International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 1
***Article Received:** 25 November 2022 **Revised:** 12 December 2022 **Accepted:** 05 January 2023*

Machine Learning based Segmentation for the Detection of Liver Disease

Parul Chhabra[a], Pradeep Kumar Bhatia[b]
[a]PhD Research Scholar, CSE Department, GJUS&T, Hisar, India
[b]Professor, CSE Department, GJUS&T, Hisar, India
[a]parul15march@gmail.com
[b]pkbhatia.gju@gmail.com

ABSTRACT- The support of Artificial intelligence (AI) can be used to update traditional healthcare services, and it can efficiently serve society. Using machine learning tools, the diagnosis process can be automated, and practitioners can process large scale clinical data to generate quick medical advisory for patients. In this paper, a segmentation based liver tumor detection scheme will be introduced, and its performance will be analyzed using different classifiers under the constraints of metric, i.e., accuracy, Recall, F1-Score, and Precision, etc.
.

Keywords- Machine Learning, Liver Tumor, Segmentation

I. INTRODUCTION

Disease recognition in medical imaging is quite complex, and specialists are required to examine it accurately. Improper observations may lead to an incorrect diagnosis. One of the most complex diseases is a Liver tumor, as shown in Figure 1.

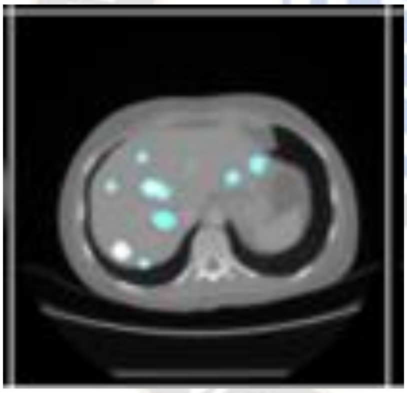

Figure: 1 Liver tumor

In this disease, Liver cells may grow abnormally, and it becomes very challenging to distinguish between healthy cells and affected cells. And error-prone interpretation directly affects the treatment plan.

There are numerous issues associated with the Liver tumor detection process, which are described below:

- Disease detection: The tumor may grow in several years, and its detection at early stages is a major challenge.
- Parameter Identification: In different patients, tumor attributes (location, size, and growth interval) may vary. So, its accurate detection is necessary for decision making.
- Disease Categorization: Tumor categorization is another major issue that may affect the diagnosis process.

Machine learning (ML) approach can be used for the automation of disease detection/diagnosis process, but the following are the limitations for the ML-based solutions as shown in figure 2:

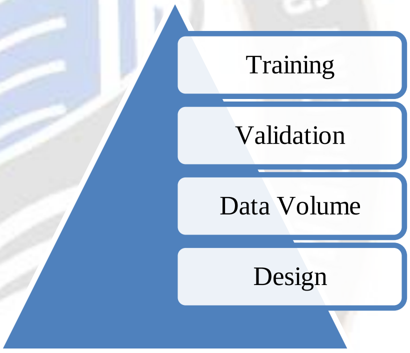

Figure: 2 Limitations of machine learning approaches

- Training dataset: A training dataset is required to build a training model that is used by ML schemes. It is prepared by a medical specialist, but its sample collection accuracy depends on the knowledge and experience in the relevant domain.
- Data Validation: Validation of the collected knowledge base is another major issue.
- Lack of Standard Solutions: For different diseases, there is a need to build training models as well as ML logic because there is no single remedy exists for all diseases.
- Dataset Volume: Training datasets may be quite large, so it may increase the infrastructure/processing cost for the implementation of ML solutions.
- Dataset Design: Various researchers are engaged in ML research in medical imaging. There is a need to define some sort of standards to design the input datasets [1-5].

Liver tumor detection is a complex process, and analysis of its growth over a certain period is very critical. To achieve this goal, this paper introduces a machine learning method using active contour based segmentation to analyze the tumor in sample input.

