshold $T(x,y)$ in

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

adaptive thresholling process the following steps were performed.

1 $Ab \times b$ region is selected within the image, where b is decided by the user. We have considered b 300 , as it give better segmented result.

2 In the nest step a weighted average or mean is calculated in the $b \times b$ region. The pivel values near to the center of the box will have higher weight representing by $WA(x,y)$.

3 Finally a constant parameter named as 'th' is subtracted from the weighted. value $WA(x,y)$ that is calculated for each pixel described in the previous step. Here "th" is defined as the mean of the image.

4 So the adaptive threshold $T(x,y)$ at each pivel position (x,y) is conmuted using the formula presented in Eq. (3.1)

$$T(x,y) = WA(x,y) - t \tag{3.1}$$

3.2.2 Particle Swarm Optimization (PSO)

The popular population-based search technique known as Particle Swarm Optimization (PSO) was created by Eberhart and Kennedy. In PSO, the fitness function—defined as the class variance _102 of the image—evaluates the particles. The number of intensity levels L will determine the search space; for example, if the frames are 8-bit pictures, the particles will be deployed between 0 and 255. It is fundamentally founded on the idea of swarm intelligence, in which the cooperative character of the advanced agents is intercommunicated locally with their environment. We discovered the optimal answer (local best) by dispersing information with nearby particles over the search space and computing the nearby hard best. The optimum solution, according to the image segmentation method, maximises the between class variance of the dispersed intensity levels inside the image. This is done by using pixels sets that match to the search space.

The PSO Algorithm

Initialize Swarm (Initialize $x_t^{\pi}, v_t^n, \overline{x_t^{\pi}}, \overline{n_e^n}$ and $\overline{g_t^n}$)
Loop:
For all partic les
Evaluate the fitness ϕ^f of each particle
Update $\overline{x_t^n}, \overline{x_t^{\pi}}$ and $\overline{x_t^{\pi}}$
Update v_t^n, x_t^n
End
until teruination (convergence)

The global best solution produced by the entire swarm is updated during each stage of processing. These findings suggest that particles perceive the location where success was obtained and are guided by success. A success function in each step evaluates the effectiveness of each algorithmic particle.

To design the Swamn. At time ' t ' each single particle moves ' n ' in a multidirectional search space with a position (x_t^n) and a velocity (v_t^n) which is depend on local best position $(\widetilde{x_t^t})$, neighbothood best (n_t^1) and global best $(\overline{g_t^*})$ data as presented in the following two Equations, Eq. (3.2) and Eq. (3.3)

$$v_{t+1}^n = wv_t^n + \rho_1 v_1(g_t^n - x_t^n) + \rho_2 r_2(\bar{x}_t^n - x_t^n) + \rho_2 v_2(\widetilde{n_t^n} - x_t^n) \tag{3.2}$$
$$x_{t+1}^n = x_t^n + v_{t+1}^n \tag{3.3}$$

Where the co-efficient $\omega_1 \rho_1, \rho_2$ and ρ_a are the weights of the inertial influence, the global best, the local best and the neighbothood best respectively. The "cognitive" and "social" conponents are represented by some integer, known as ρ_1, ρ_2 and ρ_2. Also the mensures r_1, r_2, and r_2 are represents some random numbers where each single component nomally a uniform random mumbers in-between 0 and 1. In PSO the particles are evahated by the fitness function, which is specified as between class variance σ_i^2 of the image. Initially, the velocities of the particles are consider as zero and the position is arbitrarily taken from the boundaries of the search space.

