Predicting Liver Cancer Staging Using Ensemble Learning

Ms.Sinduja S, Dr.Mohanapriya N, Ms.Nithya sree R S, Ms.Rahini N, Ms.Yazhini R
Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women Namakkal, India

***Abstract*- Prediction from the Liver Cancer Conference (PLCS), an important indicator for evaluating the development of primary liver cancer cells (PLCCS), is extremely important in the diagnosis, treatment and rehabilitation of liver cancer. This project introduces an integrated system that integrates CNN, KNN, and Gemini AI to improve liver cancer prediction and control.**

This system is based on the CNN model of CT-SCAN via the CNN model that predicts stadiums (0, A, B, C, D) based on the BCLC staging system. The KNN model checked and fine-tuned predictions by examining information on clinical and radiation reports based on the BCLC guidelines. Gemini AI recommends patient-specific treatment schemes to optimize patient care with the potential for resection, ablation, implantation, and systemic therapy.

Additionally, the web-based portal offers health instructions and prefabricated YouTube videos for patient training and access to resources. The combination of CNN, KNN, and Gemini AI improves classification accuracy and treatment suggestions, while the interactive portal increases patient awareness and participation. Experimental results demonstrate the effectiveness of the system in planning and treatment planning for liver cancer, making it a useful tool for clinicians and improving patient outcome.

I. INTRODUCTION

Primary liver cancer (PLC) accounts for the third largest cancer-related mortality rate in the world, with over 830,000 deaths per year. Proper staging of liver cancer is essential for choosing the best treatment options and predicting patient survival. Traditional staging systems such as Tumour Node Metastasis (TNM) and China Liver Cancer staging system (CNLC) are time-consuming, subjective and false, as they are based on subjective evaluations of imaging and clinical reporting by radiologists.

Unstructured complexity radiology reports and the non-uniform availability of medical resources also complicate effective staging. To overcome these limitations, machine learning (ML) and deep learning (DL) methods have picked up pace in cancer staging. Though image-based methods based on convolutional neural networks (CNN) have delivered encouraging results in tumour classification, they tend to demand large annotated datasets and manual annotation.

This work presents an ensemble learning paradigm of CNN and KNN for the task of automatic liver cancer staging. Deep features from medical images are learned using CNN, while cancer stage classification is performed by KNN from radiology report information. By using both image-based and text-based information, the suggested approach is improved in terms of classification accuracy and robustness.

The experimental results prove that the ensemble model is better than the use of a single classifier, yielding high accuracy and reliability in liver cancer staging. The system also enhances clinical benefits by providing stage-specific treatment suggestions and a web-based interface

II.RELATED WORKS

Liver cancer - a critical process in selecting the right treatment strategy and patient predictions. Traditional staging techniques, including TNM and BCLC classification systems, are based on medical imaging, clinical factors, and histopathological studies. However, these traditional methods suffer from subjectivity, time limits and heterogeneity of expert opinions. To address these challenges, a recent study investigated machine learning (ML) and deep learning (DL) to improve the accuracy and efficiency of liver cancer detection and staging.

Medical imaging is essential for staging and diagnosis of liver cancer. Deep learning-based models, including folding networks (CNNS), were widespread for segmentation and tumour classification. Napte et al. Automatic liver cancer

detection (ALCD) via a deep folding network (DCNN) to improve liver segmentation in computed tomography (CT) scans.

The proposed model, parallel UNET (ESP-UNET) of edge reinforcement (ESP-Unet), effectively dissolves the problem in subsequent excess problems due to noise and tumour structure variation, leading to improved classification accuracy. Other researchers are considering a variety of CNN architectures to improve cancer detection. For example, bread etc.

CNNs associated with the classification of brain tumours from multiphase-MRI-scans demonstrate the usefulness of deep learning in analyzing medical imaging. Fans etc. Another study also introduced a multitasking learning model combining the radiation properties of MRI images to predict the degree of tumour and KI-67 expression levels with high diagnostic accuracy.

Although successful, depth-based image analysis techniques are usually hampered by a large amount of marked data for high-quality medical imaging data, domain adaptation problems, and dependency requirements. To avoid these limitations, researchers are looking for multi-source data fusion and hybrid machine learning methods. Multisource integration improves robustness, but introduces the complexity and challenges of property selection and merger methods.

