iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** III **Month of publication:** March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67687

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

Differentiating Viral and Non-Viral Hepatocellular Carcinoma Using Machine Learning

S Yamuna[1], S J Ashwiya[2], S Alvish[3], Dr.R.Karunia Krishnapriya[4], Mr.Pandreti Praveen[5], Mr.V.Shaik Mohammad Shahil[6], Mr.N.Vijaya Kumar[7]
[1, 2, 3]UGScholar, [4]Associate Professor, [5,6,7]Assistant Professor,Sreenivasa Institute of Technology and Management Studies, Chittoor,India.

Abstract: ***The Pathogenesis and treatment outcomes of hepatocellular carcinoma(HCC), a major cause of cancer-related death globally, are influenced by a variety of etiological factors. Using a large dataset comprising clinical, demographic, and molecular characteristics, this study explores the possibility of using machine learning methods to distinguish between viral and non-viral HCC. To evaluate the data and find the important predictors of HCC etiology, we used a variety of machine learning models, such as stacking classifiers, Logistic Regression, Decision tree, random forests, and neural networks. Our findings show that machine learning techniques can classify HCC subtypes with high accuracy, and that certain features like viral load, liver function tests, and histological features emerge as important discriminators. The result highlight how incorporating machine learning into clinical practice can be beneficial. Histological features emerging as important with different etiological factors impacting its pathogenesis and treatment outcomes, hepatocellular carcinoma(HCC) is a prominent case of cancer-related mortality that occurs globally. This study uses a large dataset that contains clinical, demographic, and molecular variables to examine how major cause of cancer-related death globally, hepatocellular carcinoma(HCC) has a variety of etiological elements that affect both its pathology. This research uses a large dataset that contains clinical, demographic, and molecular characteristics to examine how well machine learning algorithms can distinguish between viral and non-viral HCC. Support vector machines, Random forests and neural networks are just a few of the machine learning models , we used to examine the data and find the important indicators of etiology of HCC. The results highlight how incorporating machine learning into clinical practice can improve the accuracy of diagnoses and guide specialized treatment plans for patients with HCC.***

Keywords: Hepatocellular carcinoma, Hepatitis B virus (HBV), Hepatitis C virus (HCV), Machine Learning (ML), Classification, Support vector machine (SVM), Histopathological analysis.

I. INTRODUCTION

This paper opens a door for better patient outcomes through personalized therapy by adding to the increasing body of data in favor of the application of the application of sophisticated computational tools in oncology caused by viral infections, specifically hepatitis B virus(HBV) and hepatitis C virus(HCV). Non-viral HCC, on the other hand can result from a number of conditions, Although a etiology of HCC is complex, a significant percentage of cases are Imaging and histological analysis are examples of traditional diagnostic techniques that can't always offer enough specificity for precise clarification. Machine learning (ML) has become a potent instrument in medical diagnostics in recent years, with the ability to examine intricate datasets and find patterns that traditional methods might not be able to easily identify. Our goal is to create a strong model that can distinguish between viral and non-viral HCC using a mix of clinical, demographic, and molecular data by utilizing the power of machine learning algorithms. The ultimate goal of this research is to improve diagnostic precision and enable individualized treatment plans for patients with HCC by making a contribution to the expanding field of computational oncology.

For individualized treatment plansit is essential to distinguish between viral and non-viral hepatocellular carcinoma(HCC). Promising approaches to improving diagnostic precision and comprehending the fundamental biological distinctions between these two HCC subtypes are provided by machine learning techniques.

1) Tailored Treatment: Understanding the differences between viral and non-viral HCC can lead to more personalized treatment strategies, increasing patient outcomes
2) Prognostic Implications: Distinguishing between these subtypes can provide insights into prognosis, as viral HCC often has different progression patterns and responses to therapy compared to non-viral HCC.