Where the co-efficient ω, ρ_1, ρ_2 and ρ_3 are the weights of the inertial influence, the global best, the local best and the neighborhood best respectively. The "cognitive" and "social" components are represented by some integer, known as ρ_1, ρ_2 and ρ_3. Also the measures r_1, r_2, and r_3 are represents some random numbers where each single component normally a uniform random numbers in-between 0 and 1. In PSO the particles are evaluated by the fitness function, which is specified as between class variance σ_B^2 of the image. Initially, the velocities of the particles are consider as zero and the position is arbitrarily taken from the boundaries of the search space.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

According to the nature of problem, the global bests and local neighborhood are initialized with the worst possible values. The other parameter like population size and stopping criteria are also needed adjustment. To get a good solution, it is important to optimize the population size within an acceptable time limit. According to the problem type the convergence can be defined the number of iterations for obtaining superior results. PSO revels an outcome of implicit communication between parties by upgrading the neighborhood and also the global information, Hence PSO algorithm has been used effectively in various applications like robotics, electric systems and sports engineering etc.

3.2.3 Darwinian PSO (DPSO)

The optimization algorithm's drawback is that it can get stuck at the local optimum. Despite the fact that a way may be designed to change the fitness function while also modifying the fitness environment, the result would be a new powerful approach with a wider range of applications. A natural selection-based technique is developed, wherein when the particles identify local optima in the search area, the finding in that space is completely deleted and a new space is introduced as a replacement. Thus, this form of search is planned as part of the trial.

In the execution, PSO utilizes a distinct swarm of test solution. To apply natural selection with a particular swarm, the method must observe the lack of development of PSO algorithm. It is quite difficult to model natural selection with a distinct swarm because it is unfit to distinguish between local and global optimum. A selection process in which after each swarm updates the present fitness of the particles is considered to arrange the particles. From the starting, one half of the particles is correlated and substitutes the position and velocity of the particles of the bottom half. The particle does not change its personal best value. So it archives better convergence applying the algorithm. Darwinian PSO is a superior approach of natural selection, where at any time multiple swarms of the test solution may exist [116]. Each swarm individually behaves like a traditional PSO but some rules are collectively governed by the swarm to stimulate the natural selection. The DPSO algorithm was implemented using the following details:

A. Initialization of Particle and Swarm

Each particle in the PSO algorithm containing an array with ' N ' numbers, suppose $\bar{x}_1$ is a particle of the swarm and the dimension of individual particle is arbitrarily computed within the range $\overline{x_{min}} \leq \bar{x}_1 \leq \overline{x_{max}}$. Likely the velocities are arbitrarily computed within the range $\overline{v_{min}} \leq \bar{v}_1 \leq \overline{v_{max}}$, that permit particles to move a remarkable fraction within the scope of $\bar{x}_1$ in one iteration while moving at $|\bar{v}_1| = v_{max}$. So in this algorithm the swarm initialization is done using a particle population. The pseudo code of main program and evolve swarm algorithm is presented below. Every single swarm is evolved using the fitness function of all of the particles. Then the individual best positions and the neighborhood of every single particle are updated. When a new global best fitness is found at that time the swarm spawns a new particle. A particle is deleted, if the swarm is unable to find more fit state.

B. Condition for Deleting a Swarm

Within a swarm's population, m is within the range of, $m_{min} \leq m \leq m_{max}$. So when the population of swarm decline below the range of m_{min}, the swarm is deleted.

C. Condition for Deleting a Particle

The particle with worst performance is deleted; when the swarms search counter goes beyond SC_c^{max} (maximum critical threshold range). At the time of swarm creation the value of search counter is set to zero. Likewise when the search counter arrives to a value near aboutSC_c^{max}, the particle is deleted during which the swarms do not formulate any improvement in the fitness. If N_{kill} represents the total amount of particles destroyed within the swarm across a period when the improvement in fitness is no more, hence the reset value of the search counter is selected as in Eq. (3.4)

$$SC_c(N_{kill}) = SC_c^{max}\left[1 - \frac{1}{N_{kill}+1}\right] \quad (3.4)$$

The Algorithm

Main Program Loop (step 1)
Evolve Swarm Algorithm
For every single swarm in the collection
For every single particle in the swarm
Evolve the swarm (right)
Update Particle Fitness
For every single swarm in the collection
For every single particle in the swarm
Allow the swann to spawn
Update Particle Bests

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

Delete 'failed' swarms.
For every single particle in the swarm
Move Particle
If the swarm gets better
Reward swarm: spawn particle: extend
The swarm life
If the swarm has not improved
Punish swarm: possibly delete particle:
Reducetheswarm life

