Annexure-4

Final Report

1. Title of the Project: **Radiomics with Machine Learning Methods towards Prediction of Gall Bladder Cancer**

2. Unique ID of the Project (provided by ICMR): File No. 5/7/1679/NER/Adhoc/2019-RBMCH

3. Principal Investigator and Co-Investigators:

   Principal Investigator: **Dr. Rosy Sarmah**, Dept. of Computer Science & Engineering, Tezpur University, Tezpur, Assam, Pin:784028

   CO- PIs: i. **Dr. Abhijit Talukdar**, Dept. of Surgical Oncology, Dr B Borooah Cancer Institute, Guwahati, Assam

   ii. **Dr. Amal Chandra Kataki**, Director, Dr B Borooah Cancer Institute, Guwahati, Assam

   iii. **Prof. Dhruba Kumar Bhattacharyya**, Dept. of Computer Science & Engineering, Tezpur University, Tezpur, Assam

   iv. **Dr. Binoy Kumar Choudhury**, Dept. of Radiology, Dr B Borooah Cancer Institute, Guwahati, Assam

   v. **Dr. Geetanjali Barman**, Dept. of Radiology, Dr B Borooah Cancer Institute, Guwahati, Assam

   vi. **Dr. Debojit Boro**, Dept. of Computer Science & Engineering, Tezpur University, Tezpur, Assam

   vii. **Dr. Sauravjyoti Sarmah**, Dept. of Computer Science & Engineering, Jorhat Engg. College, Jorhat, Assam

4. Implementing Institution and other collaborating Institutions:

   (i) **Tezpur University,** Tezpur, Assam and

   (ii) **B Barooah Cancer Institute**, Guwahati, Assam, Guwahati,India

5. Date of commencement: 03/06/2020

6. Duration: 2 and 1/2 years

7. Date of completion: 02/12/2022

8. Objectives as approved:

Year	Activities
03/06/2020 To 02/12/2020	• Data Collection • To develop a GUI based Pre-processor to facilitate: (a) An intelligent GUI based pre-processor to enhance image quality and (b) modelling data from heterogeneous sources (clinical, and image) to support integrative analysis. • Parallel feature extraction to generate structured meta data to enable the multimodal image analysis towards detection of radiomic biomarkers for GBC. • Knowledge transfer (Capacity building, outreach, sharing): trainings, Publications (book, research papers, reviews, reports). • To develop solutions utilizing block chain approach for secure yet cost

	effective storage and transmission of health-related privacy preserved data.
03/12/2020 To 02/06/2021	• Data Collection • Supervised ML tool to support identification of GWT, presence of gallstones, liver metastasis detection, lymph nodes and para-aortic node involvement and selection of radiomic biomarkers. • To develop solutions utilizing block chain approach for secure yet cost effective storage and transmission of health-related privacy preserved data. • Knowledge Transfer: Workshop
03/06/2021 to 02/12/2021	• Data Collection • To develop a robust, peer to peer prototype for clinical (Big) data management system (CDMS) using private blockchain approach to facilitate researchers in systematic analysis of multi-omics data. This private blockchain will be initially deployed only at the partner institutions. • An unsupervised ML tool to support construction of radiomic biomarkers, detection of GWT, presence of gallstones, liver metastasis detection and lymph nodes and para-aortic node involvement.
03/12/2021 to 02/06/2022	• Data Collection • Develop a correlation engine to validate the identified radiomic biomarkers using meta data with reference to other clinical traits database. • Statistical, biological and clinical validation of the proposed AI/ML tools for identification using partner institutions to ascertain their clinical usefulness and biological relevance. • A smart App for GBC patients' awareness. • Knowledge transfer: Awareness campaign • To conceptualize and propose a detailed ToC of a book on **Multimodal images towards Decision support in Gall Bladder Cancer: A ML Perspective**.
03/06/2022 to 02/12/2022	• Data Collection • Modelling data from heterogeneous sources (clinical, and image) to support integrative analysis. • Develop a tool for secure, cost-effective storage and transmission of health-related privacy preserved data utilizing block chain approach. • Develop a semi-supervised algorithm to support construction of radiomic biomarkers, detection of GWT, liver metastasis detection and lymph nodes and para-aortic node involvement. • Develop an UNET architecture to segment the malignant regions from 2D CT images. • A CNN architecture has been developed to classify the malignant regions from CT images along with data augmentation using GAN. • To conceptualize and propose a detailed ToC of a book on **Medical images towards Decision support in Gall Bladder Cancer: A ML Perspective**.

9. Deviation made from original objectives if any, while implementing the project and reasons thereof: The data has been collected for 300 GBC and 150 Normal patients instead of 300 GBC and 50 Normal patients. This has facilitated the feature selection and classification problem by providing a more informative dataset.

10. Field/ Experimental work giving full details of summary of methods adopted.

The major contributions of the project work are summarized below:

I. **Data collection:** The project involved collection of CT scan images of 300 patients at Dr. B. Borooah Cancer Institute, Guwahati. It includes 150 CT scans of patients suffering from Gallbladder Cancer (GBC) and another 150 CT scan images of cancer patients with normal gallbladder. This data is collected in Dicom image format. Also, manual segmentation has been performed for all the 300 patient's data under the supervision of the radiologists, to obtain the ground truth and stored as NRRD files. We have also collected 15 USG images of GBC patients. The dataset also involves clinical data of all the patients comprising of patient's age, sex, cancer grading, biopsy reports, staging and personal information like name, contact number and address.

Sr No.	Patient	Age	Gender	Grading	Overall.Stage
1	Patient-1	56	Male	No Adenocarcinoma	NA
2	Patient-2	43	Male	No Adenocarcinoma	NA
3	Patient-3	48	Female	No Adenocarcinoma	NA
4	Patient-4	57	Female	No Adenocarcinoma	NA
5	Patient-5	47	Female	No Adenocarcinoma	NA
6	Patient-6	70	Female	No Adenocarcinoma	NA
7	Patient-7	41	Female	No Adenocarcinoma	NA
8	Patient-8	57	Female	No Adenocarcinoma	NA
9	Patient-9	50	Female	No Adenocarcinoma	NA
10	Patient-10	51	Female	No Adenocarcinoma	NA

Figure 1: Snapshot of Clinical data

II. **Feature Extraction:** We have successfully extracted around 1000 radiomic features from the CT scan images. The radiomic features extracted include the shape based 2D and 3D features, first order statistics, Gray Level Concurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Tone Matrix (GLSTM), Gray Level Dependence Matrix (GLDM) and Neighboring Gray Tone Difference Matrix (NGTDM). These features have been found for five different filtered images such as original image, squareroot image, logarithmic image, wavelet, squared and exponential image.

Patient	Image	Mask	original_s	original_s	original_s	original_s	original_s	original_s	original_s
1	F:\GBC_DA_DAT/	F:\GBC_D/	0.500625	0.343699	27.93258	81.27049	76.56894	57.86464	59.11529
2	F:\GBC_DA_DAT/	F:\GBC_D/	0.433139	0.316412	15.02353	47.48091	43.85262	26.05125	40.53094
3	F:\GBC_DA_DAT/	F:\GBC_D/	0.342538	0.254901	16.11534	63.22192	73.41734	29.03272	42.19743
4	F:\GBC_DA_DAT/	F:\GBC_D/	0.644936	0.556071	25.80035	46.39758	48.19633	45.48338	39.98068
5	F:\GBC_DA_DAT/	F:\GBC_D/	0.610776	0.482378	19.27544	39.95922	41.4507	44.73336	28.74203
6	F:\GBC_DA_DAT/	F:\GBC_D/	0.585197	0.414732	16.02782	38.64622	29.89091	32.28288	41.14029
7	F:\GBC_DA_DAT/	F:\GBC_D/	0.573336	0.419867	18.17132	43.27881	34.4019	49.28185	37.50023
8	F:\GBC_DA_DAT/	F:\GBC_D/	0.472783	0.269055	12.22655	45.44251	38.32509	34.18169	31.80362
9	F:\GBC_DA_DAT/	F:\GBC_D/	0.529869	0.409207	15.21884	37.19106	33.89128	29.61963	34.42082

Figure 2: Snapshot of Feature Extraction

III. **Unsupervised Segmentation**

(A) **RDS Method**: This method has been developed for effective segmentation of gallbladder and gallstones from CT scan and USG images. Delineation of Gallbladder (GB) and identification of gallstones from Computed Tomography (CT) and Ultrasonography (USG) images is an essential step in the radiomic analysis of Gallbladder Cancer (GBC). We have devised a method for effective segmentation of GB from 2D CT images and Gallstones from USG images, by introducing a Rough Density based Segmentation (RDS) method. Based on the threshold value obtained using rough entropy thresholding, the image is thresholded and passed as an input to the RDS method to obtain the desired segmented regions. To evaluate the performance of RDS method, we used images from 30 patients exhibiting normal GB and 8 patients with gallstones. Additionally, the versatility of our RDS method has also been tested for segmenting lungs from a publicly available Covid-19 lung CT image dataset with cohort size of 20 patients. Our method has been compared with several well-known algorithms like hybrid fuzzy clustering, Morphological active contour without edges, modified fuzzy c means and morphological geodesic active contours and found to give significantly better results with reference to Jaccard coefficient, dice coefficient, accuracy, precision, sensitivity, specificity and McNemar's test.

Algorithm 1 Proposed RDS segmentation Pseudocode

Require: CT/USG GB image.

1: Output: Segmented GB/gallstones region.
2: Convert input image into grayscale image.
3: Apply median filter and histogram equilization for pre-processing.
4: Select ROI.
5: Apply RET on the ROI (Eq:1)
6: Apply RDS algorithm on the thresholded image (Algorithm: 2).
7: Apply morphological dilation operation as post-processing for GB segmentation.