II. LITERATURE SURVEY

M. D. Samad et al. [12] have machine learning based methods that can predict survival accuracy using a limited set of input variables for echocardiography outcomes. The study found that traditional methods use Ejection Fraction and Comorbidities based prediction models, which are less accurate as compared to machine learning algorithms.

Y. Xue et al. [13] investigated different prediction models (Regression/Support Vector Machine/Random Tree based) for patient readmission, and analytical results indicate that the Functional Independence Measure method outperforms as compared to traditional methods. Training and validation were performed using existing clinical data, and accuracy and sensitivity were both adjusted by finding the optimal cutoff point of the receiver operating characteristic curve. Results state that it can reduce the overall treatment cost as well as it can also improve the quality of healthcare services.

A. Clim et al. [14] investigated the relation between chest sounds and level of hypertension and found that a prediction scheme based on Kullback-Leibler Divergence is more accurate and can enhance clinical decision support as compared to traditional methods.

C. R. Olsen et al. [15] investigated the role of machine learning algorithms in the diagnosis of heart diseases. Study indicates that these methods can assist the practitioners in diagnosis and can also be used to develop the prediction models as per the patient's classification. Large-scale disease datasets can also be analyzed by the integration of these algorithms with the BigData framework.

F. Y. Qin et al. [16] developed a predation framework for the health analysis of elderly patients. It builds the predicates by performing the feature classification. Finally, a dataset is used for training and validation purposes. Experimental results show that it outperforms in terms of prediction accuracy, as well it can also optimize the overall data processing complexity.

H. Yin et al. [17] developed a framework that collects data from various wearable sensors as well as from computed assisted medical systems. A machine learning based scheme is used to process and classify the patients as per disease categories. Output data is further used for decision making and diagnosis purposes. Experimental results show that it can improve treatment accuracy as well, and practitioners can utilize multiple datasets to improve healthcare services.

S. Anakal et al. [18] introduced a decision support system for the diagnosis of chronic lung diseases. It uses different machine learning schemes, i.e., Decision trees/Support vector machines/neural networks/ Classifier Ensembles, etc. Experimental results show that practitioners can redefine their treatment strategies as well as also utilize the system feedback to manage the drug level for patients. It can be integrated with a cloud platform to provide telemedicine support for remote areas.

H. R. Mansilla et al. [19] presented a decision support framework that can analyze the risk of infection after surgery, and practitioners can use the alternative treatment type to avoid these side effects. It uses the combination of a support vector machine and a decision tree to balance the accuracy level in results. Experimental outcomes show that estimation of infection risk can optimize the diagnosis strategies.

A. Yahyaoui et al. [20] developed a decision support framework for diabetic patients that uses a deep learning approach for disease prediction and diagnosis support. Experimental results show its performance in terms of prediction accuracy as compared to traditional approaches (Random Forest/Support vector machine). As per the results, its accuracy can be further enhanced by integrating deep feature extraction.

C. Comito et al. [21] developed an automated decision support framework that can assist the practitioners as per the available datasets that are built using different medical data resources (Lab test/patient health records). Experimental results show that by using deep learning, at an early stage, symptoms can be detected, and diagnosis plans can be suggested for identified diseases.

L. Zhao et al. [22] investigated the challenges associated with pharmaceutical research and drug development. The study found few facts (data source/ quality/ format/ validity/authenticity/data rate/ volume/ values) having direct influence over drug development cost. Large scale data related to drugs can be analyzed through the integration of machine learning/deep learning algorithms over big data platforms, thus may lead to the reduction of the overall R&D cost of drugs.

C. Réda et al. [23] surveyed to find out the association of the diseases with different drugs and its impact on the drug development process. Analysis shows that a complete knowledgebase of diseases can be acquired using machine learning algorithms, and it may reduce the research cost as well as human trials may be conducted at earlier stages, and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 1
***Article Received:** 25 November 2022 **Revised:** 12 December 2022 **Accepted:** 05 January 2023*

feedback from several successful trials can be utilized to refine the drug modeling process/dose/accuracy level, etc.