D. Condition for Spawning Particles and Swarms

A new swarm is generated, if the maximum swarm number will not be reached and the existing swarm must satisfy the condition $N_{kill} = 0$. The probability of spawning a new swarm is $p = f/N_S$, here f represents a uniform random number within the range of $[0,1]$ and N_S represents the total number of swarms. The above factor of $1/N_S$ is used to control the creation of the swarm, when several number of swarms are presented. The parent swarm is unaffected, when it spawns a new swarm. A child swarm is created, by considering half of the particles from the parent swarm and the rest half of the particles from any random number. So in DPSO algorithm, a new particle is spawned at any time when a swarm scores a new global best fitness function.

3.2.4 Wavelet Decomposition

One efficient method for extracting features from an image to handle the classification issue is the wavelet transform. Due to the DWT's multiresolution capacity and the fact that it results in a sparser signal representation than the DCT, the DWT-based feature extraction was more effective than the DCT at classifying normal and abnormal images. Depending on the frequency properties of the data, it divides the image into various components. Using some coefficients that are separated into the horizontal, vertical, and diagonal regions of the image, it converts the data into the frequency domain. In this transform, the initial image is transformed into four components known as sub-bands named as LL, HL, LH, and HH as presented in the Fig. 3.2. The LL sub-band represents the approximation or average coefficients of the initial image. The other three sub-bands represents the details components namely the vertical details, diagonal details and horizontal details respectively. Transform resolve the signal or image at distinct scale or resolution. Wavelet transform is extremely beneficial for segmentation because it provides time-frequency response of a signal. A variety of wavelet families are Haar, Daubechies, Biorthogonal, Symlet, Meyer are commonly used. The mother wavelets are scaled and shifted to produce the higher order wavelets.

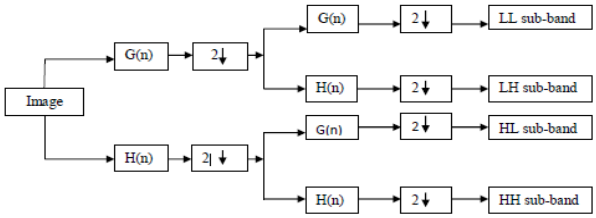

Figure 3.2: Block diagram of 2D DWT

3.2.5 Feature Extraction

Broadly speaking, feature extraction refers to processes that take the informative values out of each image pixel and provide a data set for categorising diseases. Analysis of texture is a key component of machine learning methods used in image processing.

Statistical Features

Morphologic features

3.2.6 Feature Selection Ranking

The filter method selects features independently without using learning algorithm. It selects some important features of the complete training data. Initially the algorithm ranks the features and defined a threshold to select the most significant features from the ranking. In this work the information measures based filter ranking method is used.

3.2.7 C4.5 Decision Tree Classifier

A statistical test is used by the C4.5 algorithm to assign each node's properties based on the entropy measure. The attributes are chosen to build the tree based on the attribute with the highest information gain ratio. Gain Ratio (A, S) of an attribute, which is a proportional to the sample set S, represents the information gain ratio in Eq. (3.5)

$$\text{Gain Ratio}(A,S) = \frac{\text{Gain}(A,S)}{\text{spilt Inf formation}(A,5)} \quad (3.5)$$

Where Gain

$$(A,S) = \text{Ent}(S) - \sum_{a\in A} \frac{|S_a|}{|S|}\text{Ent}(S_a) \quad (3.23)$$

And Split Information $(A,S) = -\sum_{a\in A} \frac{|S_a|}{|S|}\log_2\frac{|S_a|}{|S|}$ (3.5)

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

Here Sa is the subset of S. For discrete valued attributes, the information gain ratio can be calculated directly; however, continuous valued attributes must first be discretized before the information gain ratio can be determined.