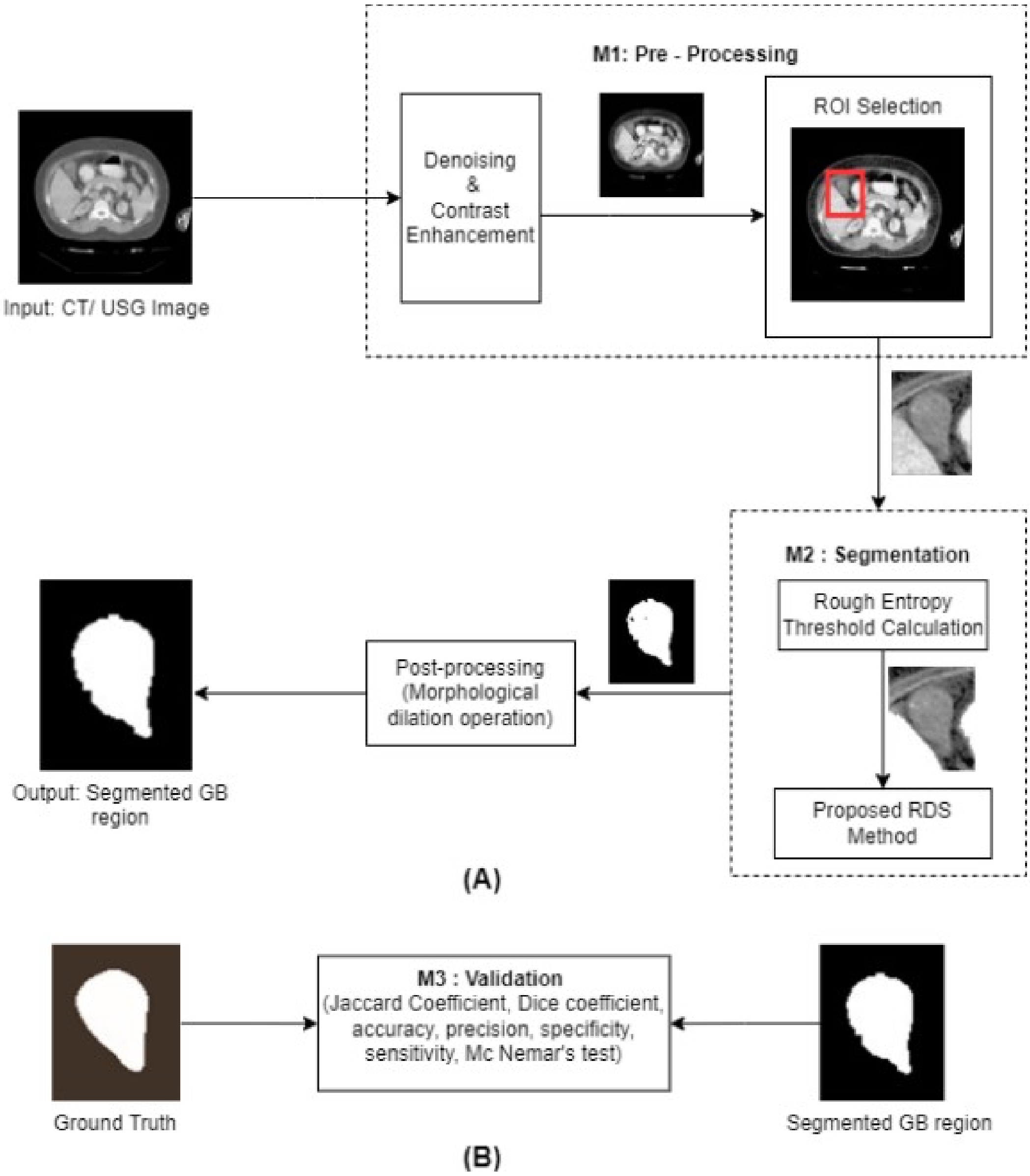


Figure 3: Workflow diagram of RDS method for segmentation of gallbladder and gallstones from medical CT images.

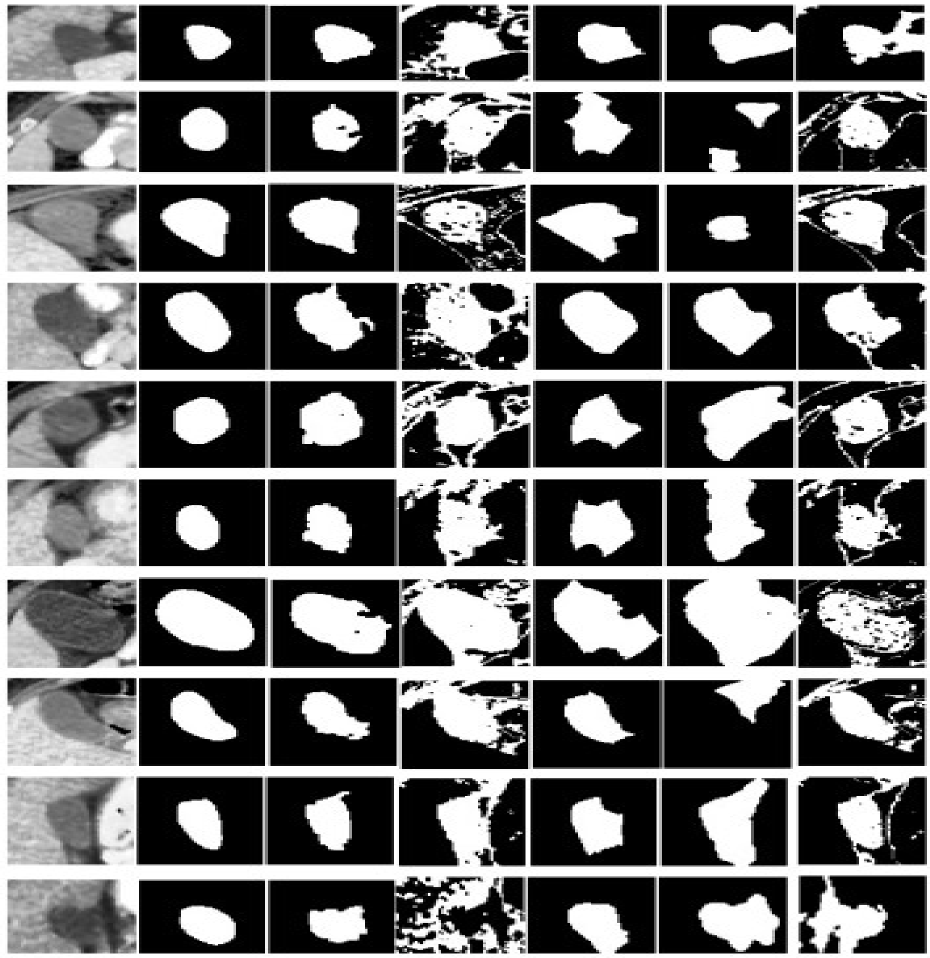
Method	Jaccard	Dice	Accuracy	Precision	Sensitivity	Specificity
MorphACWE	0.36	0.51	0.85	0.56	0.72	0.88
MorphGAC	0.57	0.73	**0.95**	0.82	**0.91**	0.96
HybridFuzzy	0.33	0.49	0.80	0.46	0.84	0.79
MFCM	0.48	0.64	0.87	0.59	0.89	0.87
Proposed	**0.78**	**0.87**	**0.95**	**0.87**	0.89	**0.97**

Figure 4: Performance comparison results using RDS segmentation for 30 2D CT images of Gallbladder.

Method	Jaccard	Dice	Accuracy	Precision	Sensitivity	Specificity
HybridFuzzy	0.03	0.06	0.58	0.91	0.48	0.84
MorphGAC	**0.46**	**0.64**	0.77	**0.99**	**0.72**	**0.97**
MFCM	0.08	0.15	0.75	**0.99**	0.70	0.96
Proposed	**0.39**	**0.57**	**0.72**	**0.99**	**0.62**	**0.96**

Figure 5: performance comparison results for USG images of gallstones using RDS segmentation.

1
2
3
4
5
6
7
8
9
10



A B C D E F G

Figure 6: Experimental Results of GB segmentation. (A) Original image ROI, (B) Ground Truth, (C) Output using Proposed RDS method, (D) Using HybridFuzzy approach, (E) Using MorphGAC, (F) Using MorphACWE, (G) using MFCM approach.