This led researchers to examine alternative approaches such as natural language processing (NLP)-based text analysis.Various studies have highlighted several combinations of data sources to improve classification performance. Zhang et al. Based on B-mode ultrasound and contrast-enhanced ultrasound (CEU), we passed the Multikernel Support Vector Machine Plus (MKL-NHSVM+) model for the diagnosis of liver cancer.

The model was used to learn multi-source transfers so that it could learn from different imaging modalities and improve diagnostic performance. It was suggested that multi-view feature conversion-based SVM+ (MFSVM+) models also include multi-phase image data. This model used a feature-transforming approach to improve the ability of support vector machines (SVMs) in the treatment of multimodal imaging functions, which significantly improved the classification of liver cancer. Medical imaging is frequently followed by radiology reports with important staging data in unstructured text form. Conventional ML models are ineffective in extracting significant features from lengthy and complicated radiology reports, and hence NLP-based techniques become crucial.

Li et al. They hired LLM-based Knowledge Network (LLM) (LKA) (LKA) to combine large-scale language models (LLM) with hierarchical attention mechanisms. Your text. This model achieved an impressive accuracy of 90.3%, the value of NLP in liver cancer automation. The deep learning and NLP models have also used other studies to process text-based clinical data. Fink et al.

We created a pulmonaryembolism assessment system from a structured radiologist who demonstrates the potential of NLP in supporting clinical decision-making. D'Anballe et al. The rule-based algorithms used (RBA) and the rule-based algorithms (RBA) used for multi-disease classification from radiologists and attention-driven repeating neural networks (RNNS) further confirm the need for NLP in the case of automated cancer.

In addition to medical imaging and textual analysis, genomic data for the discovery of important biomarkers for the staging and development of liver cancer was extremely important. Deng and Guo discovered important genes involved in liver cancer progression using network-based biomarker discovery based on protein interaction networks and differential gene expression analysis.

Important genes such as TOP2A, ESR1, and KMO were presented as important results in studies that were identified as potential biomarkers in previous liver cancer studies. By using genomic information in ML models, treatment of liver cancer can potentially be adapted by revealing individual molecular signatures of each patient. Challenges such as data heterogeneity, computing concentration requirements, and the need for large-scale biological validation are significant obstacles to the fact that these methods are practiced in clinical settings.

Singh et al. We proposed a k-nearest Neighbour (KNN)-based technology for text classification. The proposed technique used the advantages of KNN to manage high-dimensional data and maintain a high level of accuracy for text classification tasks. This model is trained using a set of 1000 text documents with different categories and holds a high accuracy rate of 94.5% in text documents classification. The

study shows the performance of KNNs in the event of text classification problems and how they can be used for medical text processing.

This solution has the promise to improve the accuracy of text classification for medical applications with less chance of human error and improved patient outcomes. One of the main strengths of KNN is that it is extremely flexible under these circumstances and is able to use a variety of domains, such as: B. Medical text processing. KNN does not include complex training processes and instead, since I have sincerely learned all the training cases, it is easy to include changes to new medical vocabulary and classification requirements.

Kim et al. (2023) AL-controlled clinical decision support system (CDS) for liver cancer. The model solves personalization problems by adapting the therapy to a specific patient profile, including age, comorbidities, and biomarker expression. With clinical guidelines such as the staging system for Barcelona Clinical Lever Cancer (BCLC) and integration of multimodal data (e.g. medical imaging, genomic databases, electronic health records), the CDSS provides evidence-based, patient-specific advice

III.EXISTING SYSTEM

The existing model follows Large Language Model to automate liver cancer staging by analysing radiology reports.The system uses a hierarchical attention network (HAN) to identify important words and phrases within radiology reports. The extracted key terms related to staging, such as tumor size, vascular invasion, metastasis, and number of lesion are used for classification model that predict the liver cancer.The output stage categories include China liver cancer staging method like Stage IIa, IIb, IIIa, and IIIb.The current method for predicting the stage of liver cancer mostly uses radiology and medical imaging records, and doctors manually review patient databases.