IV. 4. RESULTS AND DISCUSSIONS

4.1 General

It has been described how to delineate lesions, extract features, and classify the liver in CT images using segmentation and classification. 112 CT scans were acquired as part of a dataset from the imaging centres of IMS and SUM Hospitals in India. The images were obtained using a CT scanner from GE Medical Systems with slices ranging in thickness from 0.5 mm to 1.5 mm, and having a common resolution of 512 x 512. The work was carried out using a personal computer with an Intel (R) Core (Tm) i5 64 based processor, 2.40 GHz Processor speed, and 4GB RAM using Math Works (MATLAB 2015a), Weka machine learning tool with 32-bit OS.

4.2. Validation

PSO and DPSO algorithms are used to process different CT images of liver cancer in order to segment the lesion border. Two parameters, PSNR and MSE values, which are computed from ROI pictures shown, are used to assess the image quality in Eq. (4.1) and (4.2).

$$\text{PSNR} = 10\log_{10}\left(\frac{R^2}{MSE}\right)$$

$$\text{MSE} = \frac{\sum_{M,N}[I_1(m,n)-I_2(m,n)]^2}{M\cdot N} \qquad (4.1)$$

where I_1 and I_2 are the input and segmented image, (M, N) represents the row and columns and R is the maximum fluctuation in the image. The various values of the retrieved features have a direct impact on the categorization metrics. In this study, approximation (A), horizontal (H), vertical (V), and diagonal (D) components are produced as a consequence of wavelet decomposition applied to the delimited zones. These components are recorded as pictures and used for statistical and morphological feature extraction. 2D-Haar wavelets were used for the wavelet analysis. For Haar wavelets, decomposition up to three layers was done here. The morphological and statistical characteristics

The segmented lesions' statistical and morphological properties are then extracted, and their dimension was decreased using a ranking approach to lessen computational complexity. The C4.5 decision tree classifier is then instructed to classify four different forms of liver cancer using the chosen feature set. Tables 4.2 and 4.3 give the findings of 10-fold cross validation for the classifiers for the two techniques. The table shows that the segmentation procedure had average classification accuracy for the DPSO and PSO methods of 98.26% and 92.011%, respectively. In PSO, the average DSC was 85.3%, and in DPSO, it was 96.38%. As can be observed in Tables 4.1 and 4.2, additional measured metrics performed similarly well in the DPSO segmentation strategy compared to PSO. The comparison plot utilising the C4.5 classifier of all the measured parameters for the PSO and DPSO algorithms is displayed. We can see from the plot that the segmentation method based on DPSO and C4.5 classifier is superior to the PSO method for identifying liver cancer lesions.

Few recent work has been reported the automatic system for detection of liver lesion. A CAD system to diagnose liver cancer in 71 histologically-proven liver tumors that includes 49 benign and 22 malignant lesions. The grey level co-occurrence matrix (GLCM) based features of tumors were extracted from CT images classified with binary logistic regression analysis. The evaluated performance yielded the accuracy of 81.69% and sensitivity of 81.82% in combined feature set. Compared with the published approaches, our method produced better result in detecting the liver cancerous cells which is presented in Table 4.4.

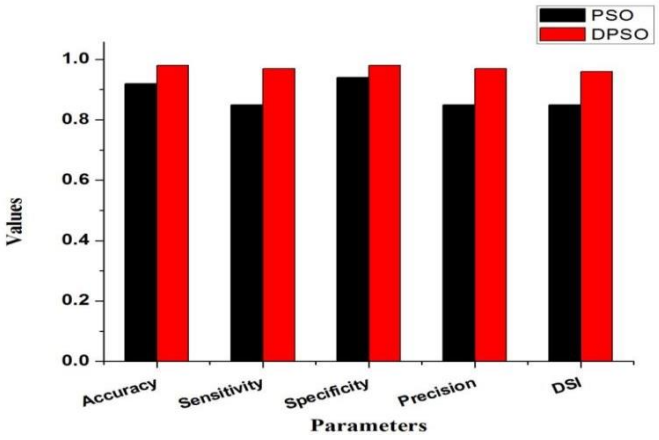

Figure 4.1**:** Comparison of different measures for PSO and DPSO methods using C4.5 classifier