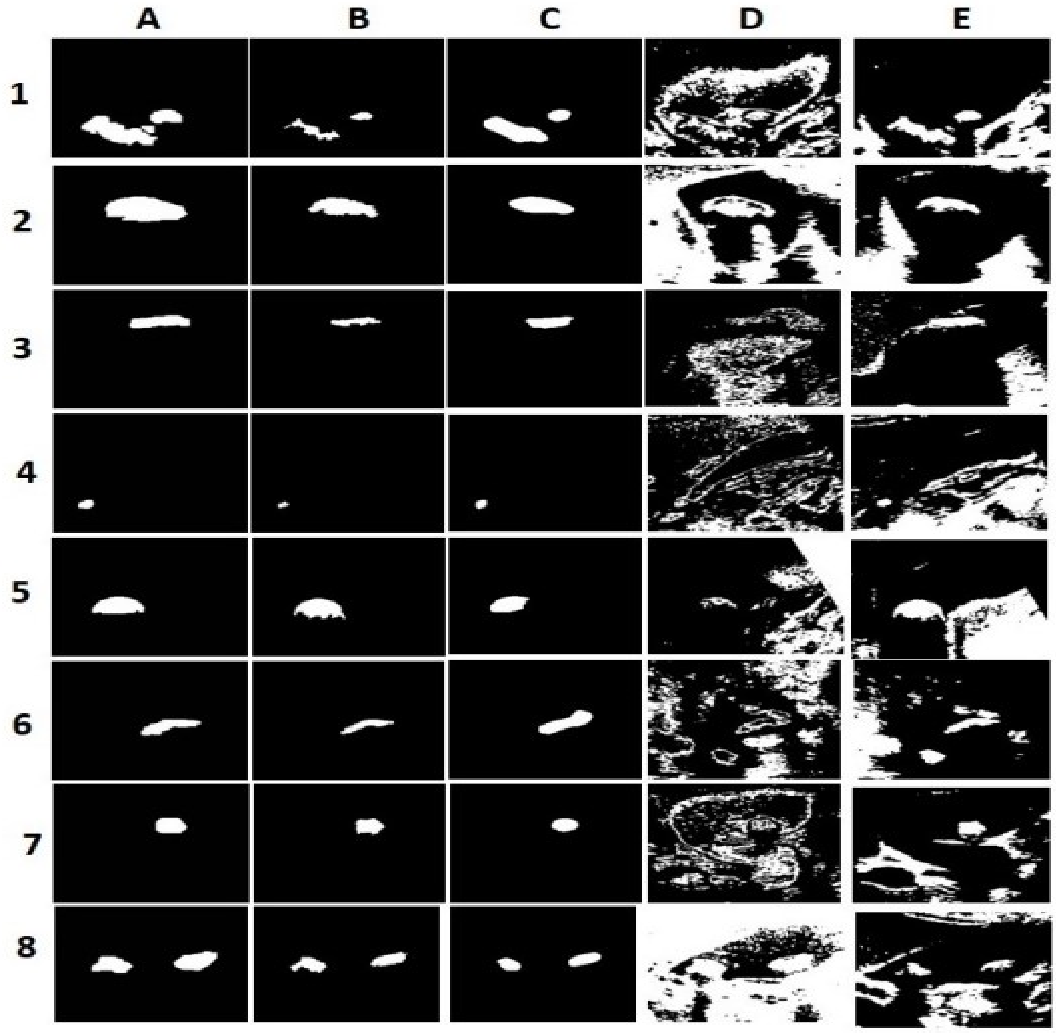


Figure 7: Experimental Results of gallstone segmentation. (A) Ground Truth, (B) Output using Proposed RDS method, (C) Using MorphGAC, (D) Using HybridFuzzy approach, (E) Using MFCM method.

(B) **GFDen Method:** Medical image segmentation is a critical process in many clinical applications, including disease diagnosis, treatment planning and medical research. Supervised learning approaches are commonly and widely exploited for medical image segmentation, but often they have been found expensive, time consuming, and demand large, annotated datasets. Semi-supervised approaches, on the other hand, are appealing since they can automatically segment images with less prior knowledge. Hence, we proposed a grid radiomic feature based density segmentation method for medical images called GFDen method. The medical image is first pre-processed and approximated as grid cells and radiomic features are extracted from each grid cell. Then, weights are assigned to each feature and an average feature weight matrix is constructed, which is used as an input to the density-based segmentation method based on a weighted cosine similarity measure. We evaluate the method on several medical computed tomography (CT) image datasets for four major applications: Liver and liver tumor segmentation, Gallbladder mass segmentation, Liver infiltrated Gallbladder mass segmentation from abdominal CT images and Covid-19 affected lung segmentation from chest CT images. The performances have been compared with four well-known segmentation methods namely, K-means clustering, FCM, EnFCM and MFCM and found to give significantly better results with reference to statistical validation parameters such as Jaccard coefficient, Dice coefficient, accuracy, F1 score, Precision, Specificity and Sensitivity.

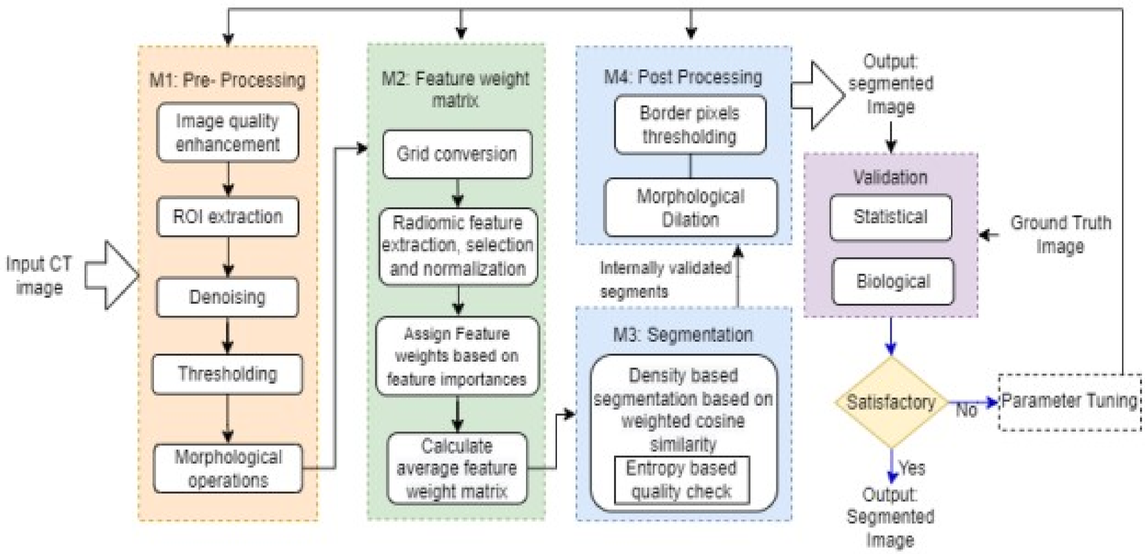


Figure 8: Workflow diagram of proposed GFDen method.

Algorithm 1 Pseudocode for proposed GFDen Segmentation method

---

Require: Medical CT scan image

1: Convert input image into grayscale image.
2: Select ROI
3: Apply pre-processing (median filter + thresholding + morphological operations).
4: Convert image into grid cells.
5: Feature extraction and normalization for each grid cell
6: Assign feature weights to all features (refer Definition 2, section 3.2.3)
7: Create average feature weighted matrix using Average feature weights (refer Definition 3)
8: Apply proposed density based segmentation algorithm. (refer Algorithm 2)
9: Apply post-processing (thresholding border pixels + morphological dilation operation).
10: **Output**: Segmented Area of Interest (AOI).

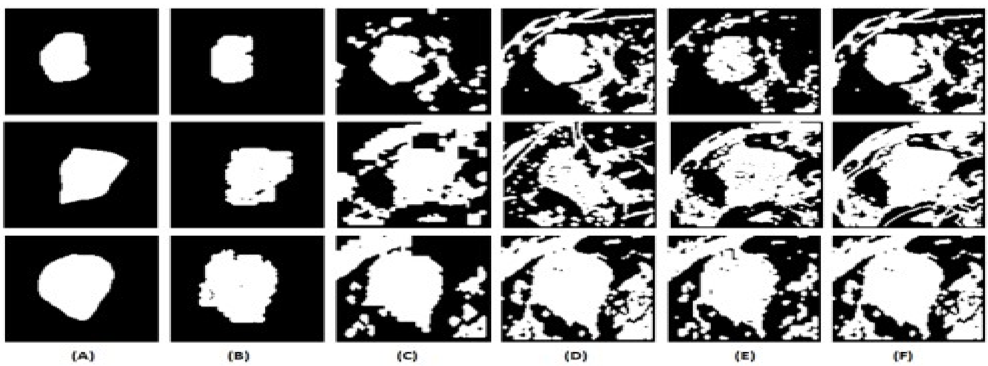


Figure 9: Experimental Results of GB mass segmentation. (A) Ground Truth, (B) Output using Proposed GFDen method, (C) Using K-means method, (D) Using FCM method, (E) Using EnFCM method, (F) Using MFCM method.

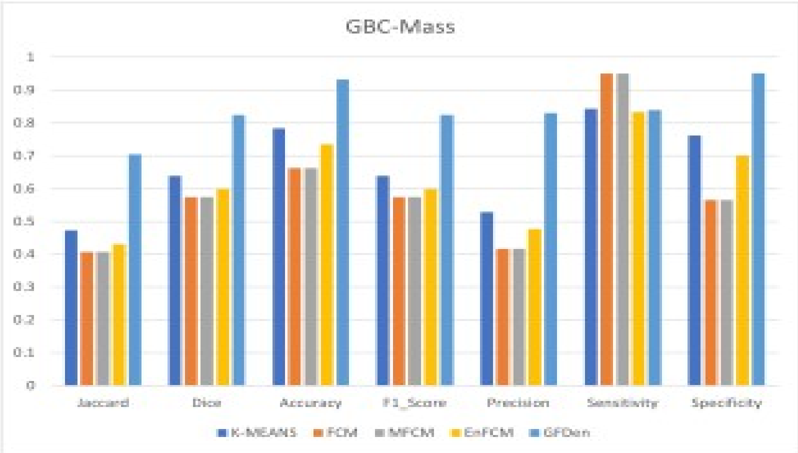


Figure 10: Performance comparison results for GB mass segmentation.