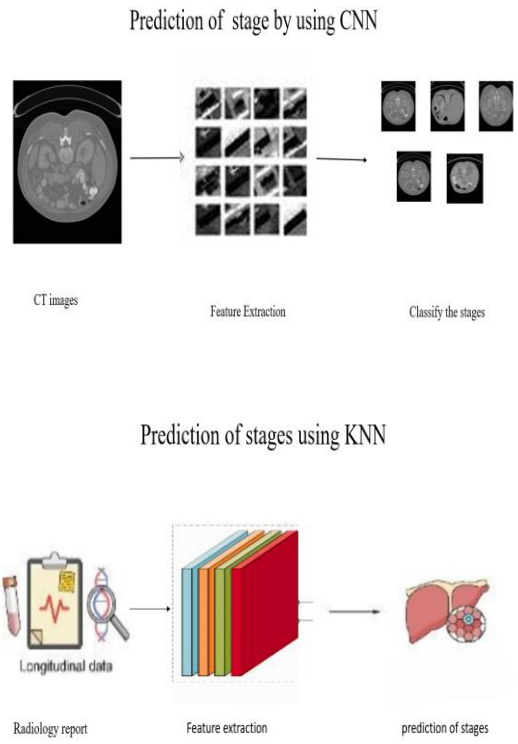
This process is time-consuming and subjective due to the unstructured nature of medical reports. One major disadvantage is that it depends on high-quality data for training, which might be difficult to get, especially in early-stage liver cancer cases.Liver cancer is often asymptomatic in its early stages, leading to late detection when symptoms become more noticed. As a result, hospitals and research institutions have fewer recorded cases of early-stage liver cancer, making it difficult to collect sufficient training data for Al models. Due to the lack of sufficient training samples, the model may struggle to distinguish early-stage cancer from normal or benign liver conditions, increasing false-negative rates.

IV.PROPOSED SYSTEM

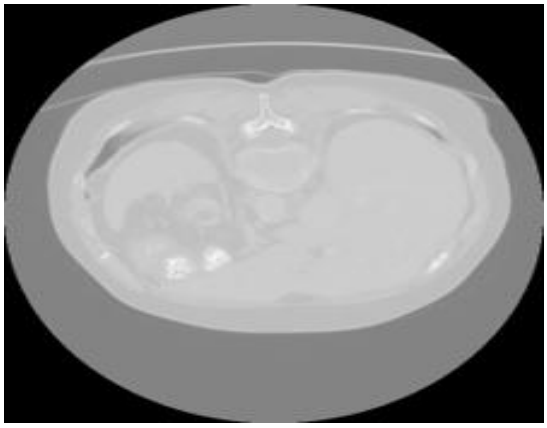
The proposed system provides a solution by incorporating technologies like deep learning algorithm and machine learning algorithm for liver cancer prediction.The CNN (Convolutional Neural Network) will classify medical images CT of the live and predict the stage of liver cancer (e.g., Stage o, Stage A, Stage B, Stage C, Stage D). The KNN (K-Nearest Neighbors) algorithm will be used to interpret medical reports and categorize the severity or provide recommendations based on patient data.The system generates automated diagnosis reports, ensuring precision and consistency. Additionally, the output includes liver cancer staging and corresponding treatment recommendations, assisting medical professionals in decision-making.

Gemini AI will perform this information and propose appropriate treatment options based on the BCCLC guidelines (hepatic cancer) for Barcelona Clinic Lever Cancer (BCCLC). The system suggests the appropriate type of intervention, for example, depending on the size, distribution of the tumor, and the condition of the patient. Corresponding type, resection, ablation, implantation, or systemic treatment. This module integrates Gemini AI through APIs and provides real-time consumption recommendations on a real-time basis. This improves the clinician's decision-making process and provides patients with accurate medical information and advice.

Additionally, a web-based interface enhances patient support by offering healthcare tips and curated YouTube videos on liver cancer treatment, medication, and lifestyle improvements. By integrating Gemini AI via API, this module ensures real-time, AI-driven treatment suggestions, improving decision-making for clinicians while empowering patients with reliable medical information and guidance.