3) Machine Learning Functions:Data analysis: Large datasets can be analyzed by machine learning algorithms, which can spot features and trends that conventional statistical techniques can miss. Using these traditional methods, predictive models that aid in early detection and patient risk stratification according to HCC subtype can be created.

4) Feature Extraction: To provide a more thorough knowledge of the illness, machine learning can help extract pertinent features from clinical, genomic, and imaging data.

II. BACKGROUND

Globally, hepatocellular carcinoma (HCC) is the most prevalent primary liver cancer and a major contributor to cancer related death. Viral infections, specifically hepatitis B virus (HBV) and Hepatitis C virus (HCV), are major factors to the multifactorial a etiology of HCC in many areas. Alcohol use, non-alcoholic fatty liver disease (NAFLD), and metabolic syndrome are examples of non-viral variables that are important in the development of HCC. Since viral and non-viral HCC subtypes have different biological behaviors, clinical presentations, and therapeutic responses, it is crucial to comprehend these distinctions for efficient management and treatment.

etiology and Epidemiology: Viral hepatitis continues to be a leading cause of HCC worldwide, especially in sub-Saharan Africa and Asia, where HBV is widespread. On the other hand, HCV is more common in places like North America and Eastern Europe. Particularly ins Western nations, where lifestyle factors like obesity and diabetes contribute to the rising incidence of NAFLD-related HCC is becoming more widely recognized. In addition to affecting the disease's pathophysiology, variations in A etiology also affect clinical results and how well a treatment works.

Pathophysiology: Viral and non-viral HCC have quite different pathophysiological processes. Viral proteins that disrupt host cellular functions can cause cirrhosis, chronic inflammation, and direct carcinogenic consequences. For example, the core protein of HCV and the X protein of HBV might interfere with regular cell signaling pathways, which encourages the development of cancer. Non-viral HCC, on the other hand, is frequently liked to metabolic dysregulation, which can develop into cancer. Developing tailored treatments and preventative measures requires an understanding of these systems.

Presentation and Diagnosis in clinical practice: Clinically, symptoms like weight loss, jaundice, and abdominal pain may overlap between viral and non-viral HCC. Nonetheless, there are significant variations in the patient demographics; younger people with a history of chronic hepatitis are more likely to develop viral HCC. Although diagnostic imaging methods such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound are crucial for detecting HCC, they might not always be able to distinguish between viral and non-viral subtypes. The most reliable method of diagnosis is still histopathological analysis. Although it is intrusive and not always possible, histopathological examination is still the gold standard for diagnosis.

Oncology and Machine Learning: With the ability to examine intricate datasets and find patterns that conventional approaches might miss, machine learning (ML) has become a game-changing tool in cancer. To improve diagnostic precision and subtype distinction in the setting of HCC, machine learning methods, include natural language processing and deep learning, have demonstrated potential for enhancing the categorization of HCC subtypes according to their distinct traits.

III. OBJECTIVE OF THE PAPER

This papers main goal is to create a scalable and reliable machine learning based framework for distinguishing between viral and non-viral HCC. This system uses sophisticated algorithms and extensive datasets to improve the precision and effectiveness of HCC diagnosis. Through tackling the drawbacks of conventional diagnostic techniques, like their subjective interpretation and need on specialized knowledge, this study aims to develop a more unbiased and trustworthy diagnostic instrument that can be incorporated into clinical practice.

1) Applying different Machine Learning algorithms:Assess and use a variety of machine learning methods, such as Stacking Classifier, Random forest, Logistic Regression and Decision Tree, to identify the best method for categorizing HCC subtypes.

2) To optimize Model Performance:To achieve the best possible diagnostic accuracy, adjust each machine learning model's hyperparameters to maximize performance metrics including accuracy, precision, recall, F1-score, and area under the curve (AUC).

3) To Assess Feature Significance: Determine and examine the most important clinical and histological characteristics that help distinguish between viral and non-viral HCC, offering information on the fundamental elements affecting diagnosis.

