



ML BASED DIAGNOSTIC PARADIGM IN HEPATOCELLULAR CARCINOMA

Sushmitha G P Harithas, Dr Keerthi Kumar H M,

Research Scholar, Associate Professor,
Artificial Intelligence and Data Science,
Malnad College of Engineering, Hassan, Karnataka, India

Abstract : In developing countries, both viral and non-viral Hepatocellular Carcinoma (HCC) is on the rise. Late detection is a major issue, contributing to high mortality rates. While traditional diagnostic methods for HCC are the standard, they have limitations, leading to delays in confirming cancer progression. Artificial intelligence (AI) tools offer promise in complementing these traditional approaches by enabling earlier and more certain diagnoses. This review contrasts traditional HCC diagnostic methods with AI, particularly Machine Learning and Deep Learning, and explores their application in histopathology, radiology, biomarker analysis, and Electronic Health Records (EHRs)-based diagnosis of HCC.

IndexTerms - Cancer, Artificial intelligence, liver cancer, Machine learning, deep learning

INTRODUCTION

Approximately, 80% of the hepatocellular carcinoma is caused by chronic viral infections including Hepatitis C Virus (HCV) and Hepatitis B Virus (HBV). World Health Organization (WHO) aims to reduce the CHV incidence by 90% and mortality by 65% before 2030 according to the strategy on viral hepatitis prepared in 2016 [27]. Globally, two million deaths are caused due to hepatocellular carcinoma (HCC). Particularly in underdeveloped nations, the hepatitis B and hepatitis C viruses are the main causes of chronic liver disease, which can progress to HCC. As early diagnosis of liver inflammation causing fibrosis and cirrhosis is crucial for better treatment outcomes.

Traditional statistical models, like conventional regression and Multiple Logistic Regression (LR), often fall short in capturing the non-linear interactions among predictor variables in large datasets, particularly in analyzing fibrotic liver conditions. These limitations hinder the ability to elucidate the complex relationships defining the severity of inflammation development. However, the emergence of machine learning-assisted methods has provided new opportunities to address these challenges and enhance the diagnosis, staging, and prognosis of various diseases, including hepatocellular carcinoma.

Incorporating machine learning techniques into medical research offers the potential to overcome the shortcomings of traditional statistical models by effectively analyzing big data. By leveraging machine learning algorithms, clinicians can better capture the intricate relationships within large datasets, leading to improved diagnostic accuracy and prognostic capabilities. Furthermore, the adept analytical capabilities of machine learning algorithms enable the identification of complex patterns and predictive models for various medical conditions, including breast cancer and postprandial glycaemic responses. Continuous evidence supports the implementation of machine learning approaches to enhance prediction systems and improve patient outcomes across a wide range of diseases.