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
Article Received: 25 November 2022 Revised: 30 December 2022 Accepted: 30 January 2023

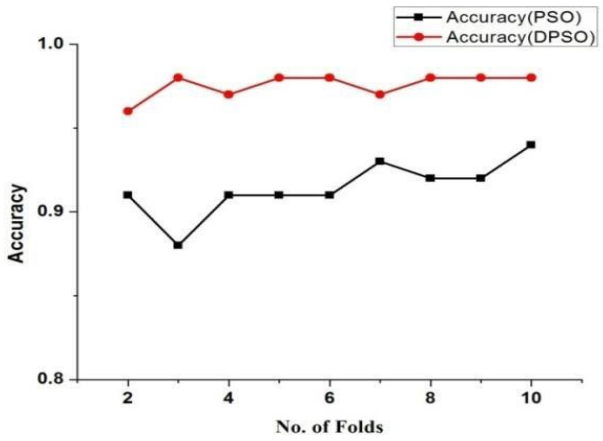

Figure 4.2**:** The accuracy test of C4.5 classifier for PSO and DPSO segmentation

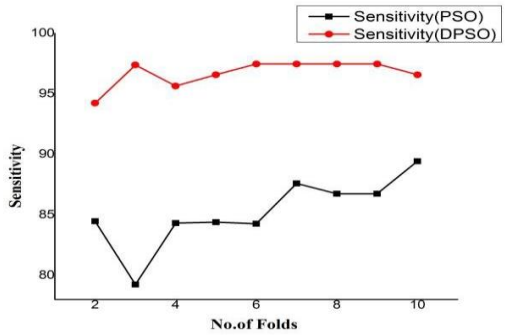

Figure 4.3: The sensitivity test of C4.5 classifier for PSO and DPSO segmentation

Table 4.1

Parameter evaluation for image quality

Parameter	PSO	DPSO
PSNR	97.56	115.75
MSE	607.9	405

Table 4.2

Estimation of result using C4.5 classifier with DPSO based segmentation

No. of folds	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	DSC (%)	MCC (%)
2	96.88	94.22	98.04	93.33	93.34	91.65
3	98.65	97.37	99.12	97.13	97.23	96.34
4	97.77	95.63	98.54	95.33	95.44	93.99
5	98.21	96.56	98.86	96.17	96.22	95.16
6	98.66	97.45	99.13	97.17	97.25	96.41
7	98.66	97.45	99.13	97.17	97.25	96.41
8	98.66	97.45	99.13	97.17	97.25	96.41
9	98.66	97.45	99.13	97.17	97.25	96.41
10	98.21	96.56	98.86	96.17	96.22	95.16
Avg.	**98.26**	**96.68**	**98.88**	**96.31**	**96.38**	**95.33**

Table 4.3

Comparison table with some existing techniques

Authors	Modality/Classes	Classifiers	Size of database	Accuracy (%)
Gletsoset al.[44]	CT/Normal Hepatic cyst, Hemangioma	Neural network classifier(NN)	147	97
Sethiet al.[47]	CT/Tumor, cyst calculi, normal liver	Genetic Algorithm, SVM, ANN	120	95.1
Kumar et al.[32]	CT/two(Hepatocellular, Hemangioma)	Probabilistic Neural Network(PNN)	150	96.7
Proposed	**CT/HCC, hemangioma, cyst, metastatic carcinoma**	**C4.5 Decision Tree Classifier**	**112**	**PSO: 92.01 DPSO: 98.26**

We have provided an automated system for segmenting and finding liver cancer in CT images that is based on improved methodologies. The PSO and DPSO methods were implemented for lesion segmentation and the features extracted from the images is classified in C4.5 classifier in k-fold cross validation process to classify four types of cancers such as cyst, hemangioma, hepatocellular

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

carcinoma (HCC), and metastatic carcinoma (MET).Feature selection technique used by filter method have been employed on the feature vector to reduce its dimensionality and to decrease the complexity of the classifier. The feature extraction module consists of 20 statistical and morphological features was derived from the segmented images. The detection results are promising. The cancerous lesions area was segmented effectively with average accuracy of 98.26% in DPSO method which was better than PSO method having average accuracy 92.01%.These obtained result is better compared to few published result as shown in Table 4. The focus of future study will be on improving the algorithm's efficacy in identifying cancer lesions utilising a greater number of slices. Expanding to 3D analysis on larger datasets will be the focus of future study.