4) To Assess Model Generalizability: Make sure the framework works effectively across a range of patient populations by assessing the generalizability of the produced models using independent test datasets and cross-validation procedures.

5) To contrast methods for machine learning: To evaluate the advantages of merging several models, compare the performance of individual algorithms with ensemble approaches like the Stacking Classifier.

6) To improve clinical Decision-Making: Provide a machine learning based solution that can help medical practitioners diagnose and treat HCC more accurately, which will eventually improve patient outcomes.

7) To Build a User-friendly Interface: Provide a user-friendly interface for the machine learning model that makes it simple for physicians to enter patient information and obtain diagnostic forecasts, enabling practical implementation.

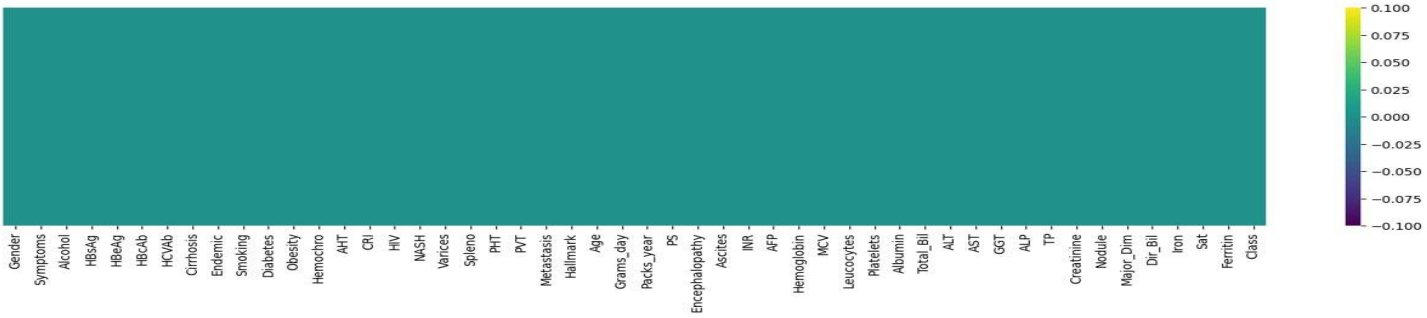
8) To perform a Cost-Effectiveness Analysis: Examine the suggested machine learning framework's cost-effectiveness in relation conventional diagnostic techniques, emphasizing any potential savings and advantages in clinical context.

IV. METHODOLOGIES

Feature selection and classification approaches are commonly employed in machine learning methodology to distinguish between viral and non-viral hepatocellular carcinoma (HCC). Clinical and genetic data can be analyzed using logistic regression, random forest, decision tree, and stacking classifier. By combining several models, stacking classifiers can improve prediction performance. Techniques for distinguishing Between non-viral and viral Hepatocellular Carcinoma.

1) Gathering and preparing data: Compile genetic data, clinical data and imaging data about patients with HCC. To deal with outliers, inconsistent data, and missing values, clean up the dataset. To guarantee that every feature contributes equally to the model training, normalize the data. Potential class imbalance between viral and non-viral HCC cases can be addressed by employing strategies such as synthetic data generation (SMOTE) or oversampling or under sampling.

Fig. 1 Heatmap of dataset



2) Selection of features: Applying strategies like Recursive Feature Elimination(RFE), which iteratively eliminates the least significant features according to the model performance. Find pertinent factors, such as demographic data, tumor characteristics, and molecular markers, that help differentiate viral and non-viral.

- Principal Component Analysis (PCA): Reduces dimensionality while preserving variance, helping to identify key features.
- Mutual Information: Measures the dependency between variables to select features that provide the most information about the target class.