SAMPLE DATASET IMAGE



V.METHODOLOGY

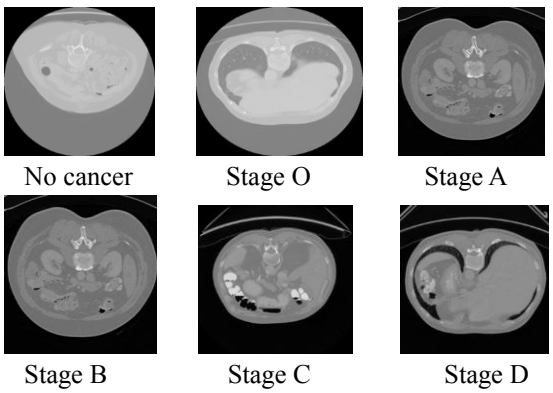
Module 1: Staging of cancer using CNN for imaging

The CNN model used to predict the stage of liver cancer uses many layers. When processing CT scan photos, each layer is extremely important. The model receives a normalized image, evenly processed at the entrance layer. The data then passes through the folding layer. This layer extracts important features such as tumor size, tumor number, metastasis, and vascular invasion. These properties are handled by rearranging layers that add nonlinearity so that the model can recognize complex patterns.

The maximum pooling layer is to complete the functional card, reduce dimensions and maintain relevant tumor-related information. In other words, the model is more invariant with tumor changes. After extracting features, the data is flattened and fully connected layers are fed, and the relationship between the extracted properties and the specific stage of liver cancer is learned. To avoid excessive adaptation, the layer of impairment randomly switches several neurons, achieving improved generalization from training to new instances. Finally, the output layer uses SoftMax's activation function to provide probability values for different liver cancer stages (BCLC levels 0, A, B, C, D), whereby the highest probability rating determines the final classification. This organized method allows for accurate and automated staging of liver cancer, facilitating faster and more reliable diagnosis.

Module 2: Staging of cancer using KNN for radiology report

This module predicts liver cancer sclerosants based on radiologists and KNN algorithms (K-nearest Neighbors). The process begins with searching for important biomarkers from radiologists such as bilirubin, albumin, AFP values, tumor size, number of nodes, vascular invasion, ECOG performance status, and child peus score. These clinical features have been pre-processed, normalized and formatted to maintain uniformity. The KNN algorithm is then used to predict cancer stages by adapting new patient information to existing marked cases. This model identifies nearest neighbors through similarity and describes patients with the most common BCLC levels (0, A, B, C, D). This module improves staging accuracy by combining structured clinical information with machine learning to provide consistent and effective predictions of liver cancer.



Module 3: Treatment Proposal

The third module is a Gemini AI-mediated treatment proposal. It is connected to the system by API to make personalized treatment suggestions based on the predicted stage of liver cancer. As soon as staging is defined by CNN-based image classification (module 1) and KNN-based radiation report analysis (module 2), Gemini AI will perform this information and propose appropriate treatment options based on the BCCLC guidelines (hepatic cancer) for Barcelona Clinic Lever Cancer (BCCLC). The system suggests the appropriate type of intervention, for example, depending on the size, distribution of the tumor, and the condition of the patient. Corresponding type, resection, ablation, implantation, or systemic treatment. This module integrates Gemini AI through APIs and provides real-time consumption recommendations on a real-time basis. This improves the clinician's decision-making process and provides patients with accurate medic al information and advice.

We provide a web-based interface enhances patient support by offering healthcare tips and curated YouTube videos on liver cancer treatment, medication, and lifestyle improvements. By integrating Gemini AI via API, this module ensures real-time, AI-driven treatment suggestions, improving decision-making for clinicians while empowering patients with reliable medical information and guidance.