R. Ietswaart et al. [24] developed a random forest approach based model to find out the association of drug reactions among patients. A number of drugs were used for training purposes, and their performance was verified using different parameters (accuracy/correlation coefficient/recall curve/ precision, etc.). Outcomes show that large scale analysis of drug reaction associations can optimize the failure rate of drug trials, as well as machine learning provides a platform for random experiments, and no human/animal is required.

N. T. Issa et al. [25] explored the drug development issues related to cancer/tumor and found that using existing large scale cell datasets can be processed through machine learning algorithms, and drug repurposing strategies can be defined to diagnose this disease as well as drug development cost can be optimized. Outcomes show that drug trial feedback can be utilized to update the training datasets to maintain the accuracy of experiments.

M. Ali et al. [26] investigated several machine learning approaches that can extract cell related data from cancer patients and can prepare training datasets for detection and diagnosis purposes. The analysis states that feature extraction of cells and drug response to patient’s health both can be used to predict the response of cancer drugs as well, and these outcomes can optimize the drug development cost also.

L. Štěpánek et al. [27] investigated the role of machine learning in plastic surgery, and multiple facial expression dataset was used to perform Multivariate linear regression using R language. Experimental results indicate that higher accuracy for facial geometry can be achieved using neural networks as well as Bayesian naive classifiers/decision trees can be used to map the facial image to emotions.

T. J. Loftus et al. [28] explored the risk associated with the surgical wards where quick assessment of high risk patients is essential and error prone diagnosis and treatment recommendations may lead to failure of healthcare services. For such types of risks, real-time health analysis and electronic health records can be generated using wearing sensors, and medical data can be further processed using machine/deep learning schemes to recognize the symptoms at earlier stages and to achieve higher accuracy in diagnosis.

A. W. Schwartz et al. [29] studied the virtual reality based surgical operations and investigated its integration with machine learning schemes. Analysis found that a combination of both technologies can be utilized to develop a large scale knowledgebase to help the stakeholders.

L. Štěpánek et al. [30] analyzed facial feature extraction and their classification using a machine learning algorithm that can refine the outcomes of facial plastic surgery. Analytical data shows that geometry features, along with sufficient datasets as evidence, can be enforced to maintain the quality of facial attractiveness.

K. Merath et al. [31] analyzed the compilations associated with the different types of surgical operations (as discussed above) and developed a solution to predict the complications using decision tree models. Experimental results show its performance in terms of higher prediction accuracy and efficient risk analysis related to diagnosis. Its scope can be enhanced using an electronic health record system.

III. PROPOSED SCHEME

In this paper, a machine learning based segmentation approach is presented to identify the liver tumor disease and its growth in the input samples. First of all, it collects the samples and applies segmentation to extract the various features related to tumor location, size region, etc., and finally, it builds a training model. Later on, classifiers are used to classify the tumor in the test data. Flow chart 1 shows the various steps of the proposed scheme, and Figure 3 provides an overview of the training and testing process.