V. CONCLUSION

Now a days, the application of machine learning techniques have been more in medical diagnosis. CT scan images provide an effective way to detect cancer lesion in the liver. The CAD methodology proposed here consists of several steps: liver segmentation, cancer lesion extraction, feature extraction or transformation and classification. The main objective of this thesis is to develop the automatic methods based on machine learning approach for accurate detection of liver cancer in order to help radiologists in the clinical practice. The thesis' primary contribution to the process of liver cancer lesion classification and automatic detection for clinical diagnosis. For the purpose of detecting liver cancer lesions, the best approaches based on PSO and DPSO have been given. With the help of the C4.5 decision tree classifier, wavelet-based statistical and morphological features were retrieved and categorised. The method was evaluated using 112 real-time CT images, which were divided into four categories: cyst, hemangioma, hepatocellular carcinoma, and metastatic carcinoma. The accuracy for the effective segmentation is 0.92 in PSO and 0.98 in DPSO.

REFERENCES

[1.] M. Hassoon, M. S. Kouhi, M. Z. Moghadam and M. Abdar, "Rule optimization of boosted c5. 0 classification using genetic algorithm for liver disease prediction," in Proc. IEEE International Conference on Computer and Applications, 2017, pp. 299-305.

[2.] M. Abdar, M. Zomorodi-Moghadam, R. Das and I. H. Ting, "Performance analysis of classification algorithms on early detection of liver disease," Expert Systems with Applications, vol. 67, pp. 239-251, 2017.

[3.] M. Isaac, Human Liver Anatomy, Function, Location, Parts & Diseases. Available: https://www.organsofthebody.com/liver/, dated 18/06/2018.

[4.] S. N. N. Alfisahrin and T. Mantoro, "Data mining techniques for optimization of liver disease classification," in Proc. IEEE International Conference on Advanced Computer Science Applications and Technologies, 2013, pp. 379-384.

[5.] R.-H. Lin and C.-L. Chuang, "A hybrid diagnosis model for determining the types of the liver disease," Computers in Biology and Medicine, vol. 40, no. 7, pp. 665- 670, 2010.

[6.] M. Clinic, Liver disease. Available: https://www.mayoclinic.org/diseases- conditions/liver-problems/symptoms-causes/syc-20374502, dated 17/03/2021.

[7.] S. K. Asrani, H. Devarbhavi, J. Eaton and P. S. Kamath, "Burden of liver diseases in the world," Journal of hepatology, vol. 70, no. 1, pp. 151-171, 2019.

[8.] R. H. Lin, "An intelligent model for liver disease diagnosis," Artificial Intelligence in Medicine, vol. 47, no. 1, pp. 53-62, 2009.

[9.] T. LeDuc, World Health Ranking. Available: https://www.worldl ifeexpectancy.com/cause-of-death/liver-disease/by-country/, dated 08/04/2019.

[10.] A. Khakhar, A liver disease in india. Available: http://www.livertransplant.org/ liver-transplantation/awareness/liver-diseases-in-india-stats, dated 08/04/2019.

[11.] K. Nagaraj and A. Sridhar, "NeuroSVM: a graphical user interface for identification of liver patients," arXiv preprint arXiv:1502.05534, 2015.

[12.] M. B. Priya, P. L. Juliet and P. Tamilselvi, "Performance analysis of liver disease prediction using machine learning algorithms," International Research Journal of Engineering and Technology, vol. 5, no. 1, pp. 206-211, 2018.