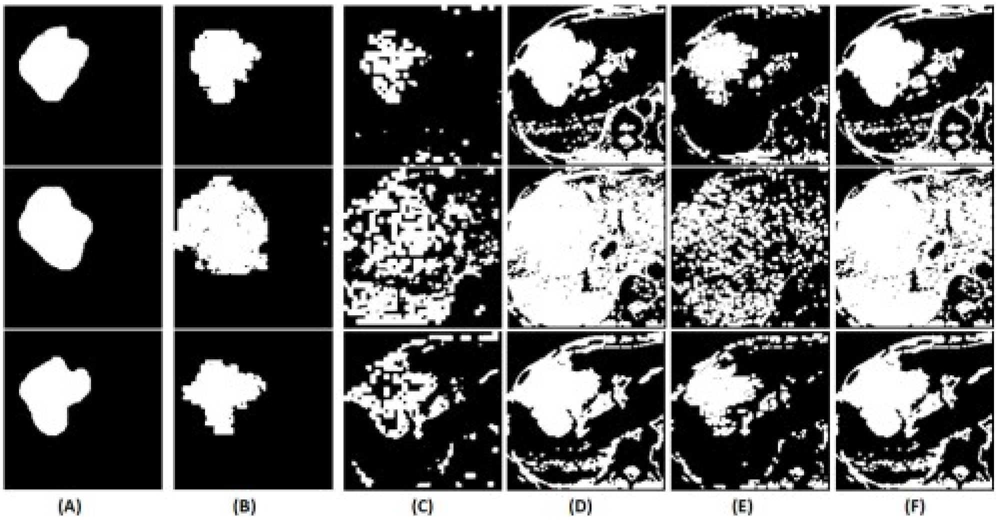


Figure 11: Experimental Results of GBC-Liver infiltration segmentation. (A) Ground Truth, (B) Output using Proposed GFDen method, (C) Using K-means method, (D) Using FCM method, (E) Using EnFCM method, (F) Using MFCM method.

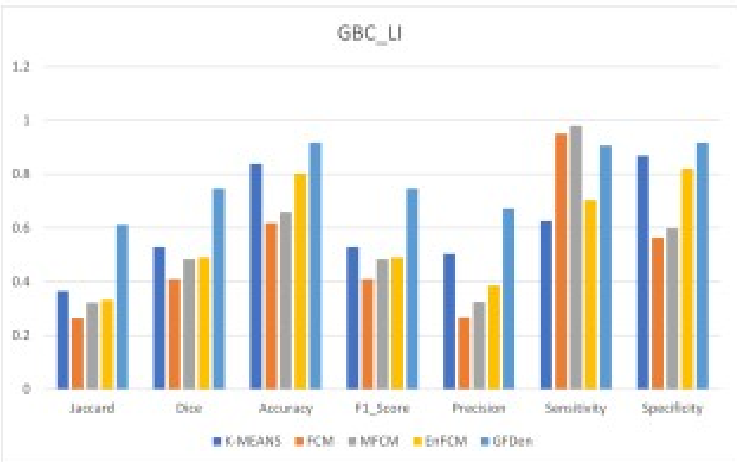


Figure 12: Performance comparison results for GB liver infiltration segmentation.

IV. **GBC_DA Application:** An effective platform called GBC Data and Analyser (GBC DA) has been developed to store CT scan images of gallbladder cancer along with its ground truth and clinical data in an efficient way, maintaining patient privacy concerns. The platform provides simple but effective graphical tools for downloading, visualizing, and analyzing datasets, using blockchain technology to provide reliable storage and faster accessibility support through a user-friendly interface for storage of 251 patients data (can be extended), with the private information encrypted using 1024-bit asymmetric encryption. GBC DA is accessible at **agnee.tezu.ernet.in:8027** for academic and medical use. The supplementary material is available at
http://agnigarh.tezu.ernet.in/~rosy8/supplementary_information_GBC_DA.pdf

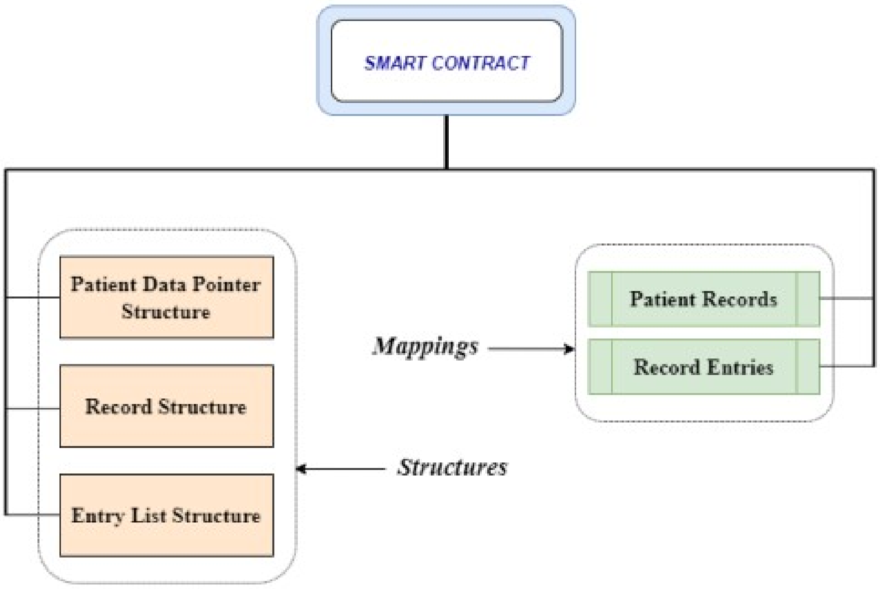


Figure 13: Proposed Smart Contract for block chain

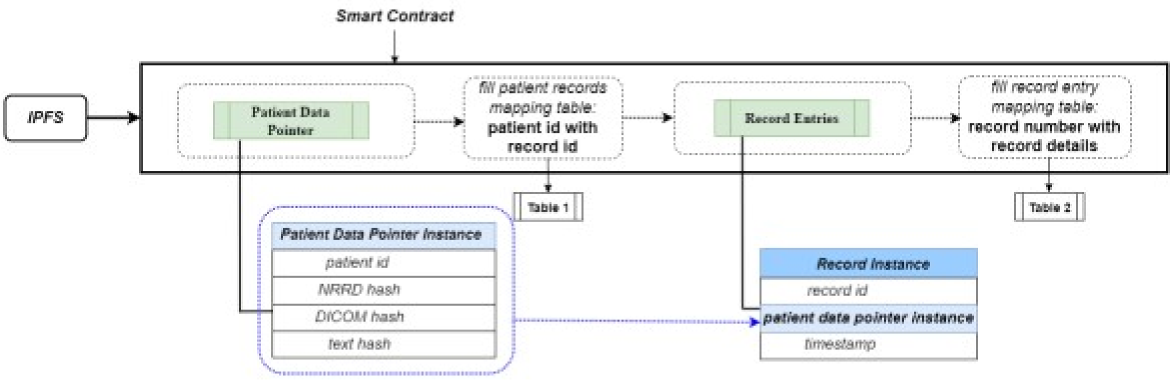


Figure 14: Sequential Flow Diagram of the proposed smart contract.

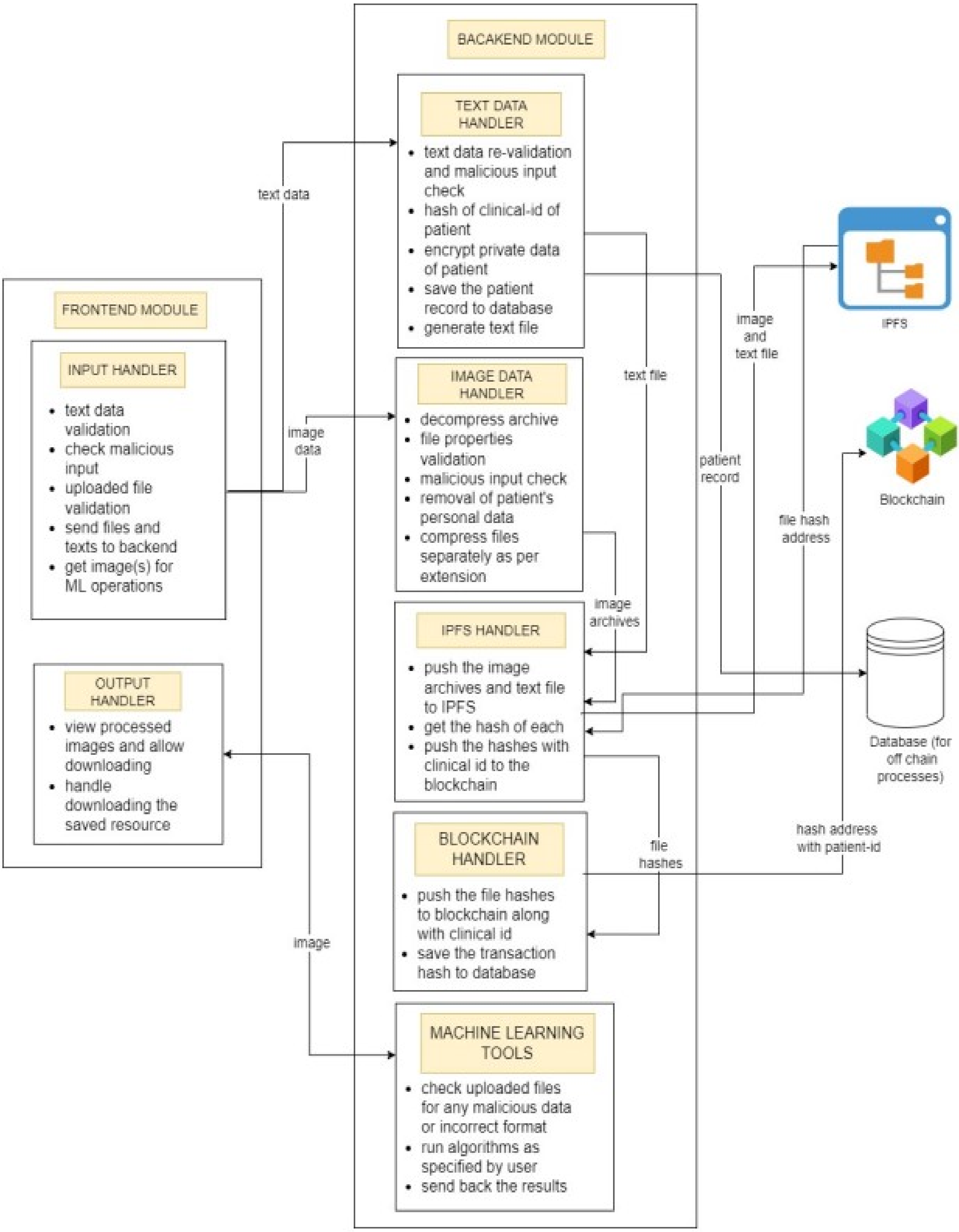


Figure 15: System architecture showing various modules and handlers for data movement around the system and performing activities specified by the user.