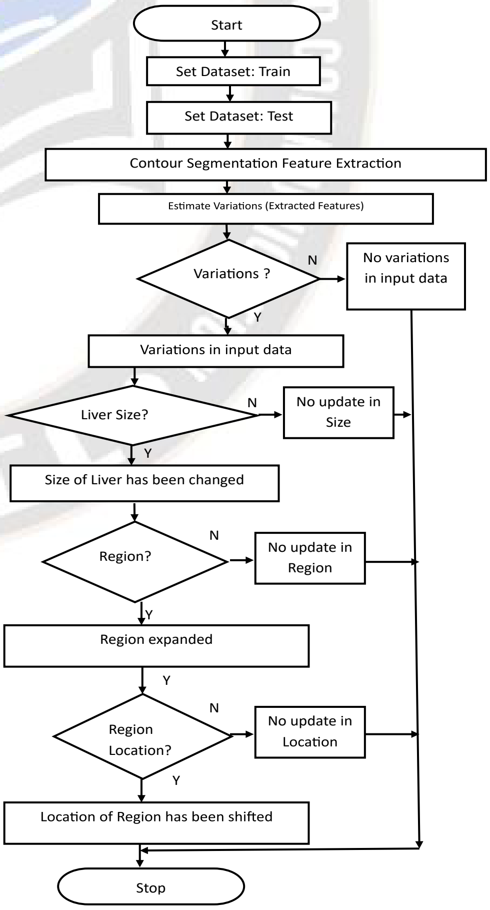

Figure 3 Overview of Training and Testing

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 1
***Article Received:** 25 November 2022 **Revised:** 12 December 2022 **Accepted:** 05 January 2023*

Figure 3 shows that segmentation is performed over the input dataset, and their features are extracted ((region/location/size)) w.r.t. each sample using Active Contour snake model that uses segmentation to mark the tumor area and finally, features are extracted in a given input, and a training model is created. It can be derived through the following equation:

$res = Lei(a_1,b_1)+Le_j|\nabla I(a_2,b_2)|^2+ \ldots + Le_n|\nabla I(a_n,b_n)$

Where (a,b) are coordinates in 2-D space
ei Line
EJ Edge
RES result
Variations in the values of *ei* and *ej* move the contour snake towards high intensity areas in the image, and finally, the spline is used to mark the identified features as:

$$V(s,t) = (a(s,t), b(s,t))$$

Where *v* Spline variable having a value between 0-1
S linear variable having a value between 0-1
T Time interval
The following are the outcomes of an active contour snake:
Size of Liver Tumor: The size of a Liver tumor may vary, and it depends on its growth, as shown in Figure 4.

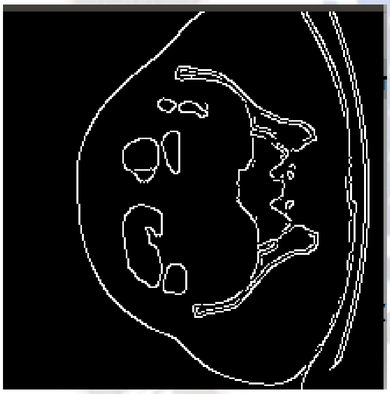

Figure 4 Size of Liver Tumor

Region of Liver Tumor: The tumor may affect a particular area in the Liver, as shown in Figure 5

Figure:5 Region of Liver Tumor

Tumor Location in the Liver: It may be anywhere in the Liver, as shown in Figure 6

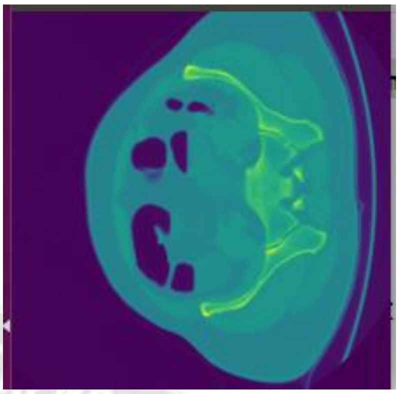

Figure: 6 Tumor Location

Table:1 Experimental Setup

Parameters	Configuration
Platform	Linux
Libraries	Python3.x, OpenCV-2
Dataset	LiTS liver tumor dataset: 130 samples [32]

Table 1 shows the experimental Setup for analysis.

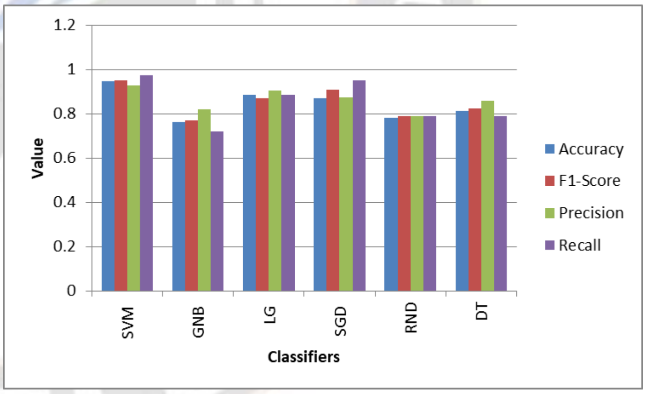

Figure 7: Performance of Classifiers

Figure 7 shows the performance of different classifiers, i.e., SVM, GNB, LG, SGD, RND DT, etc. It can be observed that SVM has the highest accuracy, followed by LG, SGD, and DT, and it is average for RND and lowest for GNB. In the case of the F1-score, it is highest for SVM, SGD classifies and average for LG and DT, and lowest for RND and GNB. SVM and LG both have the highest Precision value as compared to others. It is average for DT, SGD, and GNB, and it is lowest for RND. Recall value is higher for SVM and SGD average for GNB, LG, and DT, and it is lowest for RND.

IV. CONCLUSION

In this paper, a machine learning based segmentation approach was introduced to detect the liver tumor. Its performance was analyzed using different classifiers, i.e., SVM, GNB, LG, SGD, RND DT, etc., under the constraints of accuracy, F1-Score, Recall, and Precision. Experimental results show that SVM, SGD, and LG have higher performance. DT and GNB both are average performers, and the RND classifier could not deliver the performance at a significant level, and so on. The scope of the current work deals with machine learning based segmentation for Liver disease only. In the future, it will be extended to classify other critical diseases, i.e., Liver tumors, lung cancer, heart diseases, etc.

REFERENCES

[1] A. Smiti, "When machine learning meets medical world: Current status and future challenges," Computer Science ReviewVolume 37 August 2020, pp.1-7.
[2] T. Alsuliman, D. Humaidan, L. Sliman, "Machine learning and artificial intelligence in the service of medicine: Necessity or potentiality?" Current Research in Translational Medicine, Elsevier-2020, pp.1-7.
[3] S. Lalmuanawma, J. Hussain, L. Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review", Chaos, Solitons & Fractals, Vol.139 Elsvier-2020, pp.1-6.
[4] D. B. Israel, W. B. Jacobs, S. Casha, S. Lang, D. W. Cadotte, "The impact of machine learning on patient care: A systematic review," Artificial Intelligence in Medicine, Vol. 103, Elsevier-2020, In press.
[5] J. Waring, C. Lindvall, R. "UmetonAutomated machine learning: Review of the state-of-the-art and opportunities for healthcare," Artificial Intelligence in Medicine, Vol.104, Elsevier-2020, pp.1-12.
[6] https://www.radiologyinfo.org
[7] http://www.radtechonduty.com
[8] https://healthcare-in-europe.com
[9] https://radiopaedia.org
[10] myheart.net
[11] https://health-innovations.org
[12] M. D. Samad, A. Ulloa, G. J. Wehner, L. Jing, B.K. Fornwalt, "Predicting Survival From Large Echocardiography and Electronic Health Record Datasets: Optimization With Machine Learning," JACC: Cardiovascular Imaging, Vol.12 (4), Elsevier-2019, pp.681-689.
[13] Y. Xue, H. Liang, J. Norbury, R. Gillis, B. Killingworth, "Predicting the risk of acute care readmissions among rehabilitation inpatients: A machine learning approach," Journal of Biomedical Informatics, Vol.86, Elsevier-2018, pp.143-148.
[14] A. Clim, R. D. Zota, G. TinicĂ, "The Kullback-Leibler Divergence Used in Machine Learning Algorithms for Health Care Applications and Hypertension Prediction: A Literature Review," Procedia Computer Science, Vol.141, Elsevier-2018, pp.448-453.
[15] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, P. A. Patel, "Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure," American Heart Journal, Elsevier-2020, In press.
[16] F. Y. Qin, Z.Q. Lv, D.N. Wang, B. Hu, C. Wu, "Health status prediction for the elderly based on machine learning," Archives of Gerontology and Geriatrics, Vol.90, Elsevier-2020, In Press.
[17] H. Yin, N. K. Jha, "A Health Decision Support System for Disease Diagnosis Based on Wearable Medical Sensors and Machine Learning Ensembles," in IEEE Transactions on Multi-Scale Computing Systems, Vol.3 (4), IEEE-2017, pp.228-241.
[18] S. Anakal, P. Sandhya, "Clinical decision support system for chronic obstructive pulmonary disease using machine learning techniques, "International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), IEEE-2017, pp. 1-5.
[19] H. R. Mansilla, G. A. Solano, M. C. M. Lapitan, "deSSIde: A Clinical Decision-Support Tool for Surgical Site Infection Prediction," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), IEEE-2020, pp. 367-372.
[20] A. Yahyaoui, A. Jamil, J. Rasheed, M. Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques," 1st International Informatics and Software Engineering Conference (UBMYK), IEEE-2019, pp. 1-4.
[21] C. Comito, A. Forestiero, G. Papuzzo, "A clinical decision support framework for automatic disease diagnoses," 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE-2019, pp.933-936.
[22] L. Zhao, H. L. Ciallella, L. M. Aleksunes, H. Zhu, "Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling Drug Discovery Today," Elsevier-2020, In press.
[23] C. Réda, E. Kaufmann, A. D.Duriez, "Machine learning applications in drug development," Computational and Structural Biotechnology Journal, Vol.18, Elsevier-2020, pp.241-252.
[24] R. Ietswaart, S. Arat, A. X. Chen, S. Farahmand, L. Urban, "Machine learning guided association of adverse drug reactions with in vitro target-based pharmacology," EBioMedicine, Vol.57, Elsevier-2020, pp.1-15.
[25] N. T. Issa, V. Stathias, S. Schürer, S. Dakshanamurthy, "Machine and deep learning approaches for cancer drug repurposing," Seminars in Cancer Biology, Elsevier-2020, In Press.
[26] M. Ali, T. Aittokallio, "Machine learning and feature selection for drug response prediction in precision oncology applications, Biophys Rev 11, Springer-2019, pp.31–39.
[27] L. Štěpánek, J. Měšt'ák, P. Kasal, "Machine-learning at the service of plastic surgery: a case study evaluating facial attractiveness and emotions using R language," Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE-2019, pp.107-112.
[28] T. J. Loftus, P. J. Tighe, A. C. Filiberto, J. Balch, A. Bihorac, "Opportunities for machine learning to improve

surgical ward safety," The American Journal of Surgery, Elsevier-2020, In press.
[29] A. W. Schwartz, V. Bissonnette, N. Mirchi, N. Ponnudurai, R. F. D. Maestro, "Artificial Intelligence in Medical Education: Best Practices Using Machine Learning to Assess Surgical Expertise in Virtual Reality Simulation," Journal of Surgical Education, Vol.76 (6), Elsevier-2019, pp.1681-1690.
[30] L. Štěpánek, P. Kasal, J. Mestak, "Evaluation of facial attractiveness for purposes of plastic surgery using machine-learning methods and image analysis," 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services, IEEE-2018, pp.1-6.
[31] K. Merath, J.M.Hyer, R. Mehta, A. Farooq, F. Bagante, K. Sahara, D. I. Tsilimigras, E. Beal, A. Z. Paredes, L. Wu., A. Ejaz, T. M. Pawlik, "Use of Machine Learning for Prediction of Patient Risk of Postoperative Complications After Liver, Pancreatic, and Colorectal Surgery", Journal of Gastrointestinal Surgery, Vol.24, Springer-2019, pp.1843–1851.
[32] www.lits-challenge.com