[13.] P. Rajeswari and G. S. Reena, "Analysis of liver disorder using data mining algorithm," Global Journal of Computer Science and Technology, vol. 10, no. 14, pp. 48-52, 2010.

[14.] S. Vijayarani and S. Dhayanand, "Liver disease prediction using SVM and Naïve Bayes algorithms," International Journal of Science, Engineering and Technology Research (IJSETR), vol. 4, no. 4, pp. 816-820, 2015.

[15.] I. Yoo et al., "Data mining in healthcare and biomedicine: a survey of the literature," Journal of Medical Systems, vol. 36, no. 4, pp. 2431-2448, 2012.

[16.] M. H. Tekieh and B. Raahemi, "Importance of data mining in healthcare: a survey," in Proc. IEEE/ACM

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
Article Received: 25 November 2022 Revised: 30 December 2022 Accepted: 30 January 2023

International Conference on Advances in Social Networks Analysis and Mining, 2015, pp. 1057-1062.

[17.] U. Fayyad, G. P. Shapiro and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," Communications of the ACM, vol. 39, no. 11, pp. 27-34, 1996.

[18.] S. I. Inc. (20/04/2020). Data Mining. Available: https://www.sas.com/en_in/insights/analytics/data-mining.html

[19.] M. Srivastava, R. Garg and P. Mishra, "Analysis of data extraction and data cleaning in web usage mining," in Proc. International Conference on Advanced Research in Computer Science Engineering & Technology, 2015, pp. 1-6.

[20.] N. Khare, N. Adlakha and K. Pardasani, "An algorithm for mining multidimensional association rules using boolean matrix," in Proc. IEEE International Conference on Recent Trends in Information, Telecommunication and Computing, 2010, pp. 95-99.

[21.] N. Kumar, H. Akash, R. A. Prataap, G. Srinath and C. Mala, "Intelligent intrusion detection system using decision tree classifier and bootstrap aggregation," in Proc. IEEE 8th International Symposium on Embedded Computing and System Design, 2018, pp. 199-203.

[22.] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.

[23.] A. Dey, "Machine learning algorithms: a review," International Journal of Computer Science and Information Technologies, vol. 7, no. 3, pp. 1174-1179, 2016.

[24.] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," in Proc. IEEE Seventh International Conference on Contemporary Computing , 2014, pp. 437-442.

[25.] G. Edwards, Machine Learning: An Introduction. Available: https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0, dated 23/04/2020.

[26.] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," IEEE Transactions on Cognitive Communications and Networking, vol. 4, no. 4, pp. 648-664, 2018.

[27.] Priyadharshini, What is Machine Learning and How Does It Work. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-machine- learning, dated 26/04/2020.

[28.] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 3, no. 1, pp. 1-130, 2009.

[29.] S. B. Kotsiantis, I. Zaharakis and P. Pintelas, "Supervised machine learning: A review of classification techniques," Emerging Artificial Intelligence Applications in Computer Engineering, vol. 160, no. 1, pp. 3-24, 2007.

[30.] S. R. Kulkarni, G. Lugosi and S. S. Venkatesh, "Learning pattern classification-a survey," IEEE Transactions on Information Theory, vol. 44, no. 6, pp. 2178-2206, 1998.

[31.] U. M. Khaire and R. Dhanalakshmi, "High-dimensional microarray dataset classification using an improved adam optimizer (iAdam)," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 11, pp. 5187-5204, 2020.

[32.] R. K. Jade, L. Verma and K. Verma, "Classification using Neural Network and Support Vector Machine for Sonar Data Set," International Journal of Computer Trends and Technology, vol. 4, no. 2, pp. 116-119, 2013.

[33.] D. Chatterjee and S. Chandran, "Prediction and Classification of Heart Disease using AML and Power BI," in Proc. 11th International Joint Conference on Computational Intelligence, 2019, pp. 508-515.

[34.] A. Anil and S. R. Singh, "Effect of class imbalance in heterogeneous network embedding: An empirical study," Journal of Informetrics, vol. 14, no. 2, pp. 1-17, 2020.