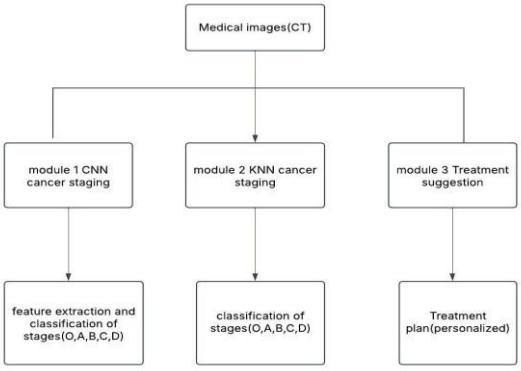


Fig: overall process of proposed system

VI. CONCLUSION

Liver cancer is a critical public health issue requiring early diagnosis, accurate staging, and tailoring treatment for improved patient outcomes. Traditional approaches are usually based on interpretations of medical images and people's reports, which can be delayed, inconsistencies in the diagnosis, and even mistakes. This initiative suggests the use of AI-based methods to integrate machine learning, deep learning, and AI-based decision-making to promote the accuracy, efficiency and patient-oriented delivery of liver cancer management. According to the Barcelona Clinic Lever Cancer (BCLC) system. This minimizes the risk of human error and supports high-precision, consistent predictions that improve clinical decisions. The system also uses Gemini AI to create patient-specific treatment proposals according to general medical guidelines and patient factors. A special website has been created to provide YouTube and nutritional advice for liver cancer patients. In this way, patients are well informed about their condition and are able to actively participate in health care.

This AI-controlled system minimizes dramatic diagnostic errors, accelerates treatment plans, and optimizes treatment choice based on real-time evidence and best congressional guidelines. By integrating computer-aided image analysis, AI-based classification, and patient-specific treatment plans, this strategy improves both the efficiency of health workflows and patient outcomes. Furthermore, the mixing of AI and medical experience improves the quality of oncology care, making detection and treatment of liver cancer more detailed, more accurate and efficient.

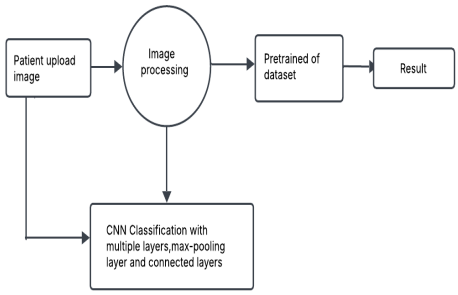


Fig: Image processing and feature extraction

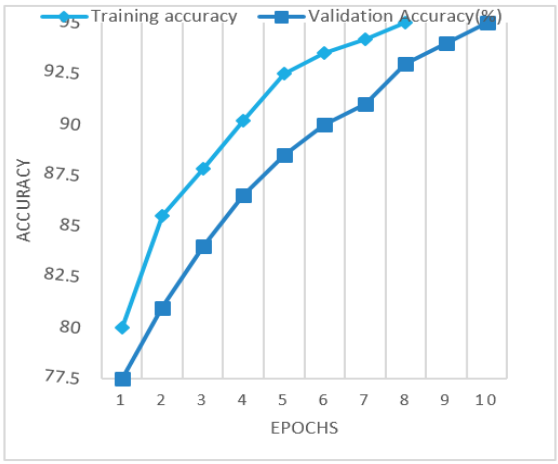
ultimately improving patient outcomes and promoting precise oncology. The main benefits of the introduced framework are improved patient outcomes, better clinical decisions, and improved efficiency. Instructions for future work include further data sources such as genomic information and electronic health files, as well as the expansion of the frame into other forms of cancer. The introduced ensemble frame is a major advance in the use of AI-based technologies in the treatment of liver cancer, providing a scalable and holistic approach to improving diagnosis, staging and treatment planning

Furthermore, the proposed framework can revolutionize oncology by enabling personalized medicine, improving treatment outcomes, and reducing the costs of the healthcare system. Using AI technology in conjunction with clinical experience can lead to improved decision-making, increased patient care, and improved quality of life for cancer patients. Additionally, frame scalability and flexibility make it a practical solution for healthcare systems that aim to improve cancer care and outcomes.The proposed framework could transform the field of oncology by enabling personalized medicine, improved outcomes, and reduced health costs. Integrating AI-controlled technologies with clinical expert knowledge can lead to improved decision-making, improved patient care, and improved quality of life for cancer patients. Additionally, frame scalability and flexibility make it an attractive solution for healthcare systems that want to improve cancer treatments and outcomes.

In summary, this project formulates an end-to-end system for liver cancer prediction and care based on CNN, KNN, and Gemini AI. CNN predicts liver cancer stages from CT scans based on the BCLC staging system, whereas KNN enhances the predictions through examination of radiology reports. Gemini AI, which is API connected, recommends personalized treatment strategies, and a web interface provides health tips and YouTube videos to inform patients. The system improves precision, minimizes errors, and speeds up clinical decision-making, offering a holistic and scalable solution for better liver cancer management.