V. **EnRAFS Method for feature selection:** Given the large dimensionality of the clinical data, it is important to choose the most significant features to aid in improved classification of patients with respect to the subtypes/grades of GBC. Hence, we proposed a novel ensemble ranking-based approach called EnRaFS, for feature selection to grade GBC patients' using CT scan images. A new ranking measure has been proposed which combines the results of multiple feature selection methods to improve the accuracy of the ranking. The ranked features are then used to train the machine learning model to predict the grade of the cancer. The proposed approach has been evaluated on a dataset of 105 patients diagnosed with GBC and compared with other state-of-the-art feature selection methods based on accuracy measure. The proposed approach can be used as an effective tool for grading GBC, which can help clinicians to make more informed decisions about the treatment of the disease.

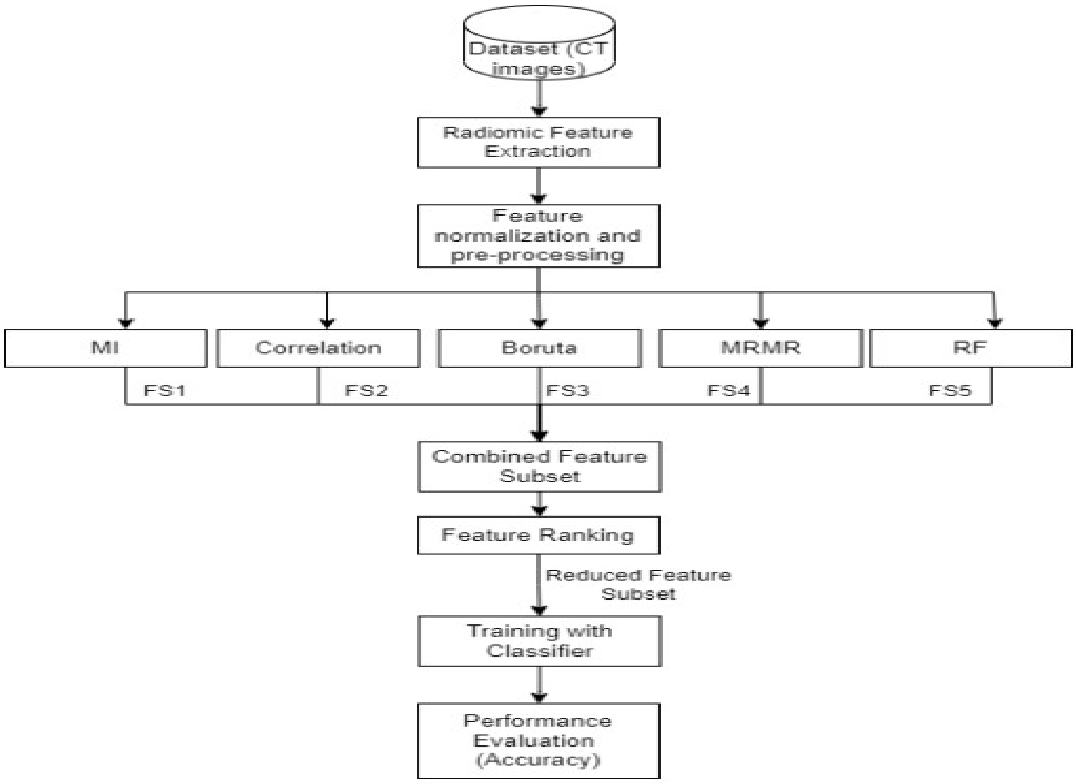


Figure 16: Workflow Diagram of EnRAFS method

FS method	Classifier	Selected Features	Accuracy
Boruta	LR	43	0.83
MI	SVM	453	0.86
Correlation	RF	32	0.82
RF	LR	224	0.85
MRMR	SVM	41	0.81
EnRaFS	LR	184	0.86

Figure 17: Performance comparison of ENRAFS with other feature selection methods. Our proposed method gives the best accuracy with the reduced feature set using LR classifier.

VI. **CNN Model for Classification:** We have proposed an approach that employs the power of Generative Adversarial Networks (GANs) for data augmentation in medical image classification tasks using CNNs. We have introduced a GAN-based augmentation method that generates synthetic images to augment the training dataset to overcome the issue of limited training data. We evaluate our models on a dataset of gallbladder medical images and compare the performance of CNN models trained with and without data augmentation. The results demonstrate that GAN-based data augmentation significantly improves the classification accuracy and performance of the model. Our findings highlight the effectiveness of data augmentation in enhancing the generalization capabilities of CNN models and overcoming the limitations of limited training data in medical image classification. This research provides valuable insights into the potential of GANs for data augmentation in the medical imaging domain in developing robust and accurate classification systems.

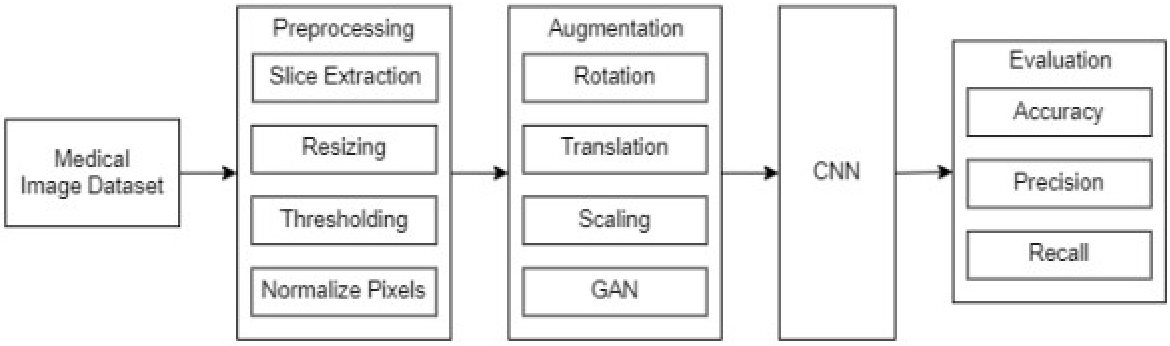


Figure 18: Workflow Diagram of proposed method

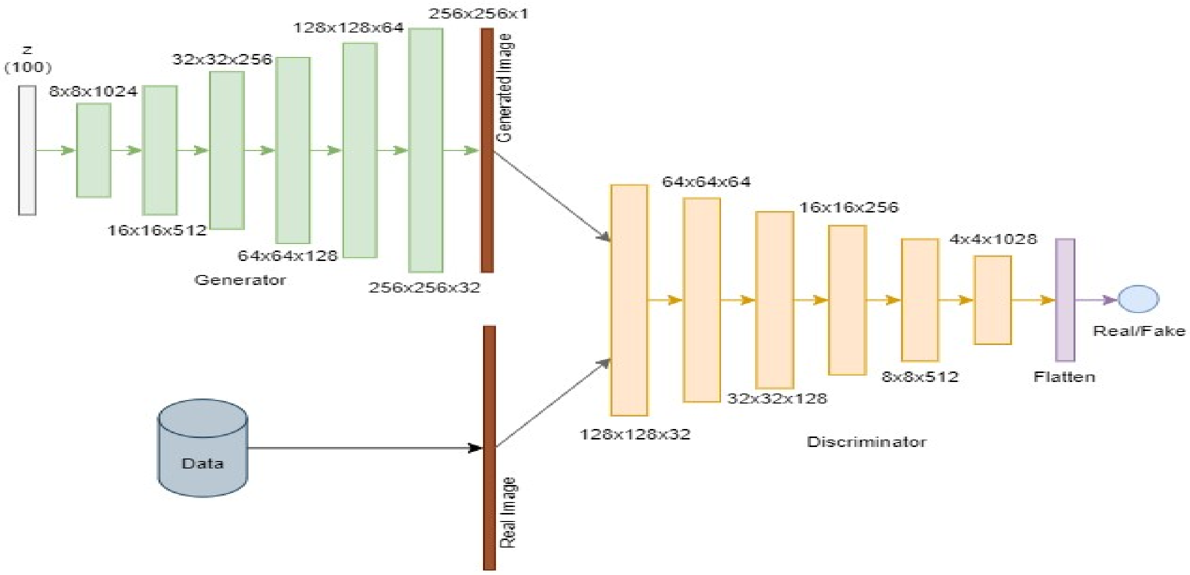


Figure 19: GAN architecture employed for data augmentation.

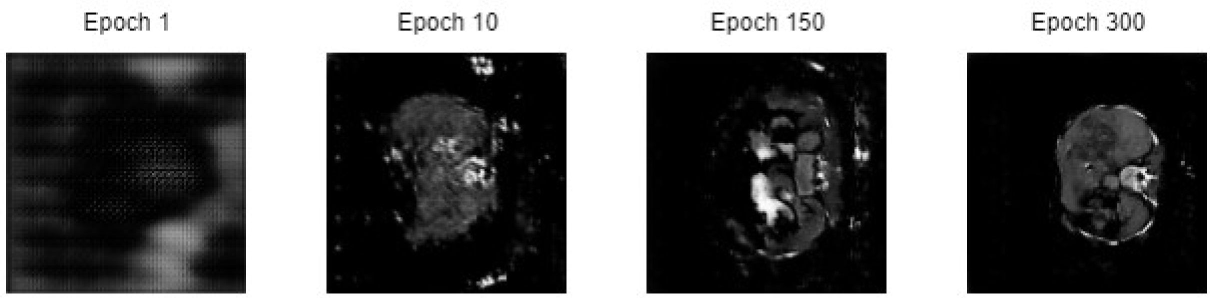


Figure 20: Visualization of progression of training using GAN.

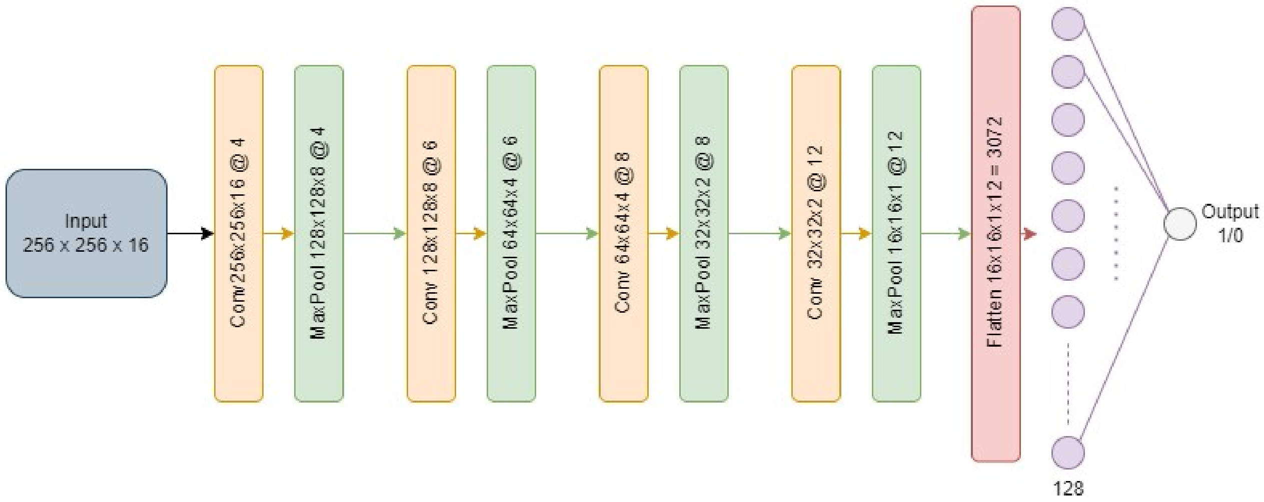


Figure 21: CNN architecture for classification of gallbladder cancer images as benign or malignant.

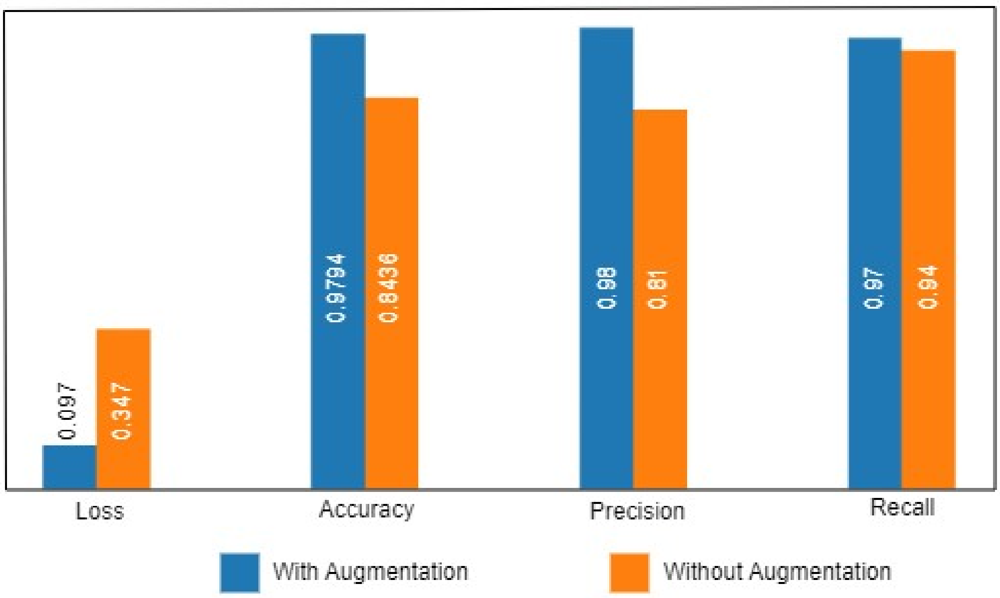


Figure 22: Performance evaluation for CNN model with and without augmentation for four evaluation metrics.

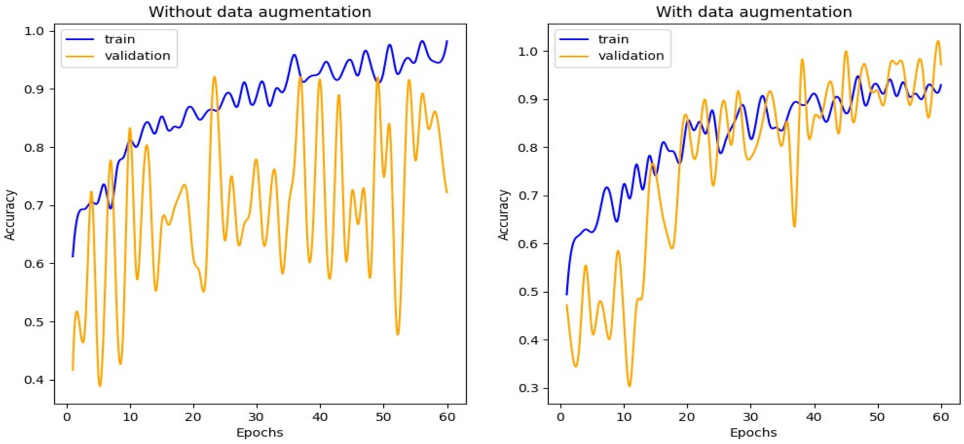


Figure 23: Accuracy plots using proposed CNN architecture.

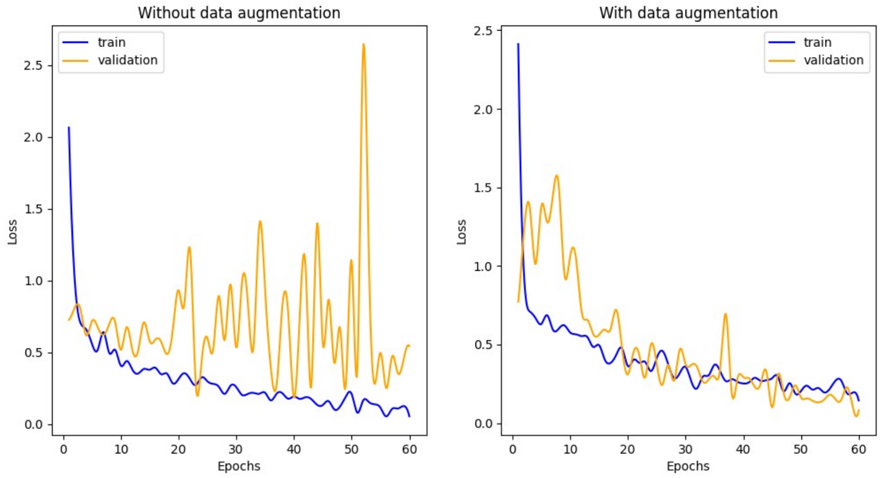


Figure 24: Loss curve using proposed CNN architecture.

VII. Supervised Segmentation

(A) EU-NET model: The segmentation of malignant regions indicative of gallbladder cancer from abdominal 2D CT scan images has been accomplished here by implementing a Deep Learning algorithm. An enhanced U-net (EU-Net) model has been introduced, obtained by training on Gallbladder cancer (GBC) dataset for segmentation of various malignancies. We compared the performance of the proposed EU-Net model to that of the U-Net, U-Net++, and Attention U-Net models and found that Eu-Net outperformed the latest state-of-the-art Deep Learning models in terms of IOU and Dice coefficient. Also, the efficiency of our proposed method has been tested on LITS dataset for liver segmentation.

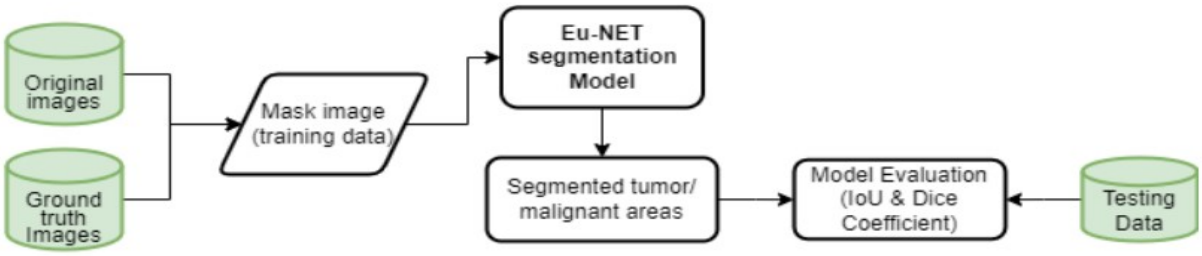


Figure 25: Workflow diagram for proposed segmentation model.

