JURNAL ILMIAH COMPUTER SCIENCE (JICS)
E-ISSN 3026-7145 P-ISSN 3030-9840
Volume 3, Nomor 2, January 2025, Page 120-134
DOI: https://doi.org/10.58602/jics.v3i2.48

Hepatocellular Carcinoma Prediction in HCV Patients using Machine Learning and Deep Learning Techniques

Fiza Saeed[1], Ashish Shiwlani[2*], Muhammad Umar[3], Zeib Jahangir[4], Anoosha Tahir[5], Sheena Shiwlani[6]
University of Texas, Arlington, (Texas-USA)
[2,3]Illinois Institute of Technology, Chicago, (Illinois-USA)
[4]William Jessup University, San Jose, (California-USA)
[5]Buch International Hospital, Multan, (Pakistan)
[6]Mount Sinai Hospital, NYC, (New York-USA)
[1]fizasaeed@mavs.uta.edu, [2*]ashiwlani@hawk.iit.edu, [3]mumar1@hawk.iit.edu,
[4]zeib.jahangir@jessup.edu, [5]anooshatahir@buchhospital.com, [6]Sheena.shiwlani@mssm.edu

Keywords:
Hepatocellular Carcinoma (HCC), Hepatitis C Virus (HCV), Alkaline Phosphate (ALP), Deep Learning (DL), Sustained Virological Response (SVR)

Abstract

Hepatitis C virus is the root cause of 78% of hepato-cellular carcinoma cases. Hepatocellular carcinoma (HCC) represents one of the primary causes of liver cancer mortality and incidence. Clinical prediction of HCC in patients suffering with hepatitis C virus infection (HCV) is challenging due to the diagnostic gold standard, liver biopsy, which is an invasive technique with several limitations. Artificial intelligence (AI) technology is being used in clinical research at a larger rate in recent years, and the field of HCC diagnosis is no exception. Several advanced and light-weight machine learning algorithms combined with less invasive blood tests have promising diagnostic potential to diagnose HCC from HCV. Deep learning algorithms are regarded as best methods for handling and processing complex, unstructured and raw data from various modalities, ranging from routine clinical variables i.e., from EMRs and laboratories to high-resolution medical images. This paper offers a thorough analysis of the most current research that has used machine learning and deep learning to diagnose, prognosticate, treat, and predict HCC risk in patients suffering with HCV.

1.INTRODUCING

Worldwide, the consequences of chronic liver disease (CLD) account for about 2 million deaths annually [1]. Chronic alcohol use, immune system abnormalities, and infections of liver with hepatitis B and C viruses are the roots of liver damage [2]. Liver damage induces a wound-healing response that can lead to cirrhosis, hepatic fibrosis (HF), and ultimately, liver failure or hepatic cancer [3]. Hepatitis C virus (HCV) infection is the most common reason of chronic liver inflammation and a common predisposing factor for the development of hepatocellular carcinoma (HCC) [4]. A malignant tumor of the liver is called HCC and it is the fifth most common cancer worldwide [5]. Also, it is the third most common cause of cancer-related deaths in the whole world [6]. The first worldwide health domain strategy on viral hepatitis was released by the World Health Organization (WHO) in June 2016 [7]. Its objectives are to lessen the incidence and mortality of CVH by 90% and 65%, respectively, by 2030 [1], [7].

Biannual ultrasonography examinations (US) are a cost-effective way to monitor hepatocellular carcinoma (HCC) and have been linked to improved survival rates in hepatitis C patients [8]. However,

Ashish Shiwlani: *Corresponding Author

JURNAL ILMIAH COMPUTER SCIENCE (JICS)
E-ISSN 3026-7145 P-ISSN 3030-9840
Volume 3, Nomor 2, January 2025, Page 120-134
DOI: https://doi.org/10.58602/jics.v3i2.48

it is still challenging to determine the precise risk at the individual level because of patient heterogeneity [9]. Moreover, a significant portion of patients enrolled in surveillance programs continue to receive an advanced HCC diagnosis, primarily due to the US's low sensitivity in identifying HCC at an extremely early stage (less than 2 cm) [8], [9]. Although early HCC detection could be significantly improved by imaging techniques like CT and MRI, the implementation of such expensive screening programs might not be cost-effective in some patient populations with hepatitis due to their comparatively low incidence of HCC annually [10]. Repeat examinations are frequently required [11]. Alpha-fetoprotein [AFP] level, presence of fibrosis, viral factors (serum HCV level), and demographic factors (gender, age) have all been found to be related with an elevated risk of HCC in cross-sectional studies [12]. However, only some people are involved in the majority of these studies [13].

The liver biopsy is the only way to confirm the clinical measurement of HCC [14]. Still, it comes with a number of drawbacks and major complications (e.g., muscular discomfort, bleeding, contamination, damage to adjacent organs, diagnostic inaccuracies, and inter- and intra-observer variability) [15]. In the clinical context, a less intrusive and more reliable method of determining the severity and progression of HCC would be extremely beneficial [16]. Thus far, basic scoring systems have been created utilizing various configurations of typical clinical characteristics without considering the presence of viral eradication or population heterogeneity [16], [17]. Machine learning algorithms can help in the prediction of HCC by identification of risk factors in the patients suffering from other chronic diseases [18].

A comprehensive tool for model development that has emerged in recent years is machine learning, which utilizes maximum data for reducing biasness and permits automatic selection of predicting variables [19]. In order to determine the risk levels of HCC in patients suffering with HCV, novel clinical, radiological, and pathological features-based prediction models using machine learning algorithms can be developed [20]. These models may be used to improve clinical evaluation and risk classification of HCC in patients with hepatitis C virus infection by integrating them into computer-based management systems [21]. One of the techniques of machine learning is the decision tree-based approach [22]. This data-driven mathematical model makes predictions or decisions that are useful in recognizing greater interactions between features that may have escaped the notice of traditional statistical methods [22].

Similarly, deep learning algorithms can be applied to predict HCC by using AI and simple statistical analysis to identify and predict patterns in large data samples [23]. The application of deep learning algorithms to many medical disciplines, including hepatology, has grown rapidly in recent years [24], [25]. Deep learning technology has made the "AI revolution" of the last ten years possible [25]. Deep learning algorithms are capable of processing a wide range of available HCV patient data, incorporating extensive data from multi-omics research and digital high-definition images, i.e., US, CT, and MRI from different radiologic and histopathologic studies, and structured numeric data like laboratory values, i.e., AFP, serum biomarkers, and vital signs [26].

This is the first review on the use of ML and DL techniques to diagnose or predict the risk of HCC in patients suffering from HCV. So, it can be a complete guide for researchers, scientists, and doctors who want to conduct studies, reviews, meta-analyses, and trials regarding the risk factors for hepatocellular carcinoma (HCC) in patients with hepatitis C virus (HCV). It consolidates current methodologies and identifies the strengths and limitations of various predictive models. The study outlines critical imaging modalities and clinical factors necessary for accurate HCC prediction, highlighting gaps in existing research. However, only one review has been published on HCC prediction using deep learning [27]. This research seeks to address the following inquiries:

RQ1: What is the current scope of AI in HCC prediction from HCV?

RQ2: What are the most used ML and Neural Networks-based algorithms being applied in the prediction of HCC in patients already suffering from HCV?

RQ3: Which factors and imaging modalities must be considered while predicting HCC in patients with HCV?

Ashish Shiwlani: *Corresponding Author

JURNAL ILMIAH COMPUTER SCIENCE (JICS)
E-ISSN 3026-7145 P-ISSN 3030-9840
Volume 3, Nomor 2, January 2025, Page 120-134
DOI: https://doi.org/10.58602/jics.v3i2.48

2.METHODOLGY

The PRISMA statement for reporting on review article studies was used in the preparation of this research [28]. We conducted this study in three stages:

A. Database Selection

Data for this study was acquired from the PubMed and Web of Science (WOS) databases. WOS is a comprehensive library of over 10000 significant scholarly publications. This database is among the most comprehensive and longstanding citation index collections, encompassing a broad spectrum of fields including natural sciences, engineering technology, biomedicine, arts and humanities, along with numerous other subjects [29].

B. Search Strategy

For the PubMed and WOS to locate published, peer-reviewed research on "Hepatocellular Carcinoma Prediction in HCV Patients using Machine Learning and Deep Learning Techniques", a search plan was first drafted given in Fig. 1. Our MESH terms included "hepatocellular carcinoma" OR "liver cancer" OR "HCC" AND "hepatitis C virus" OR "HCV" OR "hepatitis C" AND "machine learning" OR "deep learning" OR "AI" OR "predictive models" OR "predictive algorithms" to fully investigate the research title. The search was conducted between 2014 and 2024 on 15th March 2024.

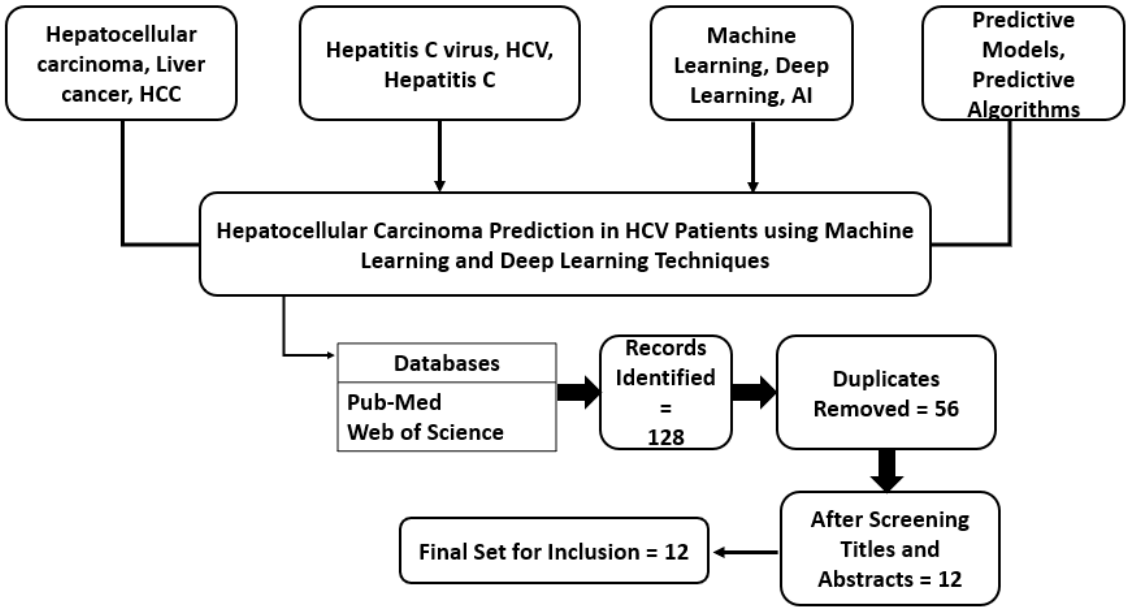

Figure 1. Strategy for articles screening and inclusion

Using these keywords, 128 publications were found. 56 of these were eliminated due to duplication. After data cleaning, considering the inclusion and exclusion criteria given in Table I, we obtained 70 articles. After reviewing the titles and abstracts for filtration, we retained 12 articles. As a result, 12 publications that showed their alignment with our research questions were selected for review under our research topic.

Table I: The inclusion and exclusion standard for the articles in the research

Inclusion Standards	Exclusion Standards
The study focuses on HCC prediction in patients suffering from HCV	The research focuses only on HCC

Ashish Shiwlani: *Corresponding Author

JURNAL ILMIAH COMPUTER SCIENCE (JICS)
E-ISSN 3026-7145 P-ISSN 3030-9840
Volume 3, Nomor 2, January 2025, Page 120-134
DOI: https://doi.org/10.58602/jics.v3i2.48

The study is a quantitative analysis and a research article	Literature is not directly related to the research topic
The study published between last ten years	The study published before 2013

3.RESEARCH FINDINGS

The research findings from this review are mentioned below:

A. Screening and Diagnostic Techniques

i. Ultrasound

Th Ultrasound stands as the prevailing imaging technique for hepatocellular carcinoma (HCC) surveillance among HCV patients because of its affordability, non-invasiveness, reasonable diagnostic accuracy, and widespread availability [30]. Furthermore, it offers extra information helpful for the observation and evaluation of the hepatic fibrosis and cirrhosis patients, like the onset of ascites and thrombosis [31]. US has the drawback of being an operator-dependent technique [30], [31]. Some conditions, like obesity, ascites, or severe liver disease, can affect the sensitivity of US, so in some cases, alternative techniques may be required [32].

ii. CT and MRI

CT and MRI are not cost-effective as screening tests, but they are helpful in the diagnosis of liver lesions [33]. Even though these tests are more sensitive at identifying lesions, particularly in the early stages, the cost increase is typically not justified by this increased sensitivity [34]. Considering the estimated doubling time of HCC, opting for ultrasound every six months seems to be favored over annual CT or MRI [35]. Aside from the expense, these methods' drawbacks include radiation exposure, nephrotoxicity risk, allergic reactions to CT contrast, the availability of MRI equipment in certain facilities, the length of the MR scan, discomfort, and the use of gadolinium-containing contrasts [33], [35]. However, the usefulness of employing alternate methods like CT or MRI, as well as its frequency, it should be tailored for patients facing difficulties with US assessments and potentially exhibiting low sensitivity [35].