[35.] A. Tiwari, S. Srivastava and M. Pant, "Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019," Pattern Recognition Letters, vol. 131, pp. 244-260, 2020.

[36.] B. Chandra and V. P. Paul, "A robust algorithm for classification using decision trees," in Proc. IEEE Conference on Cybernetics and Intelligent Systems, 2006, pp. 1-5.

[37.] J. Ker, Y. Bai, H. Y. Lee, J. Rao and L. Wang, "Automated brain histology classification using machine learning," Journal of Clinical Neuroscience, vol. 66, pp. 239-245, 2019.

[38.] T. Deepa, B. Sathiyabhama, J. Akilandeswari and N. Gopalan, "Action fuzzy rule based classifier for analysis of dermatology databases," International Journal of Biomedical Engineering and Technology, vol. 15, no. 4, pp. 360-379, 2014.

[39.] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," Neural Computing and Applications, vol. 29, no. 10, pp. 685-693, 2018.

[40.] W. Yuan, D. Guan, Q. Zhu and T. Ma, "Novel mislabeled training data detection algorithm," Neural Computing and Applications, vol. 29, no. 10, pp. 673-683, 2018.

[41.] K. B. Nahato, K. H. Nehemiah and A. Kannan, "Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets," Informatics in Medicine Unlocked, vol. 2, pp. 1-11, 2016.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2124
***Article Received**: 25 November 2022 **Revised**: 30 December 2022 **Accepted**: 30 January 2023*

[42.] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information processing & management, vol. 45, no. 4, pp. 427-437, 2009.

[43.] A. K. Verma, S. Pal and S. Kumar, "Comparison of skin disease prediction by feature selection using ensemble data mining techniques," Informatics in Medicine Unlocked, vol. 16, pp. 1-10, 2019.

[44.] D.-Y. Liu, H.-L. Chen, B. Yang, X.-E. Lv, L.-N. Li and J. Liu, "Design of an enhanced fuzzy k-nearest neighbor classifier based computer aided diagnostic system for thyroid disease," Journal of Medical Systems, vol. 36, no. 5, pp. 3243- 3254, 2012.

[45.] Q. Kang, X. Chen, S. L and M. Zhou, "A noise-filtered under-sampling scheme for imbalanced classification," IEEE transactions on cybernetics, vol. 47, no. 12, pp. 4263-4274, 2016.

[46.] K. E. Bennin, J. Keung, P. Phannachitta, A. Monden and S. Mensah, "Mahakil: Diversity based oversampling approach to alleviate the class imbalance issue in software defect prediction," IEEE Transactions on Software Engineering, vol. 44, no. 6, pp. 534-550, 2017.

[47.] Y. Yan, R. Liu, Z. Ding, X. Du, J. Chen and Y. Zhang, "A parameter-free cleaning method for SMOTE in imbalanced classification," IEEE Access, vol. 7, pp. 23537-23548, 2019.

[48.] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 40, no. 1, pp. 185-197, 2009.

[49.] D. Meng, L. Zhang, G. Cao, W. Cao, G. Zhang and B. Hu, "Liver fibrosis classification based on transfer learning and FCNet for ultrasound images," IEEE Access, vol. 5, pp. 5804-5810, 2017.

[50.] L. Peng, H. Zhang, B. Yang and Y. Chen, "A new approach for imbalanced data classification based on data gravitation," Information Sciences, vol. 288, pp. 347- 373, 2014.

[51.] S. K. A. Numerical Simulation and Assessment of Meta Heuristic Optimization Based Multi Objective Dynamic Job Shop Scheduling System. *ijfrcsce* 2022, *8*, 92-98.

[52.] S. K. A. Design Simulation and Analysis of Deep Convolutional Neural Network Based Complex Image Classification System. *ijfrcsce* 2022, *8*, 86-91.

[53.] S.K.A. (2022). Numerical Simulation and Design of Power Fluctuation Controlling of Hybrid Energy Storage System Based on Modified Particle Swarm Optimization. *International Journal on Recent Technologies in Mechanical and Electrical Engineering*, *9*(3), 88–95.