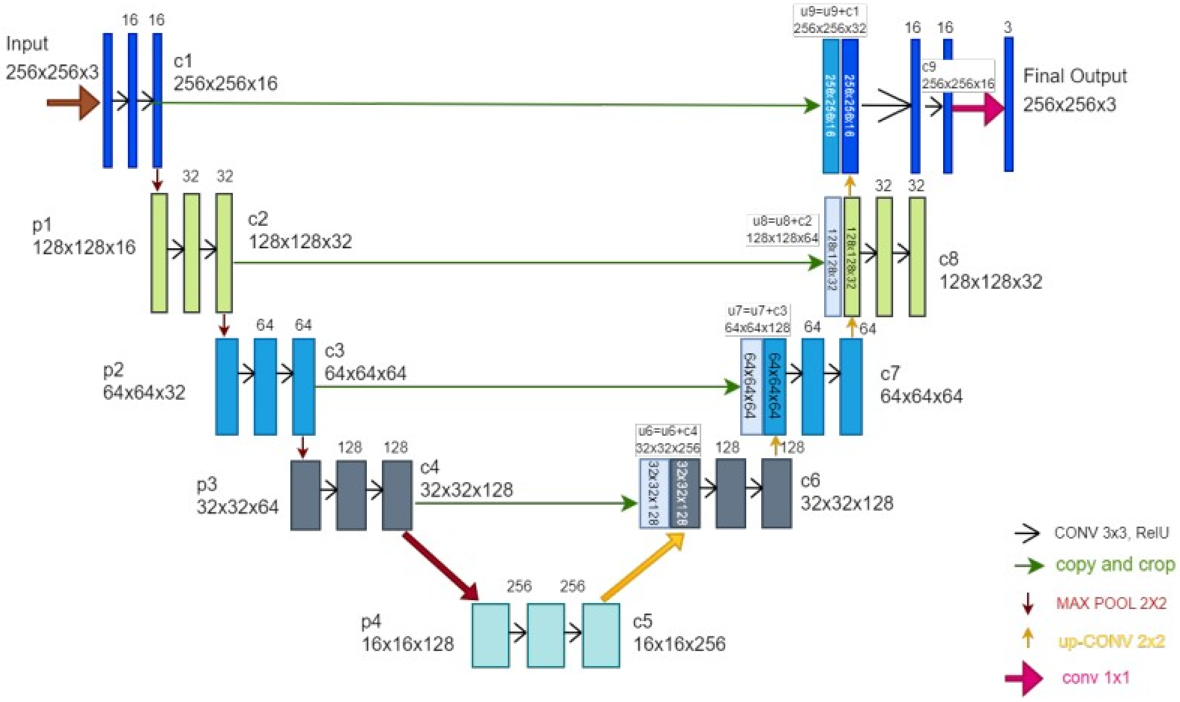


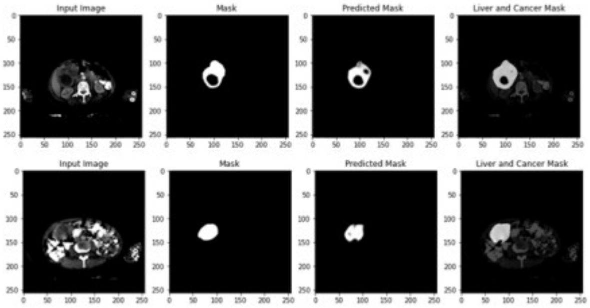
Figure 26: Architecture of proposed EU-NET model

VIII. **GBC Awareness app:** A GBC awareness app has been built which provides basic information to the common people regarding the symptoms, diagnosis, and patterns of gallbladder cancer.

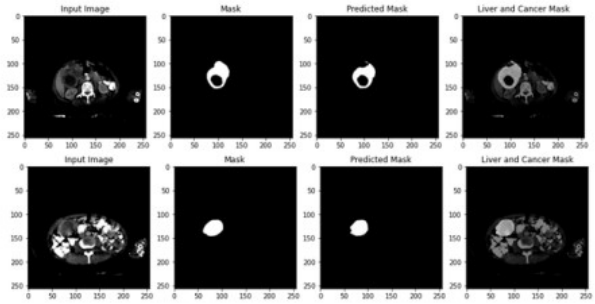
11. Supported by necessary tables, charts, diagrams, and photographs: Discussed in detail in previous section 10.

12. Detailed analysis of results: Discussed in detail in previous section 10.

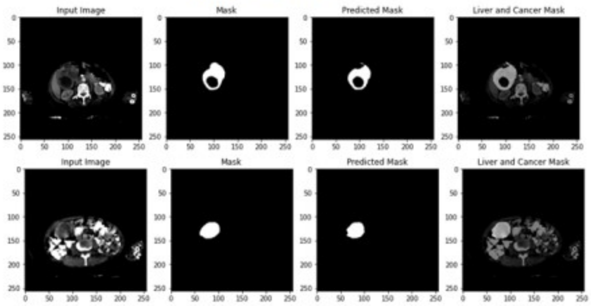
13. A summary sheet of not more than two pages under following heads (Title, Introduction, Rationale, Objectives, Methodology, Results, Translational Potential)



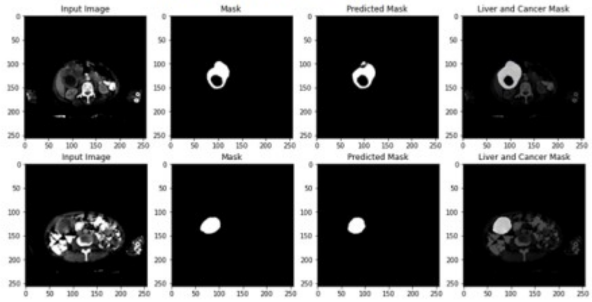
(A) U-Net++



(B) Attention U-Net



(C) U-Net



(D) EU-Net

Figure 27: Segmentation results for GBC dataset

TABLE 1: Comparison Table for LITS dataset.

Model	IoU	Dice-Coefficient
U-Net	0.8845	0.9286
U-Net++	0.8267	0.9049
Attention U-Net	0.8824	0.9374
EU-Net	0.9013	0.9480

TABLE 2: Comparison Table for GBC dataset.

Model	IoU	Dice-Coefficient
U-Net	0.7102	0.8265
U-Net++	0.6385	0.7777
Attention U-Net	0.7352	0.8465
EU-Net	0.7400	0.8494

Figure 28: Table 1 and Table 2 depicts the performance comparison of our Eu-NET model with three state-of-the-art CNN models for both LITS and GBC dataset.

14. Contributions made towards increasing the state of knowledge in the subject.

The contributions can be summarized as follows:

- A dataset of 300 GBC patients and 150 normal (300 images with ground truth (manually segmented) and 150 unannotated images) along with clinical data has been curated. It is a significant contribution to the research and medical community because there is a limited availability of data specifically dedicated to GBC analysis and it will be of immense value, allowing researchers to study and analyze Gallbladder Cancer more comprehensively.
- The radiomic features extracted from the medical CT scan images provide information of the various underlying features such as texture, shape, size, etc. which helps in better understanding the images, resulting in effective segmentation of the desired regions of interest and classification of patient images.
- The unsupervised segmentation methods, namely RDS and GFDen assist in accurately segmenting the malignant and healthy gallbladder regions from the CT scan images. They have also been validated for segmentation of liver from CT scan images and Covid-19 affected lungs from chest CT scans and proved to give significantly good results in terms of statistical validation measures.

- The supervised segmentation method, Eu-NET has outperformed state-of-the-art segmentation methods for identification of gallbladder malignancies from CT images. It has also been validated for liver and liver tumor segmentation from CT images.
- An effective ensemble feature selection method called EnRAFS has been proposed utilizing a new rank measure, RM, for selecting the most relevant feature subset from the original set of features. It has been compared to existing feature selection techniques and it outperformed significantly.
- A CNN model has been proposed for classification of the GBC dataset as benign or malignant, utilizing GAN framework for data augmentation. We have achieved classification accuracy of around 97% with data augmentation.
- An effective platform called GBC Data and Analyser (GBC_DA) has been developed to store CT scan images of gallbladder cancer along with its ground truth and clinical data in an efficient way, maintaining patient privacy concerns. The platform provides simple but effective graphical tools for downloading, visualizing, and analyzing datasets, using blockchain technology to provide reliable storage and faster accessibility support through a user-friendly interface for storage of 251 patients data (can be extended), with the private information encrypted using 1024-bit asymmetric encryption.

15. Conclusions summarizing the achievements and indication of scope for future work.

This project work involved exploring and developing Machine learning tools along with data collection and a secured storage system to assist the medical community in decision making for gallbladder cancer and its malignancies. A dataset of 300 patient + 150 normal images has been developed along with ground truth images and clinical data. The radiomic features have been extracted from these images and analyzed to detect and segment the malignant regions reflecting gallbladder cancer. An effective platform called GBC_DA has also been developed for storage of the dataset maintaining privacy concerns and security using block chain technology. In the future, more work can be done using multimodal images for detection of such malignancies.

16. Science and Technology benefits accrued:

I. List of research publications with complete details:

A. Published a research paper as a book chapter with details below:

1. N Jitani, B Singha, G Barman, A Talukdar, BK Choudhury, R Sarmah and D.K. Bhattacharyya.

"Gallbladder CT Image Segmentation by Integrating Rough Entropy Thresholding with Contours", *Advanced Computational Paradigms and Hybrid Intelligent Computing*, 651-659. DOI: 10.1007/978-981-16-4369-9_62

B. Presented in a conference and accepted as a book chapter:

1. Nitya Jitani, Anup Basumatary and Rosy Sarmah. "Deep learning-based Tumor Segmentation from CT images", *Advanced Computational and Communication Paradigms, Springer.*

C. Communicated papers:

Journal Papers:

1. N Jitani, B Singha, G Barman, A Talukdar, BK Choudhury, R Sarmah and D.K. Bhattacharyya.

“Rough density segmentation (RDS) method for medical Image segmentation, *Multimedia Tools and Applications*, Springer (under review).

2. Bhaskar Singha, Nitya Jitani, Rosy Sarmah and Dhruba k. Bhattacharyya. “EIMF: An Enhanced Iterative Mean Filter for effective denoising of medical CT images”, SN Computer Science, Springer (under review).