iii. Alpha Fetoprotein

It is the most interested HCC biomarker [36]. A positive result is defined as more than 20 ng/mL; however, this method has a low specificity at these levels and a high specificity but low sensitivity at 200 ng/mL and higher [37]. As per certain research findings, including the measurement of AFP in the imaging controls may enhance the sensitivity by 6% to 8% for early detection of HCC, albeit at the cost of a minor reduction in specificity [37]. This low yield of AFP is caused by the fact that only 10%–20% of HCC has high values in the early stages, and changed levels of this molecule can be seen in certain chronic liver diseases without any correlation to HCC [36], [37]. Furthermore, the cost of screening is greatly increased when AFP is added to the US [38]. It is not possible to establish a definitive recommendation for including AFP in the imaging test due to all these factors [39].

iv. Serum Biomarkers

Serum biomarkers have relatively low sensitivity and specificity, so they cannot be used in isolation for HCC surveillance [39]. They can, however, raise sensitivity when paired with imaging techniques, though doing so increases false-positive results [39].

Ashish Shiwlani: *Corresponding Author

JURNAL ILMIAH COMPUTER SCIENCE (JICS)
E-ISSN 3026-7145 P-ISSN 3030-9840
Volume 3, Nomor 2, January 2025, Page 120-134
DOI: https://doi.org/10.58602/jics.v3i2.48

B. Machine Learning Techniques

Machine learning (ML) is characterized as a branch of computational procedures that use data samples to comprehend and develop algorithms that learn tasks, like outcome analysis or classification [40]. Using machine learning (ML), predictive models can be fitted to data to compute objectively or approximatively identify patterns [41]. Additional categories for machine learning include supervised, unsupervised, and reinforced learning. Utilizing labeled data to train algorithms for modeling and accurate outcome prediction is known as supervised machine learning. Whereas unsupervised learning makes use of data without labels to find patterns that are hidden or unknown [40]. Consequently, reinforcement learning denotes a sophisticated form of unsupervised learning in which the learning model can predict the result by means of a series of decisions made through continuous trial and error [40]. A brief overview is given in Figure 2.

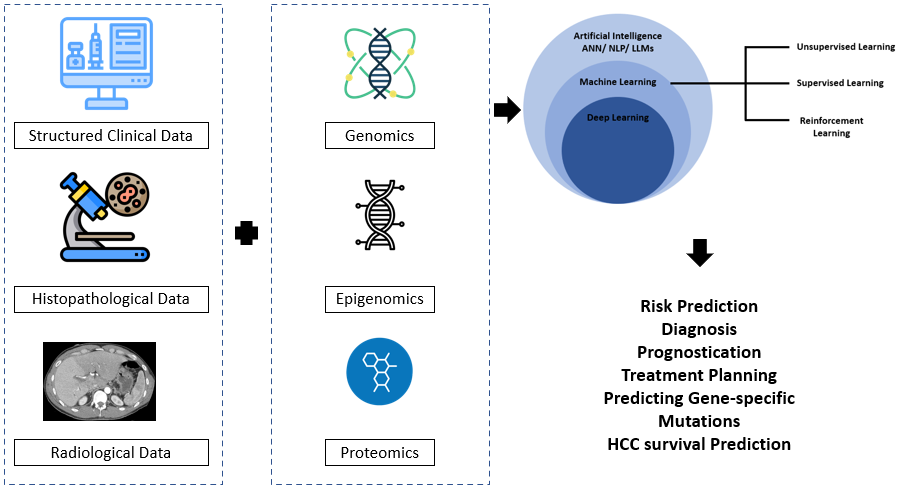

Figure 2. Integrating Multimodal Data with AI Techniques for HCC Risk Prediction

i. Decision Tree Techniques

Decision-tree learning stands out as a promising method for prediction and classification across the domains of data mining, machine learning, and statistics. One type of decision-tree algorithm is known as alternating decision tree (ADTree) which combines boosting and decision-tree algorithms [42].

ii. Multi-linear Regression

To formulate an equation depicting the linear connection, multi-linear regression evaluates how tightly linked a set of explanatory variables is to a sole response variable. The risk of developing HCC is the target variable, and albumin, age, hemoglobin, total bilirubin, and platelet count are its explanatory factors [43].

iii. Support Vector Machine (SVM)

SVM is a classifier that divides cases with different class labels into hyper planes in a multidimensional space to carry out classification tasks. Multiple continuous and categorical variables can be handled by SVM, which supports both regression and classification tasks. A dummy variable is

Ashish Shiwlani: *Corresponding Author

JURNAL ILMIAH COMPUTER SCIENCE (JICS)
E-ISSN 3026-7145 P-ISSN 3030-9840
Volume 3, Nomor 2, January 2025, Page 120-134
DOI: https://doi.org/10.58602/jics.v3i2.48

made for categorical variables, with case values of either 0 or 1. A decision plane is identified as a boundary that separates a collection of items with different class affiliations [44]. SVM has applications in feature selection and classification [45].

iv. Random Forest Model

A random forest model is an arrangement of trees. Another name for it is a random decision forest. It produces several trees for categorization. It generates a great deal of distinct decision trees. The result of random forest is the average prediction of each individual tree [44].

v. Feature Selection

During the model construction phase, variable selection, also known as feature selection or subset selection, is utilized to eliminate redundant variables and maintain simplicity and comprehensibility. Selecting a variable subset to maximize classification or prediction accuracy is part of this process. The three categories of variable selection filters, wrappers, and embedded methods which have been frequently used for HCC prediction studies are depicted in Figure 3.