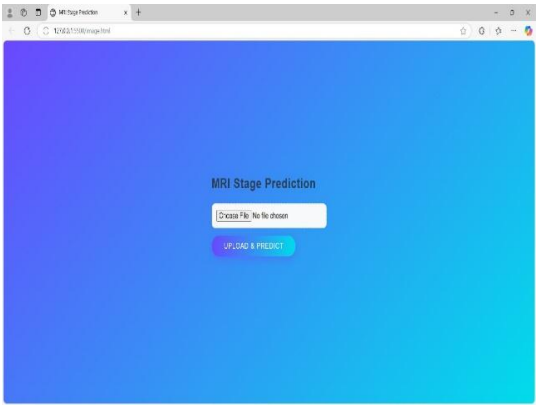
Future work will focus on further improvements to the model by adding additional data sources such as genomic data and applying them to other cancer types. The proposed framework represents a major advance in using AI-controlled techniques for liver cancer management, providing a comprehensive and scalable solution for improving diagnosis, staging and treatment planning.

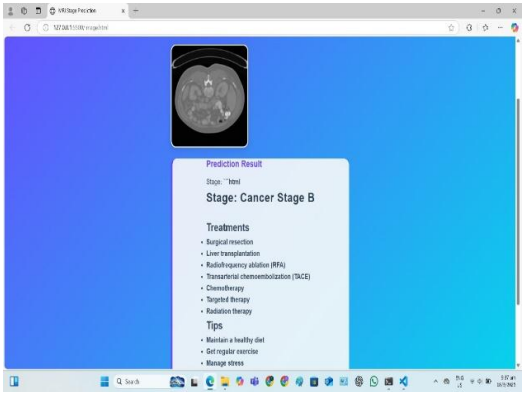
In summary, "Predicting liver cancer using an ensemble model" is a major advance in improving the accuracy of liver cancer detection and treatment. The system uses a three-module ensemble frame that integrates CNNs into image processing. This leads to a robust solution for effective treatment of liver cancer and improved patient improved patient outcome



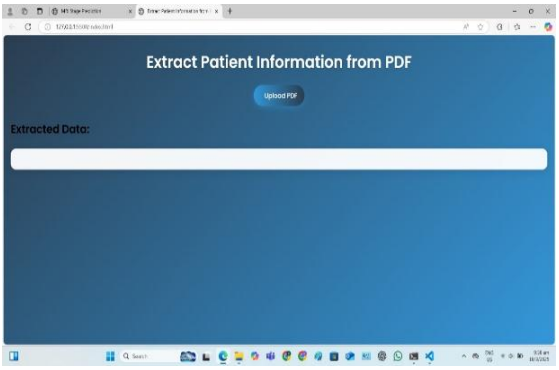
Training vs Validation Accuracy

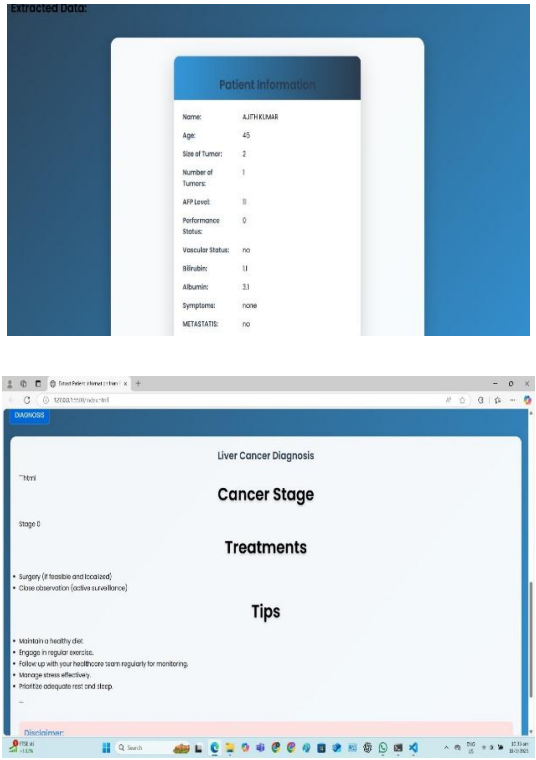
OUTPUT





(FIG: Stage prediction by using CNN)





(FIG: Stage prediction using KNN}

REFERENCES

[1] Li, Y.; Zheng, X.; Li, J.; Dai, Q.; Wang, C.-D.; Chen, M. LKAN: LLM-Based Knowledge-Aware Attention Network for Clinical Staging of Liver Cancer. IEEE J. Biomed. Health Inform. 2024.