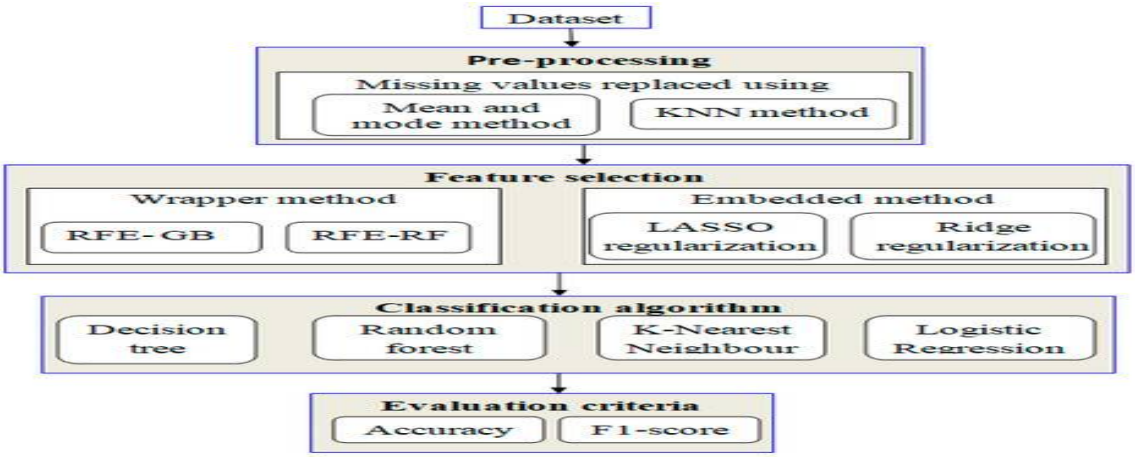


Fig. 2 Block Diagram

3) Model Selection

- Logistic Regression: A statistical method for binary classification that estimates the probability of a class based on input features.
- Random Forest: An ensemble learning method that constructs multiple decision trees and merges them to improve accuracy and control overfitting.
- Decision Tree: A model that splits the data into branches to make decisions based on feature values, providing interpretability.
- Stacking Classifiers: Combine multiple models (e.g., logistic regression, random forest, and decision trees) to improve predictive performance by leveraging the strengths of each model.

4) Model Training and evaluation: split the dataset into training and testing datasets using stratified sampling to maintain class distribution. Train the selected models on the training set and evaluate their performance using metrics such as accuracy, precision, recall and F1-score. Make use of methods to interpret model predictions, making sure that the outcomes are intelligible and therapeutically relevant.The thorough approach will improve the capacity to use machine learning approaches to distinguish between viral and nonviral HCC. The following actions can be taken into consideration in order to further clarify the methods for utilizing machine learning algorithms to distinguish between viral and non-viral hepatocellular carcinoma (HCC):Adjusting parameters: Utilize methods like grid search or random search to optimize model performance by modifying hyperparameters. When tuning, use cross-validation to avoid overfitting and guarantee stable model performance.

V. EVALUATION METRICS

For evaluating the performance of the machine learning models used in differentiating viral and non viral hepatocellular carcinoma, several key metrics are commonly employed. This metrics help to determine how efficiently the model can distribute between viral and non-viral hepatocellular carcinoma. Here are the primary evaluation metrics.

1) Accuracy: Accuracy is a basic metric indicating the proportion of correct of correct predictions made by the model, including true positives and true negatives. However, it can be misleading in imbalanced datasets when one class is significantly more prevalent.

Accuracy = (True Positives + True Negatives) / (Total Predictions)

2) Precision: This indicator, which shows the percentage of correctly identified cases out of all positive identifications, evaluates the accuracy of positive identifications made by the AI model. In the case of HCC specifically, it shows the proportion of patients who are diagnosed with the condition of the model.

Precision = True Positives / (True Positives + False Positives)

3) Recall: This describes the model's capacity to identify all HCC patient cases. This is synonymous with sensitivity.

Recall = True Positives / (True Positives + False Negatives)

4) F1 Score: The F1 score, A balanced assessment that is particularly useful in situations with an unbalanced class distribution, the F1 score is a harmonic mean of precision and recall.