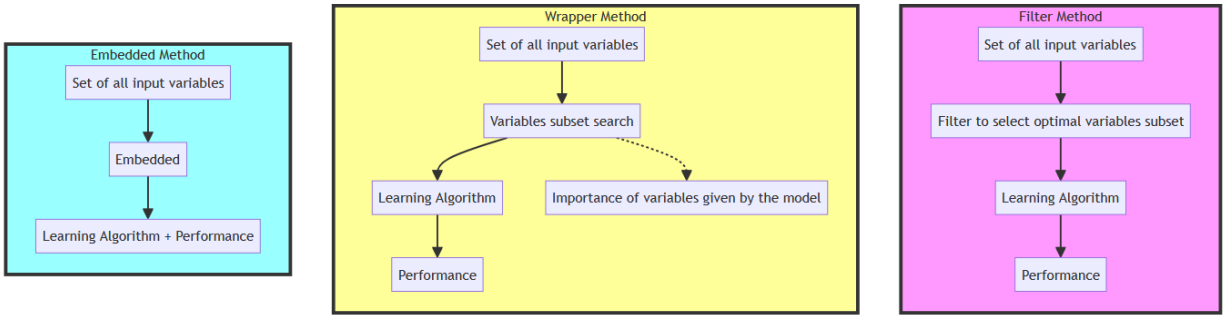

Figure 3. The commonly employed feature selection methods in the risk prediction of HCC

C. Deep Learning Techniques

DL is a subset of machine learning (ML) models that are created with neural networks (NNs, also called artificial neural networks, ANNs, and forward feed neural networks [46]. These neural networks consist of mathematical algorithms designed to mimic the functionality of neurons in the human brain. From high-dimensional data, these NNs extract essential, perplexing, and non-linear features [47]. Multiples of densely connected computing neurons are arranged in successive layers of a deep neural network, which receives input and sends it to hide or succeeding layers where it is multiplied, divided, added, or subtracted thousands of times before reaching the output layer, which indicates the algorithm's expected result [48]. Fig. 4 depicts a ML and DL-based framework being employed in the prediction of HCC.

Ashish Shiwlani: *Corresponding Author

JURNAL ILMIAH COMPUTER SCIENCE (JICS)
E-ISSN 3026-7145 P-ISSN 3030-9840
Volume 3, Nomor 2, January 2025, Page 120-134
DOI: https://doi.org/10.58602/jics.v3i2.48

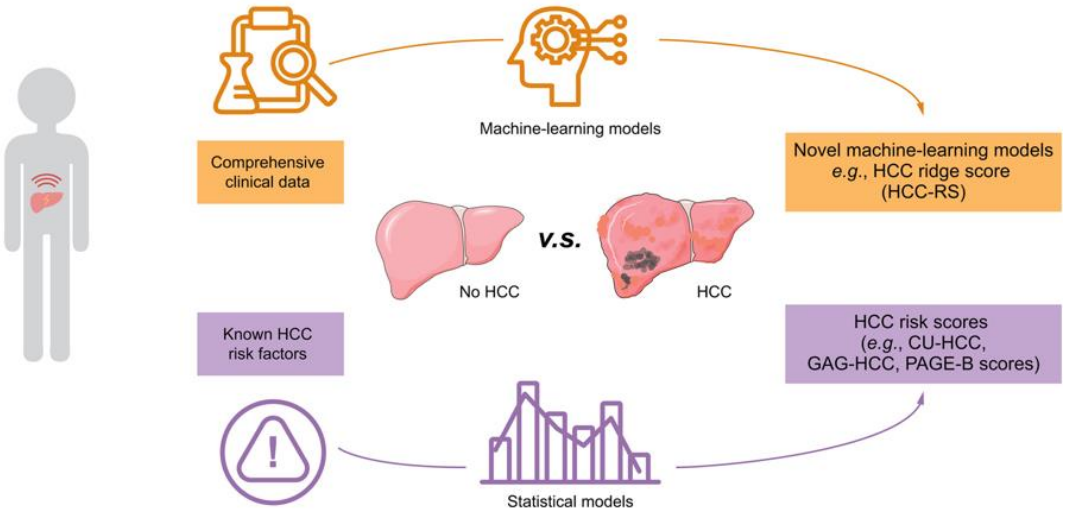

Figure 4. Calculation of Risk HCC in Chronic Viral Hepatitis (HCV) Patients Using Clinical Data

i. CNNs

Convolutional neural networks are among the most widely utilized deep learning models. Three layers make up CNN's fundamental architecture: convolutional, pooling, and fully connected layers. They are primarily employed in feature extraction tasks where it is necessary to maintain the spatial arrangement of the data and guarantee translation invariance. Their primary application lies in image analysis, particularly in cases where features pertain to object shapes, such as in the automated detection of focal liver lesions in CT or MRI scans [46], [47], [48].

ii. RNNs

The recurrent neural network is another well-liked deep learning model (RNN). RNN algorithms, in contrast to CNNs, are often employed in text analysis and in the processing of sequential data problems in order to predict a result. RNNs work by giving the neural network internal memories, which are then processed and stored in later stages [47]. RNNs process sequential data points using a recurrent hidden state, in contrast to CNNs, where each step depends on the one before it [48].

Ashish Shiwlani: *Corresponding Author

Table II. The summarized studies using machine learning algorithms for the risk prediction of HCC in HCV Patients

Ref	Target	Input	Architecture	Pre-Processing	Dataset	Outcome	Output	Result
[49]	HCC	Age, AFP, ALP, albumin, total bilirubin	Decision Tree (CART, ADTree, REP-tree), Linear Regression	Filtering, significance test (P-value), correlation coefficient analysis	HCV patients - 4,423 From two institutes in Egypt	Prediction of HCC presence	(AUROC)	AUROC: 99%, Accuracy: 95.6%
[50]	HCC occurrence	Clinical data including sex, age, platelet count, albumin, total bilirubin, ALT, AST, alpha-fetoprotein, liver disease markers, comorbidities, medications	LR, AdaBoost, RF, Ridge Regression	Feature selection with filter methods, split dataset into training and validation cohorts (7:3 ratio)	Hospital Authority Data Collaboration Lab (HADCL) in Hong Kong, including 124,006 patients with HCV	HCC-RS	HCC-RS showed higher accuracy	AUROC: 0.844
[51]	HCC, Advanced hepatic fibrosis and Cirrhosis	Clinical data: age, AST, ALT, PLT	Decision Tree, Random Forest, Gradient Boosting	10-fold cross-validation, hyper-parameter tuning, Jackknife resampling	Discovery dataset: 490 HBV patients; Validation-1: 86 HBV patients; Validation-2: 254 HCV patients; Validation-3: 230 HCV patients (pre- and post-treatment)	GB-based risk score and FIB-4 score	Gradient Boosting outperformed other models	AUROC: 0.817
[52]	HCC Risk	Age, platelet count, albumin, AFP, GGT,albumin, AST	Random Survival Forest (RSF)	Hyperparameter tuning, KNN imputation for missing values	1,742 patients- Training 977 patients -validation	Predictive risk stratification for HCC	Individualized predictive curve for HCC	C – index: 0.839
[53]	HCC Prediction	49 features including age, gender, alcohol consumption, hepatitis B, hepatitis C, liver function tests	KNN, Naive Bayes, DT, RF, SVM	Handling missing values, manual correction of data errors, feature scaling	HCC Dataset from University Hospital in Portugal, 165 clinical patient data	HCC outcome (Die/Live)	Binary classification (0: Dies, 1: Lives)	Acc.: 80.64
[54]	HCC Prediction	Patient data including liver function tests, demographics, clinical characteristics, etc.	Regression Analysis	Handling missing values, feature selection, normalization	442 cirrhotic patients from the University of Michigan (UM cohort)	HCC development	Binary classification	C – index: 0.64

					and validation on HALT-C Trial cohort			
[55]	HCC diagnosis prediction	Clinical and laboratory data (albumin, TB, age, sex, AFP, AFP-L3, DCP, AST, ALT, platelet count, ALP, GGT)	Gradient Boosting, SVM, RFs, NNs, Deep Learning	Handling missing values, feature selection, normalization, grid search for hyperparameter optimization	539 HCC and 1043 non-HCC (University of Tokyo Hospital)	Presence of HCC	Binary classification (HCC presence or not)	Acc.: 87.34
[56]	Risk of HCC occurrence	Clinical and biological features, SVR status	Random Survival Forest (RSF), Fine-Gray regression, Decision Tree (DT)	Recursive partitioning, time-dependent covariates	ANRS CO12 CirVir cohort, ANRS CO22 Hepather cohort	Prognostic models for HCC occurrence	Novel HCC risk classes based on SVR	Model / C-index FGR / 0.645 DT / 0.59 RSF / 0.69

Table III. The summarized studies using deep learning algorithms for the prediction of HCC in HCV Patients

Ref	Target	Input	Architecture	Pre-Processing	Dataset	Outcome	Output	Result
[57]	HCC prediction	Longitudinal data from electronic health records including age, sex, race, HCV genotype, BMI, and 24 laboratory tests	RNN-GRU	Imputation of missing values by mean of non-missing entries, filling missing entries at subsequent times by looking backward	48,151 patients from Veterans Health Administration, USA (2000-2019)	Prediction of HCC within 3 years	Risk probability	AUROC: 0.759
[58]	HCC prediction	Age, sex, platelet count, serum albumin, serum bilirubin, presence of diabetes, HCV DNA levels	Deep Neural Network (DNN) with 2 residual blocks, 7 layers	Numerical data were normalized; factors such as cirrhosis, etiology, and race were excluded	Derivation cohort: 424 patients Validation cohort: 316 patients	Prediction of HCC	Risk probability	C – index: 0.719
[59]	Liver cancer prediction	History of diseases in hepatitis cohort	Convolutional Neural Network (CNN)	Backward imputation for missing data	National Health Insurance Research	Prediction of HCC in hepatitis patients	Probability of liver cancer	Acc.: 0.98

					Database (NHIRD)			
[60]	Development of HCC within 3 years	Age, gender, race, HCV genotype, SVR, 21 longitudinal predictors	Recurrent Neural Network (RNN)	Backward imputation for missing data	146,218 patients with HCV infection from Veterans Health Administration	Prediction of development of HCC in HCV patients	Probability of HCC development	AUROC: 0.89

JURNAL ILMIAH COMPUTER SCIENCE (JICS)
E-ISSN 3026-7145 P-ISSN 3030-9840
Volume 3, Nomor 2, January 2025, Page 120-134
DOI: https://doi.org/10.58602/jics.v3i2.48

4.DISCUSSION AND OPEN RESEARCH AREAS

HCC diagnosis is based on several factors and standards. Clinical trials, laboratory tests (blood, urine, bilirubin content, etc.), imaging tests, biopsies, etc. are all involved in the diagnosis of disease [19]. This allows for the diagnosis of the HCC's stages, from early to critical, as well as the classification of liver cancer into groups such as transplantable, metastatic, inoperable, etc., and the analysis of survival times in relation to the degree of organ involvement [52], [56].

A review of artificial intelligence algorithms applied to different liver and HCV dataset parameters was provided by the proposed work [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61]. To find the best accuracy, a variety of data mining techniques were applied to heterogeneous datasets, including classification algorithms and other machine learning techniques. Every study provided its own analysis of HCC and other liver-related disorders, along with its own hypothesis and conclusions [5], [55], [58], [59], [60]. Tumor markers, reports from FNAC (Fine Needle Aspiration Cytology), and imaging studies can all be taken into consideration for future research [59], [61]. Further research on medical data mining can be conducted. Data mining is a technique for transforming unprocessed data into meaningful information by analyzing vast amounts of data to find hidden patterns. Data miners have a strong interest in diagnosing liver disorder diseases, but the availability of adequate data has made it difficult to diagnose the condition at an early stage [6], [57], [60].

The ML and DL analyses involve the identification of correlations between HCC risk factors in patients with liver inflammation [58]. Notably, the results differed depending on the SVR that is achieved. The analyses are carried out by applying machine learning techniques tailored to time variant features [54]. While previously developed statistical models primarily employed group of features that were independently associated with HCC as a basis for standard regression modeling approaches, the current techniques take advantage of their capacity to model complexity [57], [59].

The wide range of variations in HCC among patients with respect to their pathogenesis and risk factors broadens and restricts the approaches used for prognosis, management, and prediction [51], [56], [59], [60]. Recent developments in AI have given rise to special opportunities for enhancing HCC, which have been investigated through improved pre-screening analysis to estimate the likelihood that patients with HCV will develop HCC; ii) improved diagnosis; iii) enhances prognosis and management of the disease. The utilization of CT, MRI, and histology scans for HCC diagnosis has improved with computer-aided automation, leading to faster patient diagnosis and care, as well as a reduction in the possibility of unfavorable effects on clinical outcomes [51].

Certain studies limit serum levels of AFP [56] to create prediction models for HCC. For instance, there is a significant correlation between AFP levels and HCC, suggesting that AFP may be a useful predictor of HCC that can differentiate between early and advanced stages of the disease. AFP ≥50.3 ng/mL is a predictor of HCC [61]. [51]investigated patients who had anti-HCV antibodies, but they discovered that the AFP was not a reliable predictor of HCC on its own. Without taking AFP into account, [53] created a straightforward model for predicting the risk of HCC based on independent factors such as albumin, age, platelets, and aspartate aminotransferase.

Based on this study's findings as well as earlier research when it comes to medical prediction models, the ADTree model outperforms other machine-learning techniques, particularly in its ability to generalize and enhance predictions through boosting. Each iteration of the data has a distinct distribution due to the use of boosted algorithms in ADTs. Additionally, the ADT model incorporates feature selection methods. Moreover, this model's simplicity makes interpretation through interactive visualization easier.

Since the volume of data used to train AI models greatly influences their performance, the presence of extensive data samples is crucial to advancing the prediction models and ensuring that they

Ashish Shiwlani: *Corresponding Author

will eventually have an impact on clinical care. Large datasets should be deposited and shared in order to achieve this goal. Structured data from EHRs is used and shared within and between health systems as part of this. Additionally, there's a growing consensus among scientists and institutions regarding the ethical and scientific necessity of sharing individual-participant data (IPD) derived from clinical trials or academic research endeavors. This would help create datasets big enough and detailed enough to properly train and validate AI models.

The lack of adequate racial, ethnic, and socioeconomic diversity in the cohorts used for developing and training AI models for HCC risk prediction and prognosis is a significant issue thus far. This poses a crucial challenge as the accuracy of AI-based algorithms heavily relies on the quality and quantity of input data. Subsequent studies should ensure thorough validation of promising AI-driven tools across diverse cohorts, including individuals from varied socioeconomic backgrounds and ethnic and racial minorities. This underscores the significance of data sharing among researchers and institutions to construct representative cohorts.

5. CONCLUSION

High accuracy can be achieved in estimating the risk of HCC development through machine learning and deep learning techniques. These techniques provide doctors with effective, non-invasive means of diagnosing and tracking patients with HCC. The presence of HCC was found to be strongly correlated with age, AFP, ALP, albumin, and total bilirubin. The effectiveness of alternating decision trees, reduced error pruning trees, multi-linear regression, classification and regression trees, and neural networks in detecting the presence of HCC was evaluated in this study. The machine learning and deep learning methods that were investigated produced comparable outcomes in terms of HCC identification. To fully integrate such algorithms in clinical practice, there is a need to develop reliable methods for the gathering, sharing, and storing well-labeled and structured data as well as the requirement to show the dependability and resilience of models.

.

6. ACKNOWLEDGMENT

The authors acknowledge the support from Buch International Hospital, Multan, Pakistan. https://www.buchhospital.com/

7. REFRENCES

1. A. M. Moon, A. G. Singal, and E. B. Tapper, “Contemporary Epidemiology of Chronic Liver Disease and Cirrhosis,” Clinical Gastroenterology and Hepatology, vol. 18, no. 12, pp. 2650–2666, Nov. 2020, doi: 10.1016/J.CGH.2019.07.060.
2. N. Ali and K. Hossain, “Liver injury in severe COVID-19 infection: current insights and challenges,” Expert Rev Gastroenterol Hepatol, Oct. 2020, doi: 10.1080/17474124.2020.1794812.
3. A. D. Nardo, M. Schneeweiss-Gleixner, M. Bakail, E. D. Dixon, S. F. Lax, and M. Trauner, “Pathophysiological mechanisms of liver injury in COVID-19,” Liver International, vol. 41, no. 1, pp. 20–32, Jan. 2021, doi: 10.1111/LIV.14730.
4. A. R. Silva et al., “HBV and HCV serological markers in health professionals and users of the Brazilian Unified Health System network in the city of Resende, Rio de Janeiro, Brazil,” J Bras Patol Med Lab, vol. 53, no. 2, pp. 92–99, Jan. 2017, doi: 10.5935/1676-2444.20170016.
5. A. Villanueva, V. Hernandez-Gea, and J. M. Llovet, “Medical therapies for hepatocellular carcinoma: a critical view of the evidence,” Nature Reviews Gastroenterology & Hepatology 2012 10:1, vol. 10, no. 1, pp. 34–42, Nov. 2012, doi: 10.1038/nrgastro.2012.199.

Ashish Shiwlani: *Corresponding Author

6. S. Mohammadreza Mortazavizadeh, S. Rafatmagham, F. Tabatabaie, R. Hakimizad, S. Mohammad, and A. Hashemipour, "Frequency distribution and ten-year survival rate of patients with different malignant liver lesions in Iran," 2022, doi: 10.51847/u57GBNwQq3.
7. E. Vo Quang, Y. Shimakawa, and P. Nahon, "Epidemiological projections of viral-induced hepatocellular carcinoma in the perspective of WHO global hepatitis elimination," Liver International, vol. 41, no. 5, pp. 915–927, May 2021, doi: 10.1111/LIV.14843.
8. D. Liu et al., "Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound," Eur Radiol, vol. 30, no. 4, pp. 2365–2376, Apr. 2020, doi: 10.1007/S00330-019-06553-6/METRICS.
9. Y. Dong et al., "Preoperative prediction of microvascular invasion (MVI) in hepatocellular carcinoma based on kupffer phase radiomics features of sonazoid contrast-enhanced ultrasound (SCEUS): A prospective study," Clin Hemorheol Microcirc, vol. 81, no. 1, pp. 97–107, Jan. 2022, doi: 10.3233/CH-211363.
10. S. Hafeez, M. S. Alam, Z. Sajjad, Z. A. Khan, W. Akhter, and F. Mubarak, "Triphasic computed tomography (CT) scan in focal tumoral liver lesions," J Pak Med Assoc, vol. 61, no. 6, Jun. 2011, Accessed: May 26, 2024. [Online]. Available: https://ecommons.aku.edu/pakistan_fhs_mc_radiol/22
11. J. Drott, B. Björnsson, P. Sandström, and C. Berterö, "Experiences of Symptoms and Impact on Daily Life and Health in Hepatocellular Carcinoma Patients: A Meta-synthesis of Qualitative Research," Cancer Nurs, vol. 45, no. 6, pp. 430–437, Nov. 2022, doi: 10.1097/NCC.0000000000001044.
12. M. Serper et al., "Patient-reported outcomes in HCC: A scoping review by the Practice Metrics Committee of the American Association for the Study of Liver Diseases," Hepatology, vol. 76, no. 1, pp. 251–274, Jul. 2022, doi: 10.1002/HEP.32313.
13. S. Hashem et al., "Accurate Prediction of Advanced Liver Fibrosis Using the Decision Tree Learning Algorithm in Chronic Hepatitis C Egyptian Patients," Gastroenterol Res Pract, vol. 2016, 2016, doi: 10.1155/2016/2636390.
14. W. Wang and C. Wei, "Advances in the early diagnosis of hepatocellular carcinoma," Genes Dis, vol. 7, no. 3, pp. 308–319, Sep. 2020, doi: 10.1016/J.GENDIS.2020.01.014.
15. K. Cui, Y. Ou, Y. Shen, S. Li, and Z. Sun, "Clinical value of circulating tumor cells for the diagnosis and prognosis of hepatocellular carcinoma (HCC): A systematic review and meta-Analysis," Medicine (United States), vol. 99, no. 40, p. E22242, Oct. 2020, doi: 10.1097/MD.0000000000022242.
16. R. Sartoris, J. Gregory, M. Dioguardi Burgio, M. Ronot, and V. Vilgrain, "HCC advances in diagnosis and prognosis: Digital and Imaging," Liver International, vol. 41, no. S1, pp. 73–77, Jun. 2021, doi: 10.1111/LIV.14865.
17. D. Q. Huang, H. B. El-Serag, and R. Loomba, "Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention," Nature Reviews Gastroenterology & Hepatology 2020 18:4, vol. 18, no. 4, pp. 223–238, Dec. 2020, doi: 10.1038/s41575-020-00381-6.
18. L. Zhang et al., "Prediction of HCC microvascular invasion with gadobenate-enhanced MRI: correlation with pathology," Eur Radiol, vol. 30, no. 10, pp. 5327–5336, Oct. 2020, doi: 10.1007/S00330-020-06895-6/METRICS.
19. S. Hashem et al., "Machine Learning Prediction Models for Diagnosing Hepatocellular Carcinoma with HCV-related Chronic Liver Disease," Comput Methods Programs Biomed, vol. 196, p. 105551, Nov. 2020, doi: 10.1016/J.CMPB.2020.105551.
20. M. Yumus, M. Apaydin, A. Degirmenci, and O. Karal, "Missing data imputation using machine learning based methods to improve HCC survival prediction," 2020 28th Signal Processing and Communications Applications Conference, SIU 2020 - Proceedings, Oct. 2020, doi: 10.1109/SIU49456.2020.9302222.
21. A. Bakrania, N. Joshi, X. Zhao, G. Zheng, and M. Bhat, "Artificial intelligence in liver cancers: Decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases," Pharmacol Res, vol. 189, p. 106706, Mar. 2023, doi: 10.1016/J.PHRS.2023.106706.
22. Y. Wang, C. Ji, Y. Wang, M. Ji, J. J. Yang, and C. M. Zhou, "Predicting postoperative liver cancer death outcomes with machine learning," Curr Med Res Opin, vol. 37, no. 4, pp. 629–634, 2021, doi: 10.1080/03007995.2021.1885361.

Ashish Shiwlani: *Corresponding Author

23. S. Hashem et al., "A SIMPLE MULTI-LINEAR REGRESSION MODEL FOR PREDICTING FIBROSIS SCORES IN CHRONIC EGYPTIAN HEPATITIS C VIRUS PATIENTS," International Journal of Bio-Technology and Research (IJBTR) ISSN(P, vol. 4, pp. 37–46, 2014, Accessed: May 26, 2024. [Online]. Available: www.tjprc.org
24. C. Le Berre et al., "Application of Artificial Intelligence to Gastroenterology and Hepatology," Gastroenterology, vol. 158, no. 1, pp. 76-94.e2, Jan. 2020, doi: 10.1053/J.GASTRO.2019.08.058.
25. T. H. Su, C. H. Wu, and J. H. Kao, "Artificial intelligence in precision medicine in hepatology," J Gastroenterol Hepatol, vol. 36, no. 3, pp. 569–580, Mar. 2021, doi: 10.1111/JGH.15415.
26. J. Calderaro, T. P. Seraphin, T. Luedde, and T. G. Simon, "Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma," J Hepatol, vol. 76, no. 6, pp. 1348–1361, Jun. 2022, doi: 10.1016/J.JHEP.2022.01.014.
27. J. C. Ahn, T. A. Qureshi, D. Li, A. G. Singal, and J. D. Yang, "Deep learning in hepatocellular carcinoma: Current status and future perspectives," World J Hepatol, vol. 13, no. 12, p. 2039, Dec. 2021, doi: 10.4254/WJH.V13.I12.2039.
28. R. Sarkis-Onofre, F. Catalá-López, E. Aromataris, and C. Lockwood, "How to properly use the PRISMA Statement," Syst Rev, vol. 10, no. 1, pp. 1–3, Dec. 2021, doi: 10.1186/S13643-021-01671-Z/METRICS.
29. M. Norris and C. Oppenheim, "Comparing alternatives to the Web of Science for coverage of the social sciences' literature," J Informetr, vol. 1, no. 2, pp. 161–169, Apr. 2007, doi: 10.1016/J.JOI.2006.12.001.
30. J. Bo, F. Xiang, F. XiaoWei, Z. LianHua, L. ShiChun, and L. YuKun, "A Nomogram Based on Contrast-Enhanced Ultrasound to Predict the Microvascular Invasion in Hepatocellular Carcinoma," Ultrasound Med Biol, vol. 49, no. 7, pp. 1561–1568, Jul. 2023, doi: 10.1016/J.ULTRASMEDBIO.2023.02.020.
31. I. Nakamura et al., "Enhanced patterns on intraoperative contrast-enhanced ultrasonography predict outcomes after curative liver resection in patients with hepatocellular carcinoma," Surg Today, vol. 51, no. 5, pp. 764–776, May 2021, doi: 10.1007/S00595-020-02145-W/METRICS.
32. N. Nishida and M. Kudo, "Artificial Intelligence in Medical Imaging and Its Application in Sonography for the Management of Liver Tumor," Front Oncol, vol. 10, p. 594580, Dec. 2020, doi: 10.3389/FONC.2020.594580/BIBTEX.
33. F. Vernuccio et al., "Benign and malignant mimickers of infiltrative hepatocellular carcinoma: tips and tricks for differential diagnosis on CT and MRI," Clin Imaging, vol. 70, pp. 33–45, Feb. 2021, doi: 10.1016/J.CLINIMAG.2020.10.011.
34. Y. Liang et al., "Diagnostic performance of LI-RADS for MRI and CT detection of HCC: A systematic review and diagnostic meta-analysis," Eur J Radiol, vol. 134, p. 109404, Jan. 2021, doi: 10.1016/J.EJRAD.2020.109404.
35. S. Lee et al., "Percentages of hepatocellular carcinoma in LI-RADS categories with CT and MRI: A systematic review and meta-Analysis," Radiology, vol. 307, no. 1, Apr. 2023, doi: 10.1148/RADIOL.220646/ASSET/IMAGES/LARGE/RADIOL.220646.FIG4.JPEG.
36. F. Özdemir and A. Baskiran, "The Importance of AFP in Liver Transplantation for HCC," J Gastrointest Cancer, vol. 51, no. 4, pp. 1127–1132, Dec. 2020, doi: 10.1007/S12029-020-00486-W/METRICS.
37. P. Luo et al., "Current Status and Perspective Biomarkers in AFP Negative HCC: Towards Screening for and Diagnosing Hepatocellular Carcinoma at an Earlier Stage," Pathology and Oncology Research, vol. 26, no. 2, pp. 599–603, Apr. 2020, doi: 10.1007/S12253-019-00585-5/METRICS.
38. N. Mehta et al., "AFP-L3 and DCP are superior to AFP in predicting waitlist dropout in HCC patients: Results of a prospective study," Liver Transplantation, vol. 29, no. 10, pp. 1041–1049, Oct. 2023, doi: 10.1097/LVT.0000000000000149.
39. J. D. Debes, P. A. Romagnoli, J. Prieto, M. Arrese, A. Z. Mattos, and A. Boonstra, "Serum Biomarkers for the Prediction of Hepatocellular Carcinoma," Cancers 2021, Vol. 13, Page 1681, vol. 13, no. 7, p. 1681, Apr. 2021, doi: 10.3390/CANCERS13071681.
40. I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput Sci, vol. 2, no. 3, pp. 1–21, May 2021, doi: 10.1007/S42979-021-00592-X/FIGURES/11.