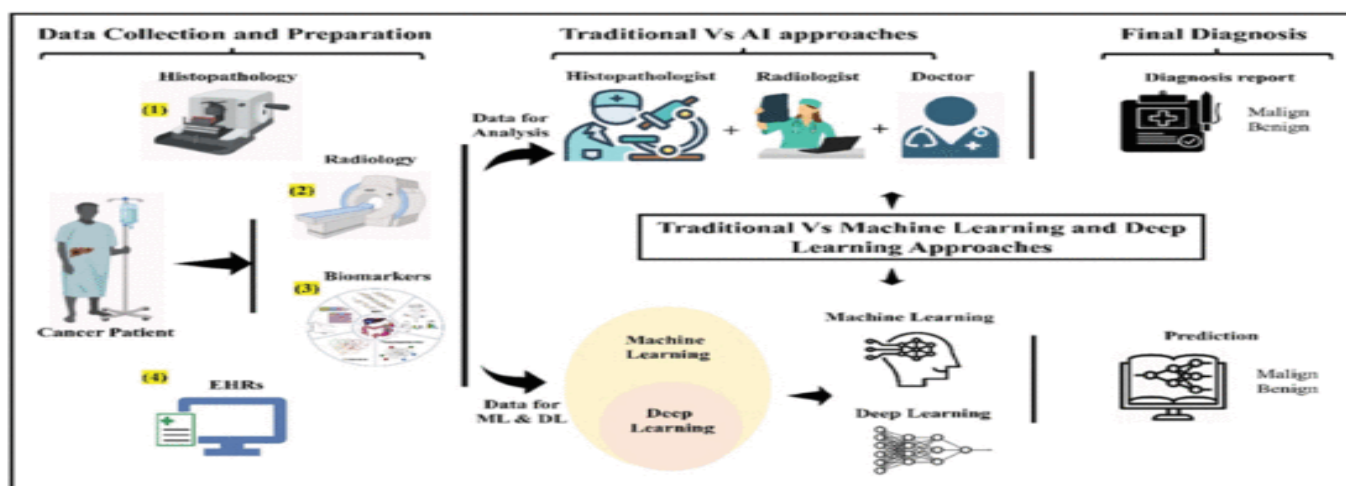


Figure1: Approaches to obtain the data related to cancer patients

NEED OF THE STUDY.

The objective of this paper is to examine the limitations of traditional statistical models in diagnosing and prognosticating hepatocellular carcinoma (HCC) and other diseases, particularly in analyzing fibrotic liver conditions. By highlighting the challenges associated with conventional regression and Multiple Logistic Regression (LR) models in capturing non-linear interactions within large datasets, the paper aims to underscore the need for alternative approaches. Specifically, the focus is on exploring the potential of machine learning-assisted methods to address these limitations and enhance the accuracy of diagnosis, staging, and prognosis for HCC and other diseases. Through a comprehensive review of existing research and evidence, the paper seeks to demonstrate the promise of machine learning algorithms in revolutionizing medical diagnostics and improving patient outcomes.

LITERATURE SURVEY

Conventional methods for hepatocellular carcinoma (HCC) diagnosis, such as histopathology, radiological analysis, biomarker assessment, and electronic health records, are time-consuming and subject to interpretational errors and biases, impacting the accuracy of diagnosis and prognosis [1], [2], [3]. These limitations affect patient management and survival outcomes, highlighting the need for more efficient diagnostic approaches.

Histopathology has been pivotal in managing liver disorders, but its invasive nature and limitations in non-invasive techniques for HCC diagnosis necessitate histological examination, particularly for tumors with atypical imaging features or to rule out other liver pathologies [4]. However, histopathological analysis also has drawbacks, including the need for substantial tissue samples and susceptibility to personal biases among pathologists, leading to inconsistencies in characterization [5], [6]. To address these challenges, researchers have turned to artificial intelligence (AI) methods to enhance the accuracy of liver tumor diagnosis.

RESEARCH METHODOLOGY

Machine learning (ML) and deep learning (DL) are integral components of artificial intelligence (AI), offering powerful tools for analyzing complex datasets in various fields, including medicine. ML employs statistical, probabilistic, and optimization techniques to enable algorithms to learn intricate patterns from large, noisy datasets, making it particularly suitable for medical applications. This capability allows ML and DL algorithms to search through multidimensional data spaces and derive meaningful insights, facilitating tasks such as cancer diagnosis and detection.

The ML learning process comprises two primary phases: approximation of non-classified dependencies from provided data and the utilization of these approximated dependencies to predict novel outcomes. Supervised learning involves training ML models with labeled data, enabling them to classify or regress future inputs, while unsupervised learning relies on unlabeled data to uncover underlying patterns through clustering or association techniques. The recent advancements in cloud computing and graphics processing unit (GPU) technology have significantly enhanced the application of ML in medical fields, including oncology.

In medical diagnostics, AI techniques are increasingly utilized in histopathology, radiology, biomarker-based diagnostics, and statistical modeling to improve accuracy and efficiency. This review discusses the benefits, challenges, and future prospects of employing AI in these areas, emphasizing its potential to revolutionize cancer diagnosis and treatment.

A. AI in histopathology:

Histopathology involves the examination and diagnosis of tissue diseases by analyzing tissue or cell samples under a microscope. Typically, biopsy tissues are processed overnight and then visualized using advanced imaging techniques. However, traditional histopathology methods have limitations, including time consumption and the need for highly experienced histopathologists, often with over a decade of expertise, to accurately identify diseased tissue. To address these challenges, machine learning (ML) and deep learning (DL) methods have emerged as promising tools for cancer diagnosis, offering potential solutions to enhance efficiency and accuracy.

Metrics:

1. Accuracy: Accuracy is a basic metric indicating the proportion of correct predictions made by the model, including true positives and true negatives. However, it can be misleading in imbalanced datasets where one class is significantly more prevalent.
2. Sensitivity: Sensitivity gauges the ratio of correctly identified positive cases (patients with HCC) by the AI model among all actual positive cases. It's essential in medical diagnostics to prevent missing patients with the disease.
3. Specificity: Specificity evaluates the ratio of correctly identified negative cases (healthy individuals or those without HCC) by the AI model among all actual negative cases. A high specificity indicates the model's ability to minimize false alarms.
4. Precision: This metric assesses the accuracy of positive identifications made by the AI model, indicating the proportion of correctly identified cases out of all positive identifications. Specifically in the context of HCC, it indicates the percentage of patients identified by the model as having HCC who actually have the disease.
5. Recall: This is synonymous with sensitivity and refers to the model's capability to detect all relevant instances, specifically, all patients with HCC.
6. F1 score: The F1 score, a harmonic mean of precision and recall, offers a balanced evaluation, especially valuable in scenarios with imbalanced class distribution.
7. Area Under the Receiver Operating Characteristic (ROC) Curve (AUC-ROC): The Receiver Operating Characteristic (ROC) curve evaluates classification performance across different threshold settings, while the Area Under the Curve (AUC) indicates the model's ability to differentiate between classes. Conversely, the Area Under the Precision-Recall Curve (AUC-PR) is employed when there's a substantial class imbalance, offering insights into model performance.
8. Confusion metrics: The confusion matrix, though not a standalone metric, provides a structured view of algorithm performance by tabulating true positives, false positives, true negatives, and false negatives. This framework facilitates cell detection, counting, and the identification of cancerous cell morphology, enabling early diagnosis. Moreover, ML techniques can classify tissue images from microscopy as either healthy or cancerous, broadening their diagnostic applications.

Table1: application of AI (Machine learning and deep learning) for diagnosis of Hepatocellular carcinoma using AI in histopathology, radiology, biomarker based, and EHRs based diagnosis

SL. NO.	Method	Datasets	Evaluation criteria	Reference
1. AI in histopathology				
1	DL based multi-class classification	TCGA-LIHC dataset KMC dataset	Accuracy and F1 score	[7]
2	DL based architecture for segmentation	H&E stained histopathology images and KMC liver	Dice and Jaccard Index, F1 score	[8]
3	DL based classification on global labels	TCGA	Precision, recall, F1 score	[9]
4	DL attention mechanism	Liver cancer histopathological images were obtained from research dataset	Accuracy, F1 score, precision, recall	[10]
5	Convolutional auto-encoder	PAIP challenge 2019- MICCAI 2019	Aggregated Jaccard index, F1 score	[11]
2. AI in radiology				
1	DL multi network model	MRI Images	AUC, F1 score	[12]
2	Deep CNN with U-net architecture	MRI images	Accuracy, precision, dice similarity coefficient	[13]
3	Extreme learning machine(ELM)	CT images	ROC curve, sensitivity	[14]
4	Image processing algorithm-fuzzy linguistic constant(FLC)	CT images	Accuracy	[15]
5	SVM and ELM	Ultrasonic tissue data	ROC-AUC	[16]
3. Biomarker based disease prediction				
1	Automated machine learning for diagnostic	TG-GATES	Precision, recall	[17]
2	ML in TCGA	TCGA, GEO	Log2FC, P-value, FDR	[18]
3	efDNA scoring system	Korea biobank network	Accuracy	[19]
4	Multiple feature selection-Alkaike information criterion	TCGA	-	[20]
5	Analysis of gene co-expression network	GEO	Accuracy, sensitivity and specificity	[21]
4. EHRs based diagnosis				
1	Deep learning model	A total of 52 983 samples (98% men)	Specificity and sensitivity	[22]

2	Ontological features based EHRs	112 abdominal computed tomography imaging examination reports	Sensitivity and accuracy	[23]
3	ML and ridge regression scores	N=86,804; 6,821 HCC(dataset) N=37,202; 2,875 HCC (validation)	AUROC	[24]
4	DL model for minimal feature extraction	47,945 individuals containing 9553 patients with HCC	AUROC, sensitivity and specificity	[25]
5	Algorithmic validation for an observational study	Patients = 339 HCCs	Accuracy, sensitivity and specificity	[26]

B. AI in radiology

Radiology, akin to histopathology, employs medical imaging to diagnose diseases in humans and animals, utilizing techniques such as X-ray, MRI, ultrasound, CT scan, and PET scan. However, traditional radiology faces limitations, including reproducibility issues, particularly concerning the textural features of hepatocellular carcinoma, which impairs radiomic signatures. Additionally, the absence of standardized protocols for texture feature quantification and model development hinders result generalizability. Moreover, limited data availability and restricted data sharing between institutions due to legal concerns further complicate matters. Fortunately, advancements in computational power and the remarkable progress of DL and ML frameworks have addressed many of these challenges, facilitating the integration of radiology/radiomics with ML and DL techniques, although data scarcity and proper labeling remain significant hurdles.

C. Biomarker based disease prediction

Methods like integrating multi-omics analysis and employing ML and DL have emerged as vital tools for biomarker screening and diagnosing HCC. Once trained, ML and DL algorithms can effectively diagnose the disease from patient samples. Various algorithms trained on diverse biomarker data types share the common objective of early HCC diagnosis, crucial for effective treatment. Typically, early tumor detection in HCC involves monitoring high-risk groups using abdominal ultrasonography, often supplemented with serum α -fetoprotein (AFP) analysis.

D. EHRs based diagnosis

Electronic health records (EHR) offer precise, current, and thorough patient information, facilitating coordinated and efficient care. Sharing EHRs streamlines disease diagnosis, minimizes errors, and accelerates delivery of safe and effective care by ensuring access to reliable patient health data. A structured EHR system plays a pivotal role in supporting crucial clinical decisions for disease diagnosis.

RESULTS AND DISCUSSION

EHRs present challenges due to their raw, noisy, and heterogeneous nature, which can hinder ML model performance without extensive pre-processing efforts. Transforming EHRs for ML requires significant data cleaning, normalization, and feature extraction, alongside addressing privacy and sharing concerns. ML and DL models face hurdles in healthcare due to high dimensionality, variability, and the "black box" nature of DL models, impacting trust and interpretability. Generalizability to new data sources and seamless integration into clinical workflows are also critical challenges.

To address these challenges, advancements in data processing, sophisticated algorithms, and federated learning can improve EHR quality and accessibility for ML analysis. Explainable AI models can enhance transparency and trust, while collaboration among stakeholders ensures ML tools meet clinical needs effectively.

CONCLUSION

Integrating AI into oncological diagnostics, particularly for HCC, enhances traditional methods through advanced ML and DL techniques, bolstered by increased computational power. This synergy allows for nuanced analysis of image-based and biomarker datasets, crucial for early HCC detection, while AI supplements existing clinical methodologies. AI not only promises improved diagnostic accuracy but also aids prognostic assessment, guiding personalized treatment approaches. Future efforts should prioritize developing interpretable AI models to foster trust, ensuring seamless clinical integration, conducting rigorous validation studies, and addressing ethical and regulatory concerns. Collaboration among data scientists, clinicians, ethicists, and policymakers is essential to fully harness AI's potential in transforming HCC diagnostics and patient care.