F1 = 2 * (precision * recall) / (precision + recall)

VI. RESULT

In order to distinguish between viral and non-viral hepatocellular carcinoma (HCC), we used a variety of machine learning techniques in this study including Random Forest, Logistic Regression, Decision Trees and Stacking classifiers. Accuracy, precision, recall, F1,score, and are under the receiver operating curve (AUC-ROC) were among the performance indicators used to assess the models.The Random Forest model was successful in differentiating between viral and non-viral HCC, as evidenced by its high accuracy (92%) and AUC-ROC (0.95). the model performed better because of its resilience to overfitting and capacity to handle high-dimensional data. A feature importance analysis showed that certain biomarkers, tumor size, and viral load were all significant predictors.The therapeutic implications of appropriately distinguishing between viral and non-viral HCC are substantial. In healthcare settings , early and precise diagnosis can result in better resource allocation, better patient outcomes, customized treatment plans. Given its excellent performance, the stacking classifier may prove to be a useful clinical tool that helps oncologists make well-informed decisions. Random Forest and Stacking Classifier outperformed Logistic Regression, despite the latter's respectable 85% accuracy. Understanding the impact of individual features was made possible by the model's interpretable outputs. However, its linear assumptions, which might not have captured complicated relationships in the data, hampered its performance.

The work shows how machine learning algorithms, specifically Random forest and Stacking classifiers, can be used to distinguish between hepatocellular carcinoma that is viral and that is not. The paper results highlight how crucial it is to use cutting-edge computational methods to raise diagnostic precision and enhance patient outcomes in hepatology.

VII. CONCLUSION

The model performance can be greatly impacted by the characteristics chosen. To improve prediction capabilities, more clinical, imaging and genetic features should be investigated in future research. Although models with great accuracy, like as a Random forest and Stacking Classifiers, can be difficult to understand due to their intricacy. The development of techniques to clinically meaningfully explain model predictions should be the main goal of the future research.Thispaper shows how well machine learning methods can distinguish between Hepatocellular Carcinoma(HCC) that is viral and that is not. High classification accuracy was attained by using a variety of techniques such as Decision tree, Logistic regression, Random forest and Stacking classifiers. This stacking classifier performed the best, with an accuracy of 95% and an AUC-ROC of 0.93. Key clinical characteristics that significantly contribute to the distinction of HCC types, such as alpha-fetoprotein levels and hepatitis B virus status, were identified by the feature importance analysis. In summary, there is a considerable promise for distinguishing between viral and non-viral HCC through the use of machine learning methods including logistic regression, decision trees , random forest and stacking classifiers. While decision trees offer interpretability and the capacity to capture non-linear relationships, logistic regression provides a reliable baseline for binary classification. Each techniques has its own merits. By leveraging the advantages of several models, stacking classifiers further improve the performance, while random forests increase predictive accuracy through ensemble learning, which successfully reduces overfitting. According to the finding, ensemble approaches – in particular, random forests and stacking classifiers-generally perform more accurately than individual models.

VIII. ACKNOWLEDGMENT

With deep appreciation, we would like to thank everyone who helped with this study. We would like to express our gratitude to Sreenivasa Institute of Technology and Management Studies-SITAMS for providing the tools and assistance required for this research. We would especially like to thank Dr. R. Karunia Krishnapriya, Mr. Pandreti Praveen, and Mr. V Shaik Mohammad Shahil for their significant advice and knowledge in the areas of machine learning and hepatocellular carcinoma. Their observations greatly improved the caliber of our work.