[2] A Machine Learning Framework for Early Prediction of Breast Cancer. "Patel, R.; Mehta, S.; Joshi, A." Comput. Biol. Med. 2020, 120, 103754. [Google Scholar] [CrossRef]

[3] An End-to-End Deep Learning Approach for Lung Cancer Detection from CT Scans. "Liu, L.; Wang, R.; Xie, C.; Yang, P.; Wang, F.; Sudirman, S.; Liu, W." IEEE Access 2019, 7, 45301–45312. [Google Scholar] [CrossRef]

[4] Early Detection of Skin Cancer Using Convolutional Neural Networks. "Alfarisy, A.A.; Chen, Q.; Guo, M." In Proceedings of the 2018 International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 21–25. [Google Scholar]

[5] A Comprehensive Review on Machine Learning Techniques for Early Cancer Prediction. "Kumar, R.; Singh, P.; Gupta, S." J. Cancer Res. 2019, 20, 789–805. [Google Scholar] [CrossRef]

[6] Real-Time Cancer Detection Using Deep Learning and Medical Imaging. "Zhang, Y.; Li, X.; Wang, H.; Chen, Z." Med. Image Anal. 2020, 62, 101665. [Google Scholar] [CrossRef]

[7] Advanced Techniques for Early Detection of Colorectal Cancer. "Balleda, K.; Satyanvesh, D.; Sampath, N.; Varma, K.; Baruah, P." In Proceedings of the 2014 IEEE 8th International Conference on Bioinformatics and Biomedicine, Belfast, UK, 2–5 November 2014; pp. 262–268. [Google Scholar]

[8] Machine Learning Models for Early Prediction of Prostate Cancer. "Wadhai, M.; Gohokar, V.; Khaparde." In Proceedings of the 2015International Conference on Health Informatics, Lisbon, Portugal, 14–16 October 2015; pp. 544–547. [Google Scholar]

[9] Deep Learning for Automated Detection of Early-Stage Cancer in Histopathology Images. "Xia, D.; Chen, P.; Wang, B.; Zhang, J.; Xie, C." J. Pathol. Inform. 2018, 9, 4169. [Google Scholar] [CrossRef] [Green Version]

[10] A Survey on Deep Learning Techniques for Early Cancer Detection. "Chen, L.; Wang, J.; Li, H." Sensors 2022, 22, 5678. [Google Scholar] [CrossRef]

[11] Application of Convolutional Neural Networks for Early Detection of Oral Cancer. "Gupta, A.; Sharma, V.; Kumar, P." J. OralPathol.Med 2020,49,123-135[google scholar]

[12] Early Detection of Pancreatic Cancer Using Machine Learning. "Ding, W.; Taylor, G." Comput. Biol. Med. 2016, 123, 17–28. [Google Scholar] [CrossRef] [Green Version]

[13] Deep Learning Techniques for Early Prediction of Ovarian Cancer. "Wang, X.; Liu, Y.; Zhang, Z." IEEE Trans. Biomed. Eng. 2021, 68, 2345–2356. [Google Scholar] [CrossRef[

[14] A Review on Early Detection of Brain Tumors Using Machine Learning. "Li, Y.; Zhang, X.; Wang, Q." J. Neurosci. Methods 2021, 180, 105890. [Google Scholar] [CrossRef]

[15] Early Detection of Cervical Cancer Using Deep Learning Models. "Naeem Ullah; Javed Ali Khan; Lubna Abdulaziz Ahlarbi; Asaf Raza Wahab Khan." In Proceedings of the 2022 International Conference on Health Informatics, Lisbon, Portugal, 14–16 October 2015; pp. 544–547. [Google Scholar]