REFERENCES

- [1] H. B. El-Serag, J. A. Marrero, L. Rudolph and K. R. Reddy, "Diagnosis and treatment of hepatocellular carcinoma", *Gastroenterology*, vol. 134, no. 6, pp. 1752-1763, 2008.
- [2] M. Sherman, "Hepatocellular carcinoma: Epidemiology surveillance and diagnosis", *Seminars Liver Disease*, vol. 30, no. 1, pp. 3-16, Feb. 2010.
- [3] J. Neuberger and O. Cain, "The need for alternatives to liver biopsies: Non-invasive analytics and diagnostics", *Hepatic Med. Evidence Res.*, vol. 13, pp. 59-69, Jun. 2021.
- [4] V. Granata, R. Grassi, R. Fusco, A. Belli, C. Cutolo, S. Pradella, et al., "Diagnostic evaluation and ablation treatments assessment in hepatocellular carcinoma", *Infectious Agents Cancer*, vol. 16, no. 1, pp. 1-22, Dec. 2021.

- [5] P. Angulo, D. E. Kleiner, S. Dam-Larsen, L. A. Adams, E. S. Bjornsson, P. Charatcharoenwittaya, et al., "Liver fibrosis but no other histologic features is associated with long-term outcomes of patients with nonalcoholic fatty liver disease", *Gastroenterology*, vol. 149, no. 2, pp. 389-397, Aug. 2015.
- [6] E. C. Stack, C. Wang, K. A. Roman and C. C. Hoyt, "Multiplexed immunohistochemistry imaging and quantitation: A review with an assessment of tyramide signal amplification multispectral imaging and multiplex analysis", *Methods*, vol. 70, no. 1, pp. 46-58, Nov. 2014.
- [7] A. A. Aatresh, K. Alabhya, S. Lal, J. Kini and P. P. Saxena, "LiverNet: Efficient and robust deep learning model for automatic diagnosis of sub-types of liver hepatocellular carcinoma cancer from H&E stained liver histopathology images", *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, no. 9, pp. 1549-1563, Sep. 2021.
- [8] M. M. Abdelsamea, U. Zidan, Z. Senousy, M. M. Gaber, E. Rakha and M. Ilyas, "A survey on artificial intelligence in histopathology image analysis", *WIREs Data Mining Knowl. Discovery*, vol. 12, no. 6, pp. e1474, Nov. 2022.
- [9] C. Sun, A. Xu, D. Liu, Z. Xiong, F. Zhao and W. Ding, "Deep learning-based classification of liver cancer histopathology images using only global labels", *IEEE J. Biomed. Health Informat.*, vol. 24, no. 6, pp. 1643-1651, Jun. 2020.
- [10] C. Chen, C. Chen, M. Ma, X. Ma, X. Lv, X. Dong, et al., "Classification of multi-differentiated liver cancer pathological images based on deep learning attention mechanism", *BMC Med. Informat. Decis. Making*, vol. 22, no. 1, pp. 1-13, Dec. 2022.
- [11] M. Roy, J. Kong, S. Kashyap, V. P. Pastore, F. Wang, K. C. L. Wong, et al., "Convolutional autoencoder based model HistoCAE for segmentation of viable tumour regions in liver whole-slide images", *Sci. Rep.*, vol. 11, no. 1, pp. 1-10, Jan. 2021.
- [12] T. He, J. N. Fong, L. W. Moore, C. F. Ezeana, D. Victor, M. Divatia, et al., "An imageomics and multi-network based deep learning model for risk assessment of liver transplantation for hepatocellular cancer", *Computerized Med. Imag. Graph.*, vol. 89, Apr. 2021.
- [13] S. M. Kamrul Hasan and C. A. Linte, "A modified U-Net convolutional network featuring a nearest-neighbor Resampling-based elastic-transformation for brain tissue characterization and segmentation", *Proc. IEEE Western New York Image Signal Process. Workshop*, pp. 1-5, Oct. 2018.