3. Nitya Jitani, Rosy Sarmah and D.K. Bhattacharyya. “GFDen: A Radiomic Grid-Feature based Density Segmentation Method for Medical Images using Unsupervised Approach”, Expert Systems with Applications, Elsevier (Submitted).

4. Ankit Sharma, Nitya Jitani and Rosy Sarmah. “Gallbladder CT image Classification using 3D CNN integrating GAN Augmentation”. (To be Communicated).

5. Abinash Dutta, Nitya Jitani, Kausthav P. Kalita, Geetanjali Barman, Abhijit Talukdar, Binoy K. Choudhury, Debojit Boro, Sauravjyoti Sarmah, Dhruba K. Bhattacharyya and Rosy Sarmah. “GBC_DA: A Platform for Storage and Analysis of Gallbladder Cancer CT Images”. Health and Techmology, Springer (Submitted).

Conference papers:

6. Nitya Jitani, Vivek K. Verma and Rosy Sarmah. “EnRaFS: An Ensemble Ranking-based Feature Selection Approach for Grading Gallbladder Cancer using Radiomic Analysis.”, Submitted to ICEEE 2023.

II. Manpower trained in the project:

a. Research Scientists or Research Fellows

b. No. of PhDs produced: 1.

c. Other Technical Personnel trained: 1.

III. Patents taken, if any: No

IV. Products developed, if any. Yes (GBC_DA: A Platform for Storage and Analysis of Gallbladder Cancer CT Images).

D. Abstract (300 words for possible publication in ICMR Bulletin).

Gallbladder Cancer (GBC) is a highly lethal malignancy with a high prevalence among the people of North and North East (NE) India and is predominantly found in females. Modalities commonly used to image the gallbladder are ultrasonography (USG) and computed tomography (CT). CT and USG can detect liver metastases/invasion, biliary or portal vein involvement, and enlarged lymph nodes. Early diagnosis is rare and many GBCs are not diagnosed preoperatively. This project planned to use AI/ML methods to detect gallbladder lesions and in stratifying the risk of GBC. A strategic multidisciplinary partnership between two institutions (one academic and the other medical) was set up which helped benchmark the oncology practice in NE India through AI/ML based analysis. The study involved collection of CT and USG images of 300 GBC patients and 150 non-GBC patients with the aim to use various ML methods to predict cancer radiomic biomarkers, gall bladder lesions and help guide treatment decisions for patients with GBC. To this end, a secure, distributed network database (GBC Data and Analyser (GBC_DA)) was curated to securely store, visualize, and analyze all

the clinical and image data ensuring the privacy and ethical responsibilities using blockchain technology. Further, 01 image denoising algorithm, 02 unsupervised ML segmentation methods, 01 feature selection method, 01 supervised ML method for segmentation of GBC malignancies, 01 Deep learning based augmentation and classification method, and an android GBC awareness App have been developed for diagnosis and prediction of Gallbladder cancer and it's various subtypes. The secured GUI based Gall bladder Cancer (GBC) diagnostic and predictive ML tools developed are capable at (i) enhancing image quality, (ii) extracting and selecting the most informative radiomic markers using various ML algorithms, (iii) identifying malignant areas using various proposed segmentation methods, and (iv) classifying the patient images into disease and non-disease as well as grade the malignancies.

E. Procurement/usage of Equipment

a.

Sl. No	Name of Equipment	Make/ Model	Quantity	Cost FE/₹	Date of Installation	Utilization rate %	Remarks regarding maintenance/breakdown
1	Dell Workstation with Intel Xeon W-2245 Processor 64 GB with 22inch Dell monitor	Dell Precision 5820 Tower	2	Rs 4,09,320/-	09-11-2020	95%	
2	Brother Mono Laser Multi-function Centre-PS Printer	DCP-B7535DW-PS	1	Rs 19,990/-	04/12/2020	90%	
3	ADATA 16 GB RAM Random Access Memory (RAM) SD RAM	LPX DDR4 PC4-19200 2400MHZ	4	Rs 89,999/-	05-04-2021	95%	

b. Suggestions for disposal of equipment. Once the equipment's become obsolete, they will be sent to central store for disposal.

Name and signature with date

1. Rosy Sarmah 17/07/2023

(Principal Investigator)

2. Abhijit Talukdar

(Dr. Abhijit Talukdar)

(Co-Investigator)

3. Geetanjali Barman

(Dr. Geetanjali Barman)

(Co-Investigator)

4. Debojit Boro

(Dr. Debojit Boro)

(Co-Investigator)

5. S Sarmah

(Dr. Sauravjyoti Sarmah)

(Co-Investigator)

UTILIZATION CERTIFICATE
(ANNUAL)

Certified that out of Rs. **22,39,182/- <u>(Twenty two lakhs Thirty Nine Thousand One Hundred Eighty Two Rupees)</u>** of grants-in-aid sanctioned during the year <u>03/06/2020 to 02/12/2022</u> in favour of <u>Registrar, Tezpur University</u> under ICMR Letter No. <u>5/7/1675/NER/Ad-hoc/2019-RBMCH</u>, total funds received is **<u>Rs 19,22,801/-</u>** and a sum of **<u>Rs 19,99,700/-</u>** has been utilized for the purpose of <u>Salary, Recurring, Non-recurring and overhead charges of the project (separate Sheet attached)</u> for which it was sanctioned and that the balance of **<u>Rs -76,899/-</u>** remaining unutilised at the end of the final year. The negative balance will be cleared after receiving the 10% retained amount from ICMR.

Rosy Sarmah Signature of Principal Investigator with date	30/5/23 Signature of Registrar/ of the Institute with date Registrar Tezpur University	30/5/23 Signature of Accounts Officer of the Institute with date Finance Officer Tezpur University

Annual Statement of Accounts

(Period 03/06/2020 to 03/12/2022)

1. Sanction letter No. : 5/7/1675/NER/Ad-hoc/2019-RBMCH
2. Total Project Cost : Rs. **19,99,700**
3. Sanction /Revised Project cost (if applicable) :
4. Date of Commencement of Project : 03/06/2020
5. Statement of Expenditure : From 03/06/2020 – 02/12/2022

S.No	Sanctioned/ Heads	Funds			Expenditure incurred			Remarks
					1st Year	2nd year	3rd year	
1.	Salaries	1,398,780/-			4,07,666/-	6,74,420/-	2,03,806/-	
2.	Permanent Equipment	5,20,000/-			5,19,309/-	-	-	
3.	Travel	80,000/-			-	3,432/-	-	
4.	Contingencies	1,72,000/-			37,896/-	63,012/- + 42,104/-	-	(Rs 42,014 is the amount committed in the first year)
5.	Overhead Expenses	68,402/-			21,055/-	27,000/-	-	
7.	**Total**	**22,39,182/-**			**9,85,926/-**	**8,09,968/-**	**2,03,806/-**	
		1st Year	2nd year	3rd year				
8.	**Funds allocated**	**12,67,448/-**	**7,68,994/-**	**2,02,740 /-**				
9.	**Funds received**	**12,67,448/-**	**4,52,677/-**	**1,96,670/- + 6006 = 2,02,676 /-**				**Rs 6006** is the bank interest adjusted by ICMR. Balance of 1st year, **Rs 2,39,418 was adjusted in 2nd year's** sanction amount. So, 2nd years total fund received is **Rs 4,52,677 + 2,39,418 = Rs 6,92,095.**
10.	**Balance**	**0**	**-76,899/-**	**0**				1st years balance was sent back to ICMR.
11.	**10% amount retained in 2nd year (yet to be received)**	**76,899/-**						The payment for the 2nd year negative balance was made using the 3rd years JRF salary. Hence, the JRF's salary for 3rd year can be only cleared after receiving the 10% retained amount.

*Since there is negative balance of Rs 76,899 from 2nd year due to 10% retained amount, the payment for the same has been made using the JRF's salary amount for 3rd year. After receiving the 10% retained amount of Rs 76,899, the remaining salary of the JRF will be disbursed.

Signature of Principal Investigator
with date

Signature of Accounts Officer
With date 3/5/23

Finance Officer
Tezpur University

Check list for covering note to accompany Utilization Certificate of grant for the project for the period ending 2nd December 2022)

1) Title of the project: ***Radiomics with Machine Learning Methods towards Prediction of Gall Bladder Cancer***

2) Name of the Institutions: **(i) Tezpur University, Tezpur, Assam and (ii) B Barooah Cancer Institute, Guwahati, Assam, Guwahati,India**

3) Principal Investigator: **Dr. Rosy Sarmah, Assoc. Prof., Dept. of Computer Science & Engineering, Tezpur University, Tezpur, Assam, Pin:784028**

4) ICMR letter No. and date sanctioning the project.
File No. 5/7/1679/NER/Ad-hoc/2019-RBMCH dated 06/03/2020

5) Head of account as given in the original sanction letter.
Budget Head: Salary, Recurring, Non-recurring, Travel and Overhead charges

6) Amount received during the financial year (Please give No. & Date of ICMR's sanction letter for the amount):
Rs. (1267448 (Year 1)+768994 (Year 2)+202740 (Year 3))= Rs. 2239182

7) Total amount that was available for expenditure (excluding commitments) during the financial year (SI.No.6+7) :
Rs. (1267448 (Year 1)+452677 (Year 2)+1,96670 (Year 3)+6006(Bank Interest))= Rs. 1922801

8) Actual expenditure (excluding commitments) incurred during the financial year (up to 31st March).
Rs. 985926(Year 1)+809968 (Year 2)+203806 (Year 3))= Rs. 1999700

9) Balance amount available at the end of the financial year. **- Rs. 76899**

10) Amount already committed, if any. **Rs. 42104 (Year 1) + Rs. 1066(Year 2) = Rs. 43170**

11) Amount to be carried forward to the next financial year (if applicable). Indicate the amount already committed with supporting documents. **NA**