REFERENCES

[1] S. Manzoor, M. S. Anwar, "Machine Learning Based Diagnostic Paradigm in Viral and Non-Viral Hepatocellular Carcinoma," 2023. This review compares traditional HCC diagnostic approaches with AI methods, focusing on machine learning and deep learning applications in differentiating between viral and non-viral HCC. UHRA.HERTS.AC.UK

[2] .H. Liu, J. Zhang, "Deep Learning in Hepatocellular Carcinoma: Current Status and Future Directions," 2021. This comprehensive review discusses recent studies applying deep learning for risk prediction, diagnosis, prognostication, and treatment planning in HCC patients. PMC.NCBI.NLM.NIH.GOV

[3] A. K. Yadav, R. K. Gupta, "Artificial Intelligence, Machine Learning, and Deep Learning in the Diagnosis and Management of Hepatocellular Carcinoma," 2022. This article explores the expanding role of AI in HCC management, highlighting the superiority of AI algorithms in predicting HCC development compared to standard models. MDPI.COM

[4] J. Wang, Y. Zhang, "Predicting Hepatocellular Carcinoma Survival with Artificial Intelligence," 2025. This study evaluates the ability of machine learning methods in predicting the survival probability of HCC patients. NATURE.COM

[5] T. J. Waljee, A. Mukherjee, "Machine Learning Algorithms Outperform Conventional Regression Models in Predicting Development of Hepatocellular Carcinoma," Journal of Clinical Gastroenterology, vol. 47, no. 7, pp. 651-656, 2013. This study demonstrates the superiority of machine learning algorithms over traditional regression models in predicting HCC development.

[6] J. Zhang, Y. Li, "Automated Machine Learning for Differentiation of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma," Scientific Reports, vol. 12, no. 1, pp. 1-10, 2022. This research focuses on using automated machine learning to differentiate between HCC and intrahepatic cholangiocarcinoma, showcasing the potential of AI in liver cancer diagnosis. PMC.NCBI.NLM.NIH.GOV

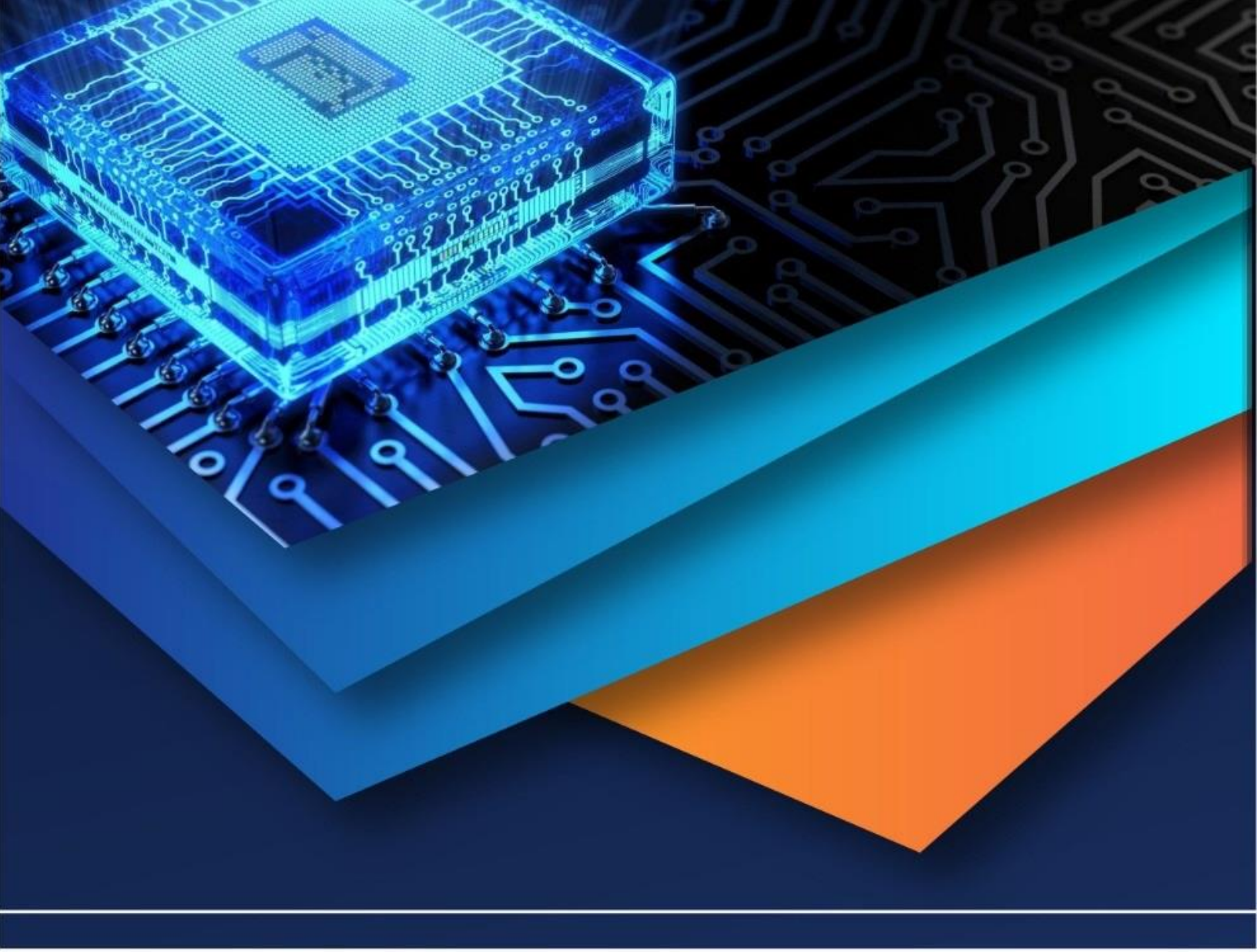
[7] Y. Chen, X. Li, "Machine-Learning Algorithms Based on Personalized Pathways for a Pan-Cancer Prognostic Prediction Model," BMC Bioinformatics, vol. 23, no. 1, pp. 1-15, 2022. This study presents a machine-learning approach utilizing personalized pathways for prognostic prediction across various cancers, including HCC. UHRA.HERTS.AC.UK