- [14] W. Huang, N. Li, Z. Lin, G.-B. Huang, W. Zong, J. Zhou, et al., "Liver tumour detection and segmentation using kernel-based extreme learning machine", *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 3662-3665, Jul. 2013.
- [15] A. A. Khan and G. B. Narejo, "Analysis of abdominal computed tomography images for automatic liver cancer diagnosis using image processing algorithm", *Current Med. Imag. Formerly Current Med. Imag. Rev.*, vol. 15, no. 10, pp. 972-982, Nov. 2019.
- [16] M. Biswas, V. Kuppli, D. R. Edla, H. S. Suri, L. Saba, R. T. Marinho, et al., "Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm", *Comput. Methods Programs Biomed.*, vol. 155, pp. 165-177, Mar. 2018.
- [17] M. Puri, "Automated machine learning diagnostic support system as a computational biomarker for detecting drug-induced liver injury patterns in whole slide liver pathology images", *ASSAY Drug Develop. Technol.*, vol. 18, no. 1, pp. 1-10, Jan. 2020.
- [18] J. Lv, J. Wang, X. Shang, F. Liu and S. Guo, "Survival prediction in patients with colon adenocarcinoma via multiomics data integration using a deep learning algorithm", *Biosci. Rep.*, vol. 40, no. 12, pp. 28-40, Dec. 2020.
- [19] T. Lee, P. A. Rawding, J. Bu, S. Hyun, W. Rou, H. Jeon, et al., "Machine-Learning-Based clinical biomarker using cell-free DNA for hepatocellular carcinoma (HCC)", *Cancers*, vol. 14, no. 9, pp. 2061, Apr. 2022.
- [20] Z. Zhang and Z.-P. Liu, "Robust biomarker discovery for hepatocellular carcinoma from high-throughput data by multiple feature selection methods", *BMC Med. Genomics*, vol. 14, no. S1, pp. 1-12, Aug. 2021.
- [21] H. Tian, D. Guan and J. Li, "Identifying osteosarcoma metastasis associated genes by weighted gene co-expression network analysis (WGCNA)", *Medicine*, vol. 97, no. 24, 2018.
- [22] G. N. Ioannou, W. Tang, L. A. Beste, M. A. Tincopa, G. L. Su, T. Van, et al., "Assessment of a deep learning model to predict hepatocellular carcinoma in patients with hepatitis C cirrhosis", *JAMA Netw. Open*, vol. 3, no. 9, Sep. 2020.
- [23] L. W. C. Chan, S. C. C. Wong, C. C. Chiau, T.-M. Chan, L. Tao, J. Feng, et al., "Association patterns of ontological features signify electronic health records in liver cancer", *J. Healthcare Eng.*, vol. 2017, pp. 1-9, May 2017.
- [24] G. L.-H. Wong, V. W.-K. Hui, Q. Tan, J. Xu, H. W. Lee, T. C.-F. Yip, et al., "Novel machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis", *JHEP Rep.*, vol. 4, no. 3, Mar. 2022.
- [25] C.-W. Liang, H.-C. Yang, M. M. Islam, P. A. A. Nguyen, Y.-T. Feng, Z. Y. Hou, et al., "Predicting hepatocellular carcinoma with minimal features from electronic health records: Development of a deep learning model", *JMIR Cancer*, vol. 7, no. 4, Oct. 2021.
- [26] R. J. Driver, V. Balachandrakumar, A. Burton, J. Shearer, A. Downing, T. Cross, et al., "Validation of an algorithm using inpatient electronic health records to determine the presence and severity of cirrhosis in patients with hepatocellular carcinoma in England: An observational study", *BMJ Open*, vol. 9, no. 7, Jul. 2019.
- [27] J. D. Yang, P. Hainaut, G. J. Gores, A. Amadou, A. Plymoth and L. R. Roberts, "A global view of hepatocellular carcinoma: Trends risk prevention and management", *Nature Rev. Gastroenterol. Hepatol.*, vol. 16, no. 10, pp. 589-604, Oct. 2019.