Ashish Shiwlani: *Corresponding Author

41. J. G. Greener, S. M. Kandathil, L. Moffat, and D. T. Jones, "A guide to machine learning for biologists," Nature Reviews Molecular Cell Biology 2021 23:1, vol. 23, no. 1, pp. 40–55, Sep. 2021, doi: 10.1038/s41580-021-00407-0.
42. B. Taha Jijo and A. Mohsin Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
43. K. K. B and D. S. S, "A Survey on Predicting Advanced Liver Fibrosis Using Different Machine Learning Algorithms," Int J Sci Res Sci Eng Technol, vol. 7, no. 1, pp. 177–183, 2020, doi: 10.32628/IJSRSET207138.
44. M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," IEEE J Sel Top Appl Earth Obs Remote Sens, vol. 13, pp. 6308–6325, 2020, doi: 10.1109/JSTARS.2020.3026724.
45. D. M. Abdullah and A. M. Abdulazeez, "Machine Learning Applications based on SVM Classification A Review," Qubahan Academic Journal, vol. 1, no. 2, pp. 81–90, Apr. 2021, doi: 10.48161/QAJ.V1N2A50.
46. P. M. Oestmann et al., "Deep learning–assisted differentiation of pathologically proven atypical and typical hepatocellular carcinoma (HCC) versus non-HCC on contrast-enhanced MRI of the liver," Eur Radiol, vol. 31, no. 7, pp. 4981–4990, Jul. 2021, doi: 10.1007/S00330-020-07559-1/METRICS.
47. G. N. Ioannou 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, pp. e2015626–e2015626, Sep. 2020, doi: 10.1001/JAMANETWORKOPEN.2020.15626.
48. J. C. Ahn, T. A. Qureshi, D. Li, A. G. Singal, and J. D. Yang, "Deep learning in hepatocellular carcinoma: Current status and future perspectives," World J Hepatol, vol. 13, no. 12, p. 2039, Dec. 2021, doi: 10.4254/WJH.V13.I12.2039.
49. S. Hashem et al., "Machine Learning Prediction Models for Diagnosing Hepatocellular Carcinoma with HCV-related Chronic Liver Disease," Comput Methods Programs Biomed, vol. 196, p. 105551, Nov. 2020, doi: 10.1016/J.CMPB.2020.105551.
50. G. L. H. Wong et al., "Novel machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis," JHEP Reports, vol. 4, no. 3, p. 100441, Mar. 2022, doi: 10.1016/J.JHEPR.2022.100441.
51. R. Wei et al., "Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning," EBioMedicine, vol. 35, pp. 124–132, Sep. 2018, doi: 10.1016/j.ebiom.2018.07.041.
52. T. Minami et al., "Machine learning for individualized prediction of hepatocellular carcinoma development after the eradication of hepatitis C virus with antivirals," J Hepatol, vol. 79, no. 4, pp. 1006–1014, Oct. 2023, doi: 10.1016/J.JHEP.2023.05.042.
53. S. Rajesh, N. A. Choudhury, and S. Moulik, "Hepatocellular Carcinoma (HCC) Liver Cancer prediction using Machine Learning Algorithms," 2020 IEEE 17th India Council International Conference, INDICON 2020, Dec. 2020, doi: 10.1109/INDICON49873.2020.9342443.
54. A. G. Singal et al., "Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma," American Journal of Gastroenterology, vol. 108, no. 11, pp. 1723–1730, Nov. 2013, doi: 10.1038/AJG.2013.332.
55. M. Sato et al., "Machine-learning Approach for the Development of a Novel Predictive Model for the Diagnosis of Hepatocellular Carcinoma," Scientific Reports 2019 9:1, vol. 9, no. 1, pp. 1–7, May 2019, doi: 10.1038/s41598-019-44022-8.
56. E. Audureau et al., "Personalized surveillance for hepatocellular carcinoma in cirrhosis – using machine learning adapted to HCV status," J Hepatol, vol. 73, no. 6, pp. 1434–1445, Dec. 2020, doi: 10.1016/J.JHEP.2020.05.052.
57. G. N. Ioannou 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, doi: 10.1001/jamanetworkopen.2020.15626.

Ashish Shiwlani: *Corresponding Author

58. J. Y. Nam, D. H. Sinn, J. Bae, E. S. Jang, J. W. Kim, and S. H. Jeong, "Deep learning model for prediction of hepatocellular carcinoma in patients with HBV-related cirrhosis on antiviral therapy," JHEP Reports, vol. 2, no. 6, p. 100175, Dec. 2020, doi: 10.1016/J.JHEPR.2020.100175.
59. D. Van Phan, C. L. Chan, A. H. A. Li, T. Y. Chien, and V. C. Nguyen, "Liver cancer prediction in a viral hepatitis cohort: A deep learning approach," Int J Cancer, vol. 147, no. 10, pp. 2871–2878, Nov. 2020, doi: 10.1002/IJC.33245.
60. [60] G. N. Ioannou et al., "498 DEEP LEARNING MODELS ACCURATELY PREDICT DEVELOPMENT OF HCC IN 146,218 PATIENTS WITH CHRONIC HEPATITIS C".
61. A. G. Singal et al., "Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma," American Journal of Gastroenterology, vol. 108, no. 11, pp. 1723–1730, Nov. 2013, doi: 10.1038/AJG.2013.332.

Ashish Shiwlani: *Corresponding Author