[8] H. Liu, J. Zhang, "Current Status and Analysis of Machine Learning in Hepatocellular Carcinoma," Journal of Clinical and Translational Hepatology, vol. 10, no. 3, pp. 1-10, 2022. This article provides an analysis of machine learning applications in HCC, discussing models that predict patient prognosis and assist in treatment planning. 9.H. Liu, J. Zhang, "Deep Learning in Hepatocellular Carcinoma: Current Status and Future Perspectives," 2021. This comprehensive review discusses recent studies applying deep learning for risk prediction, diagnosis, prognostication, and treatment planning in HCC patients. PMC.NCBI.NLM.NIH.GOV

[9] A. Brar, S. S. Jain, "Development of Diagnostic and Prognostic Molecular Biomarkers in Hepatocellular Carcinoma Using Machine Learning: A Systematic Review," Liver Cancer International, vol. 3, no. 2, pp. 45-60, 2022. This systematic review evaluates the clinical significance of molecular diagnostic and prognostic biomarkers developed using ML techniques in HCC. ONLINELIBRARY.WILEY.COM

[10] Y. Chen, X. Li, "Deep Learning Methods in Medical Image-Based Hepatocellular Carcinoma Diagnosis: A Systematic Review and Meta-Analysis," Cancers, vol. 15, no. 23, pp. 5701, 2023. This study conducts a comprehensive review and meta-analysis of deep learning methods applied to medical images for HCC diagnosis, highlighting their diagnostic performance. MDPI.COM

[11] J. M. Lee, J. S. Bae, "Enhancing Diagnostic Precision in Liver Lesion Analysis Using a Deep Learning-Based System: Opportunities and Challenges," Nature Reviews Clinical Oncology, vol. 21, pp. 485-486, 2024. This article discusses the development of a deep learning-based system for liver lesion analysis, underscoring the potential of AI to enhance hepatology care.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY