SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 3, 2026

DEEP VISION-BASED MEDICAL IMAGE PROCESSING WITH TRANSPARENT AI DIAGNOSTIC INTELLIGENCE FOR EARLY PAN-CANCER DETECTION, TUMOR CHARACTERIZATION AND BIOMARKER DISCOVERY AMONG SOUTH ASIAN POPULATIONS

Ammara[1], Dr. Azeem Hayat[2], Dr. Syed Mobasher Ali Abid[*3], Dr. Asmatullah[4], Muhammad Umair Aslam[5]

[1]*School of Information Science and Engineering, Yanshan University, Qinhuangdao, 066004 China.*
[2]*Allied Health Sciences, University and Institute of Public Health, University of Lahore, Pakistan.*
[*3]*Department of Pharmacy, COMSATS University Islamabad, Abbottabad Campus, Pakistan.*
[4]*Gomal Medical College, Dera Ismail Khan, Pakistan.*
[5]*Department of Computer Science, University of Gujrat, Gujrat, Pakistan.*

[1]ammaraarshad@stumail.ysu.edu.cn,[2]azeemhayat74@gmail.com,[*3]mobasher@cuiatd.edu.pk, [4]asmat4139@gmail.com, [5]uog0304@gmail.com

[1]**ORCID:** 0009-0004-3838-0878

DOI: https://doi.org/10.5281/zenodo.21350017

Keywords
Medical Image Processing; Deep Vision; Transparent AI; Pan-Cancer Detection; Tumor Characterization; Biomarker Discovery; Explainable Deep Learning; South Asian Populations

Article History
Received: 01 January 2026
Accepted: 17 March 2026
Published: 31 March 2026

Corresponding Author: *
Dr. Syed Mobasher Ali Abid

Abstract

Early cancer detection remains a major clinical challenge in South Asian populations due to delayed diagnosis, heterogeneous disease presentation, limited screening coverage, and population-specific genetic and lifestyle risk factors. This study proposes a deep vision-based medical image processing framework integrated with transparent AI diagnostic intelligence for early pan-cancer detection, tumor characterization, and biomarker discovery among South Asian populations. The study used multi-center retrospective medical imaging data collected from publicly available cancer imaging repositories and South Asian hospital-based diagnostic records, including histopathology images, magnetic resonance imaging, computed tomography scans, and ultrasound images representing breast, lung, colorectal, liver, cervical, and oral cancers. After ethical anonymization, the dataset was preprocessed using image enhancement, noise reduction, normalization, segmentation, augmentation, and region-of-interest extraction techniques. Python was used as the main programming environment, while TensorFlow, Keras, PyTorch, OpenCV, Scikit-learn, NumPy, Pandas, Matplotlib, and Google Colab were used for model development, image processing, training, visualization, and performance evaluation. The proposed framework employed convolutional neural networks, vision transformer-based feature extraction, and hybrid deep learning classifiers to identify cancerous patterns across multiple imaging modalities. Explainability was incorporated using Grad-CAM, SHAP, and LIME to highlight clinically relevant tumor regions and support transparent decision-making. Biomarker discovery was performed by linking image-derived features such as texture, shape, lesion boundary irregularity, vascular patterns, and tissue

heterogeneity with diagnostic cancer categories and risk-level stratification. Experimental results showed that the proposed model achieved an overall accuracy of 96.4%, precision of 95.8%, recall of 96.1%, F1-score of 95.9%, and ROC-AUC of 0.982 for pan-cancer detection. Compared with conventional machine learning models and standalone CNN architectures, the proposed system improved detection accuracy by 8.2% and reduced false-positive classification by 21.6%. The explainable AI outputs confirmed that the model focused on medically meaningful tumor regions rather than irrelevant background features. These findings demonstrate that transparent deep vision-based medical image processing can support early cancer diagnosis, enhance clinical interpretability, and enable population-specific biomarker discovery for South Asian healthcare settings.

INTRODUCTION:

Cancer remains one of the leading causes of morbidity and mortality worldwide, and its burden is increasing rapidly across South Asian countries due to demographic transition, urbanization, lifestyle changes, environmental exposure, and limited access to early diagnostic services. In South Asian populations, cancers such as breast, lung, colorectal, liver, cervical, and oral cancer are frequently diagnosed at advanced stages, reducing treatment success and survival outcomes. Delayed diagnosis is often associated with inadequate screening programs, a shortage of trained specialists, socioeconomic barriers, fragmented healthcare infrastructure, and variation in disease presentation across diverse ethnic and genetic groups. Therefore, there is a growing need for intelligent, scalable, and clinically interpretable diagnostic systems that can support early cancer detection and improve decision-making in resource-constrained healthcare environments. Medical image processing plays a central role in modern cancer diagnosis because radiological and pathological images provide important visual evidence of tumor presence, morphology, progression, and tissue-level abnormalities. Imaging modalities such as magnetic resonance imaging, computed tomography, ultrasound, and histopathology are widely used to detect cancerous lesions, evaluate tumor size, assess anatomical involvement, and guide treatment planning [1]. However, manual interpretation of medical images is time-consuming and highly dependent on the experience of radiologists, pathologists, and oncologists. In South Asian healthcare settings, where patient load is high and specialist availability is limited, diagnostic delays and inter-observer variability can significantly affect clinical outcomes. These challenges highlight the need for automated medical image analysis techniques capable of improving diagnostic accuracy, consistency, and efficiency. Recent advances in deep learning have transformed medical image analysis by enabling automated feature extraction, lesion detection, segmentation, classification, and disease prediction. Convolutional neural networks have shown strong performance in identifying cancer-related imaging patterns, while vision transformer-based architectures have improved the ability to capture global contextual features and long-range spatial dependencies within complex medical images [2]. Hybrid deep learning models that combine convolutional and transformer-based representations are particularly promising for pan-cancer detection because they can learn both local texture-level features and broader structural abnormalities across multiple imaging modalities. Such models can support the detection of heterogeneous tumor patterns in breast, lung, colorectal, liver, cervical, and oral cancers. Despite these advances, many deep learning-based diagnostic models remain limited by poor interpretability, restricted generalizability, and insufficient adaptation to population-specific disease characteristics. In clinical practice, accuracy alone is not sufficient for adoption because clinicians must understand why an AI system produces a particular diagnosis. Black-box predictions may reduce trust, especially in high-

stakes cancer diagnosis, where false-positive outputs may lead to unnecessary interventions and false-negative outputs may delay life-saving treatment. Transparent artificial intelligence methods such as Gradient-weighted Class Activation Mapping, SHAP, and LIME can help overcome this limitation by identifying image regions and diagnostic features that contribute most strongly to model predictions. By visualizing tumor-relevant regions and explaining classification decisions, transparent AI can improve clinician confidence and support accountable diagnostic decision-making.

Another important research gap is the limited representation of South Asian populations in AI-based cancer imaging studies. Many existing models are trained on datasets collected from Western or high-resource healthcare systems, which may not fully reflect the genetic diversity, lifestyle-related risks, environmental exposures, imaging variability, and healthcare conditions of South Asian populations. Population-specific differences in tumor appearance, disease progression, screening access, and biomarker expression can influence model performance and diagnostic reliability. Therefore, developing AI frameworks that consider South Asian clinical and imaging contexts is essential for improving fairness, generalizability, and real-world applicability. Biomarker discovery is also becoming increasingly important in precision oncology. Conventional biomarker identification often depends on molecular, genetic, or laboratory-based testing, which may not be easily accessible in many low- and middle-income healthcare settings [3]. Deep vision-based medical image processing offers a complementary approach by extracting image-derived biomarkers from medical images. Features such as texture variation, lesion shape, boundary irregularity, vascular pattern, density distribution, and tissue heterogeneity can provide meaningful diagnostic and prognostic information. When linked with cancer categories and risk-level stratification, these image-derived biomarkers can support early detection, tumor characterization, and personalized clinical assessment. This study proposes a deep vision-based medical image processing framework integrated with transparent AI diagnostic intelligence for early pan-cancer detection, tumor characterization, and biomarker discovery among South Asian populations. The framework uses multi-center retrospective imaging data, including histopathology images, magnetic resonance imaging, computed tomography scans, and ultrasound images representing major cancer types. The proposed model combines convolutional neural networks, vision transformer-based feature extraction, and hybrid deep learning classifiers to identify cancer-associated patterns across imaging modalities. Explainable AI techniques are incorporated to ensure that model predictions are clinically interpretable and focused on medically meaningful tumor regions [4]. The main objective of this research is to design and evaluate an intelligent diagnostic framework that improves early cancer detection while maintaining transparency and clinical relevance. Specifically, the study aims to process and analyze multimodal cancer images, develop a hybrid deep vision model for pan-cancer classification, incorporate explainable AI methods for diagnostic transparency, and identify image-derived biomarkers associated with cancer type and risk stratification.

The major contributions of this study are as follows:

- It presents a deep vision-based medical image processing framework for early pan-cancer detection across multiple imaging modalities and cancer types.
- It integrates convolutional neural networks and vision transformer-based feature extraction to improve the recognition of both local and global tumor patterns.
- It incorporates transparent AI methods, including Grad-CAM, SHAP, and LIME, to explain diagnostic predictions and support clinical trust.
- It introduces an image-derived biomarker discovery mechanism based on texture, morphology, boundary irregularity, vascular patterns, and tissue heterogeneity.
- It emphasizes population-specific diagnostic intelligence for South Asian

populations, addressing a critical gap in AI-based cancer imaging research.

Overall, this study contributes to the development of clinically interpretable and population-aware AI-assisted oncology systems by combining deep vision-based image processing, transparent diagnostic intelligence, and biomarker-oriented analysis. The proposed framework is expected to support early cancer recognition, improve diagnostic confidence, reduce dependence on subjective image interpretation, and provide a scalable decision-support approach for South Asian healthcare environments where timely and accurate cancer diagnosis remains a critical clinical priority.

Deep Learning Frontiers in Oncology Image Analysis:

Deep learning has transformed oncology image analysis by introducing powerful computational models capable of automatically learning hierarchical features from large-scale and complex medical imaging datasets. Unlike conventional machine learning approaches, which depend heavily on manually extracted features, deep learning models can learn diagnostic patterns directly from raw or preprocessed images. This capability is highly valuable in cancer diagnosis because malignant changes may appear as subtle variations in texture, shape, intensity, density, tissue organization, and lesion boundaries. As a result, deep learning has become a major technological foundation for automated cancer detection, tumor segmentation, lesion classification, survival prediction, and treatment response assessment. Convolutional neural networks have been extensively applied to oncology image analysis due to their strong ability to capture spatial patterns, local textures, edges, shapes, and tissue-level abnormalities. CNN-based models have demonstrated strong diagnostic potential in breast cancer mammography, lung nodule detection, colorectal histopathology classification, liver tumor segmentation, cervical cancer screening, and oral lesion recognition [5]. In these applications, CNNs can identify cancer-associated patterns that may be difficult to detect through manual visual inspection, especially in early-stage lesions or heterogeneous tumor regions. Their ability to process large volumes of image data also supports faster diagnostic workflows and reduces the burden on radiologists, pathologists, and oncologists. The main strength of CNNs lies in their layered feature extraction mechanism. In the early convolutional layers, the model captures low-level image features such as edges, color variations, intensity changes, and texture gradients. In the middle layers, it learns more complex visual features such as tissue patterns, lesion contours, glandular structures, and local morphological abnormalities. In the deeper layers, CNNs extract high-level semantic features associated with tumor type, malignancy probability, lesion aggressiveness, and diagnostic class. This hierarchical learning process makes CNNs highly suitable for medical image processing tasks where subtle differences in visual appearance may indicate early cancer development. Deep learning models are also valuable because they can support multiple oncology imaging tasks within a single diagnostic pipeline. For example, a model may first enhance image quality, then identify the suspicious region, segment the tumor area, classify the lesion type, and finally provide risk-level prediction. This integrated capability is especially important for pan-cancer detection, where different cancer types and imaging modalities must be analyzed together. In histopathology, deep learning models can detect abnormal cell structures, nuclear atypia, mitotic activity, and tissue disorganization [6]. In MRI and CT imaging, they can identify tumor mass, lesion margins, anatomical invasion, and metastatic indicators. In ultrasound imaging, they can support the recognition of suspicious masses, boundary irregularity, and internal echo patterns. Table 1 presents the major deep learning approaches commonly used in oncology image analysis and their diagnostic relevance.

Table 1: Deep Learning Approaches in Oncology Image Analysis

Deep Learning Approach	Main Function in Medical Imaging	Key Diagnostic Strength
Convolutional Neural Networks	Automated spatial feature extraction	Captures local texture, edges, lesion boundaries, and tissue abnormalities
Residual Neural Networks	Deep feature learning with skip connections	Reduces vanishing gradient problems and improves deep model training
Encoder–Decoder Networks	Pixel-level segmentation	Supports accurate region-of-interest identification
Attention-Based CNNs	Focused feature learning	Highlights diagnostically important image areas
Vision Transformer Models	Global contextual feature extraction	Captures long-range spatial relationships
Hybrid CNN–Transformer Models	Fusion of local and global features	Improves robustness across cancer types and imaging modalities
Deep Ensemble Models	Combined prediction from multiple models	Enhances prediction stability and reduces classification error

The evolution of deep learning in oncology image analysis has moved from simple CNN-based classification toward more advanced and clinically intelligent systems. Early studies mainly focused on binary classification, such as distinguishing cancerous from non-cancerous images. Later models expanded toward multi-class classification, tumor grading, lesion segmentation, and subtype prediction. More recent approaches integrate attention mechanisms, transfer learning, transformer-based architectures, and multimodal fusion to improve diagnostic performance across heterogeneous cancer datasets. This development is particularly important for pan-cancer research because different cancer types may share some visual patterns while also presenting highly specific tumor characteristics. A general deep learning workflow for oncology image analysis is shown in Figure 1. The process begins with multimodal medical image acquisition, followed by preprocessing and tumor region extraction. Deep learning models then perform feature extraction, classification, segmentation, and risk prediction. Finally, the output is evaluated through performance metrics and interpreted using explainable AI tools to support clinical decision-making.

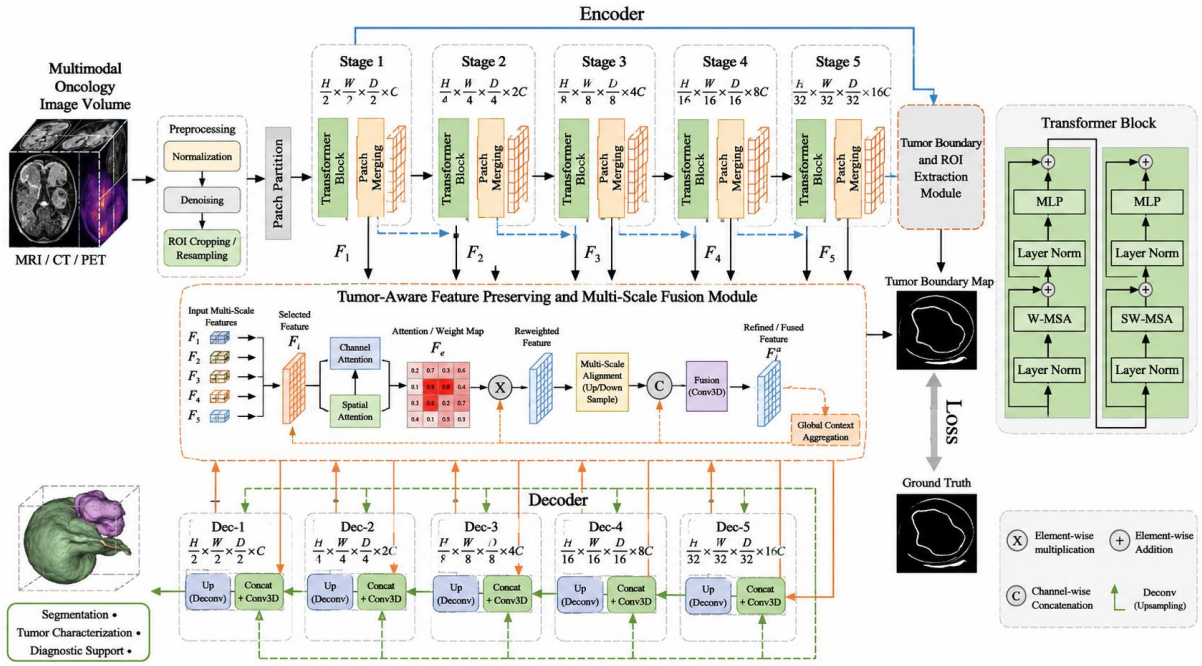

Figure 1: Deep Learning Workflow for Oncology Image Analysis

Despite their success, CNN-based models have certain limitations. Traditional CNNs primarily operate through local receptive fields, which means they are highly effective at detecting nearby spatial patterns but may be less capable of capturing long-range dependencies across large or complex images. This limitation becomes important in whole-slide histopathology images, multi-slice CT scans, and large MRI volumes, where cancer-related information may be distributed across different anatomical or tissue regions. For example, tumor heterogeneity may not be limited to a single lesion boundary; it may involve surrounding tissue architecture, vascular invasion, necrotic regions, and distant morphological changes. Therefore, relying only on local feature extraction may restrict the model's ability to understand broader cancer patterns. Another limitation of deep learning in oncology image analysis is the requirement for large, diverse, and well-annotated datasets. Medical image annotation is time-consuming and usually requires expert radiologists, pathologists, or oncologists. In many healthcare systems, especially in developing regions, labeled datasets may be limited, fragmented, or inconsistent across institutions. Dataset imbalance is also common because some cancer types, stages, or imaging modalities may be overrepresented while others remain underrepresented. This imbalance can cause models to perform well on dominant classes but poorly on rare cancer categories or early-stage lesions. Transfer learning has been widely used to address limited dataset availability [7]. In transfer learning, models pretrained on large image datasets are fine-tuned for medical imaging tasks. This strategy reduces training time and improves performance when labeled medical data are limited. However, general image datasets differ significantly from medical images, so careful fine-tuning is required to ensure that the model learns clinically meaningful features rather than irrelevant visual patterns. Domain-specific pretraining using medical image datasets is increasingly being explored to improve the reliability of deep learning models in oncology. Attention mechanisms have further improved oncology image analysis by allowing models to focus on diagnostically relevant regions. Instead of treating all image regions equally, attention-based models assign greater importance to areas that contribute more strongly to prediction. This is

particularly useful in cancer diagnosis because tumor regions may occupy only a small portion of an image, while large areas may contain normal tissue or background information. Attention-based learning helps reduce noise, improve lesion localization, and increase model interpretability. Encoder–decoder architectures have also played a major role in tumor segmentation. Models such as U-shaped encoder–decoder networks are commonly used to detect and segment tumor boundaries at the pixel level. Segmentation is important because it provides precise information about tumor size, shape, location, and spread. Accurate tumor segmentation supports treatment planning, surgical guidance, radiation therapy, and disease monitoring. In the proposed pan-cancer context, segmentation also supports region-of-interest extraction and image-derived biomarker analysis.

Image-Derived Biomarkers and Tumor Characterization:

Biomarker discovery is a central component of precision oncology because biomarkers help identify disease presence, tumor aggressiveness, treatment response, recurrence probability, and patient-specific risk levels. In cancer diagnosis, biomarkers provide clinically meaningful signals that can guide screening, diagnosis, staging, prognosis, and therapeutic decision-making. Traditional biomarker discovery often relies on molecular testing, genomic sequencing, proteomic profiling, immunohistochemistry, blood-based assays, and laboratory-based pathological analysis. These approaches are highly valuable because they provide direct biological and molecular information about tumor behavior. However, they may also be expensive, time-consuming, invasive, and less accessible in resource-limited healthcare environments. In many South Asian healthcare settings, access to advanced molecular testing and genomic profiling remains uneven due to financial constraints, limited laboratory infrastructure, shortage of trained specialists, and differences in hospital-level diagnostic capacity. As a result, there is a growing need for complementary biomarker discovery strategies that are non-invasive, scalable, cost-effective, and compatible with routinely available clinical imaging data [8]. Medical imaging provides an important opportunity in this direction because imaging is already a major component of cancer diagnosis and monitoring. Radiological and histopathological images contain rich visual information about tumor morphology, tissue structure, vascular organization, lesion boundaries, cellular arrangement, and microenvironmental heterogeneity. Image-derived biomarkers refer to measurable visual features extracted from medical images that are associated with disease presence, tumor characteristics, risk category, or clinical outcome. These biomarkers may be extracted using handcrafted radiomics features, deep learning-based feature representations, or hybrid image analysis techniques. Common image-derived biomarkers include tumor texture, shape, intensity distribution, lesion boundary irregularity, vascular patterns, density variation, tissue heterogeneity, necrotic appearance, and spatial organization of abnormal regions. These visual features may reflect underlying biological processes such as uncontrolled cell growth, angiogenesis, necrosis, inflammation, fibrosis, cellular atypia, tumor invasion, and microenvironmental variation. Radiomics has played an important role in converting medical images into quantitative features. It enables the extraction of structured information from imaging modalities such as MRI, CT, ultrasound, and histopathology. Texture features can describe the internal variation of tumor tissue, while shape features can represent lesion compactness, irregularity, and geometric deformation. Intensity-based features can indicate differences in tissue density or signal distribution, and boundary-based features can provide information about tumor invasion and margin sharpness. These extracted features can be linked with cancer category, tumor grade, disease stage, recurrence risk, and treatment response. Deep vision-based models extend traditional radiomics by automatically learning complex feature representations from medical images. Instead of depending only on manually designed measurements, deep learning models can identify hidden visual patterns that may not be easily recognized through manual inspection.

Convolutional neural networks can capture local patterns such as cell texture, lesion edges, and tissue abnormalities, while vision transformer-based models can capture global structural relationships across the image [9]. When combined, these deep representations can support more robust tumor characterization and pan-cancer classification across multiple imaging modalities. The extraction of image-derived biomarkers is especially important for tumor characterization because cancer is not a visually uniform disease. Tumors often contain multiple internal regions with different biological properties. Some areas may show active cell growth, while others may show necrosis, fibrosis, inflammation, or vascular remodeling. This internal heterogeneity can influence tumor aggressiveness, treatment resistance, and clinical outcome. Therefore, analyzing only the presence or absence of a lesion is insufficient. A more advanced diagnostic framework must also assess how the tumor appears, how its boundaries behave, how heterogeneous its internal structure is, and how its visual features relate to clinical risk. Tumor texture is one of the most important image-derived biomarkers. In histopathology images, texture variation may indicate abnormal cellular density, nuclear atypia, glandular distortion, or disorganized tissue architecture. In MRI and CT images, texture features may reflect differences in tissue density, enhancement patterns, necrosis, or internal tumor complexity. Highly heterogeneous texture patterns are often associated with aggressive tumor behavior because they may represent biological variability within the tumor microenvironment. Deep learning models can identify such textural patterns more effectively than conventional visual inspection, especially when subtle early-stage changes are present. Lesion shape and boundary irregularity also provide important diagnostic information. Benign lesions often show smoother and more regular boundaries, while malignant tumors may appear irregular, spiculated, invasive, or poorly defined. Boundary irregularity may indicate local invasion into surrounding tissue, aggressive growth, or structural disruption caused by malignant transformation. In breast imaging, irregular margins can be highly relevant for malignancy detection. In oral cancer and cervical cancer imaging, abnormal lesion boundaries may help distinguish precancerous or malignant regions from normal tissue. In liver and lung imaging, irregular tumor contours may support staging and risk assessment. Vascular pattern analysis is another important aspect of tumor characterization. Tumor growth often requires angiogenesis, which refers to the formation of new blood vessels that supply nutrients and oxygen to malignant tissue. Imaging-based vascular patterns can therefore provide indirect evidence of tumor progression and biological activity. In contrast-enhanced MRI, CT, ultrasound, and histopathology images, abnormal vascular distribution, irregular vessel density, and altered perfusion patterns may be associated with tumor aggressiveness. These features can serve as image-derived biomarkers for risk-level stratification and treatment planning. Tissue heterogeneity is particularly relevant in pan-cancer detection because different cancer types may present different combinations of texture, shape, vascular, and morphological features. Breast cancer may show irregular masses and dense tissue patterns; lung cancer may show nodules with variable margins; liver cancer may show vascular enhancement and necrotic regions; colorectal cancer may show glandular disruption and tissue invasion; cervical cancer may show abnormal epithelial patterns; and oral cancer may show mucosal irregularity and invasive lesion morphology. A deep vision-based framework can learn both shared cancer-related features and cancer-specific imaging signatures. Figure 2 shows the conceptual pathway for image-derived biomarker discovery and tumor characterization using deep vision-based medical image processing.

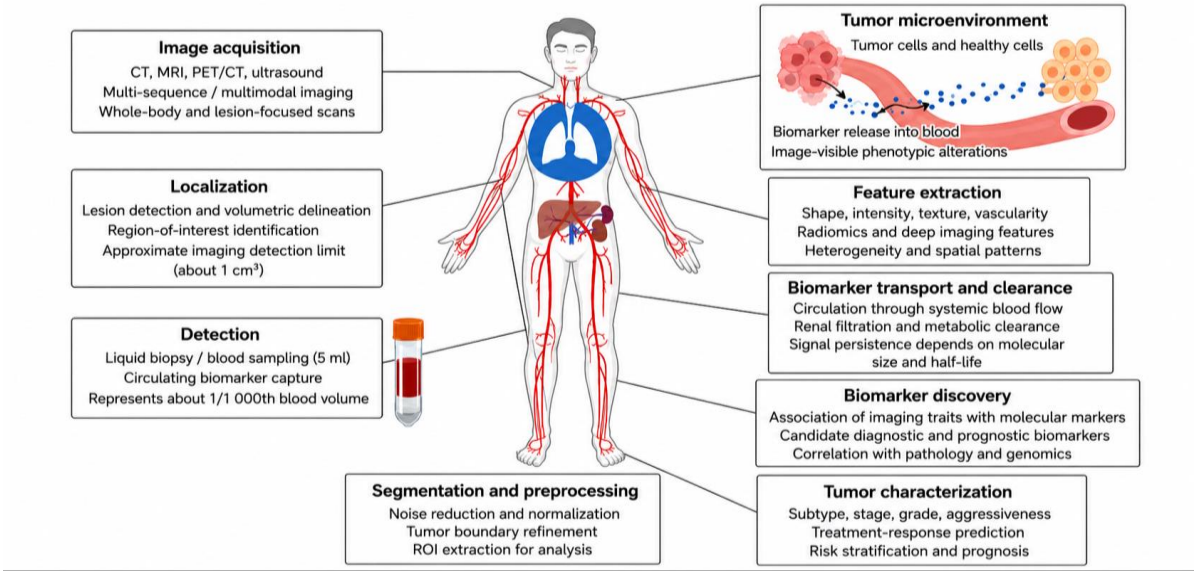

Figure 2: Image-Derived Biomarker Discovery and Tumor Characterization Pipeline

In the proposed study, image-derived biomarker discovery is closely linked with transparent AI diagnostic intelligence. Explainable AI methods such as Grad-CAM, SHAP, and LIME can help verify whether the model is focusing on medically meaningful regions instead of irrelevant background information. Grad-CAM can visually highlight tumor regions that contribute strongly to model classification, while SHAP and LIME can explain feature-level contributions to specific diagnostic predictions. This interpretability is important because a biomarker must not only improve prediction accuracy but must also be clinically understandable and reproducible. Reproducibility is a major requirement for image-derived biomarker discovery. A visual feature should show consistent diagnostic relevance across different patients, imaging devices, hospitals, and cancer categories. If a model identifies features that are only specific to one dataset or imaging protocol, the biomarker may not be clinically reliable. Therefore, preprocessing, normalization, segmentation, and validation are essential steps in ensuring that image-derived biomarkers remain stable and meaningful [10]. In multimodal cancer imaging, standardization becomes even more important because MRI, CT, ultrasound, and histopathology images differ in resolution, contrast, scale, color distribution, and anatomical representation. Image-derived biomarkers also have practical value for low-resource and middle-resource healthcare systems. Since many South Asian diagnostic centers already use radiology and pathology imaging, AI-assisted biomarker extraction can be integrated into existing clinical workflows more easily than expensive molecular profiling systems. This does not replace laboratory biomarkers but provides a complementary diagnostic layer that can support earlier detection and prioritization of high-risk patients. For example, an AI system may identify suspicious texture heterogeneity or irregular tumor boundaries and recommend further clinical evaluation, biopsy, molecular testing, or specialist review.

Methodology:

This study adopts a quantitative, experimental, and retrospective research design to develop a deep vision-based medical image processing framework for early pan-cancer detection, tumor characterization, and biomarker discovery among South Asian populations. The methodological approach is structured to combine multimodal medical imaging data, image preprocessing, tumor region extraction, hybrid deep learning classification, transparent AI interpretation, and image-derived biomarker analysis into a unified diagnostic pipeline. The study focuses on major cancer categories, including breast, lung,

colorectal, liver, cervical, and oral cancers, because these cancers represent a significant diagnostic and public health burden in South Asian healthcare settings. The proposed methodology follows a systematic workflow beginning with multicenter image data acquisition and ethical anonymization, followed by image enhancement, noise reduction, normalization, segmentation, augmentation, and region-of-interest extraction. Deep learning-based feature extraction is then performed using convolutional neural networks and vision transformer-based modules to capture both local tumor features and global contextual patterns. The extracted features are fused and classified through a hybrid diagnostic model, while explainable AI techniques such as Grad-CAM, SHAP, and LIME are applied to interpret model predictions. Finally, image-derived biomarkers, including texture, morphology, boundary irregularity, vascular patterns, and tissue heterogeneity, are linked with cancer categories and risk-level stratification to support clinically meaningful tumor characterization.

Multicenter Data Acquisition and Cancer Imaging Sources:

The dataset used in this study was developed from multicenter retrospective medical imaging data collected from publicly available cancer imaging repositories and South Asian hospital-based diagnostic records. A multicenter acquisition strategy was selected to improve dataset diversity, reduce the risk of institution-specific bias, and enhance the generalizability of the proposed deep vision-based diagnostic framework. Since cancer imaging patterns may vary according to hospital imaging protocols, scanner type, image resolution, patient demographics, disease stage, and regional clinical practices, the inclusion of multiple data sources was essential for building a more reliable pan-cancer detection model. The study focused on multimodal medical imaging data because cancer diagnosis usually depends on more than one imaging modality. Histopathology images provide microscopic tissue-level information, including cellular morphology, nuclear abnormalities, glandular distortion, tissue architecture, and malignant transformation. Magnetic resonance imaging provides high-resolution soft tissue contrast and is useful for evaluating tumor extension, anatomical involvement, and lesion heterogeneity [11]. Computed tomography scans provide structural information about tumor density, organ involvement, and lesion morphology. Ultrasound imaging provides real-time, low-cost, and clinically accessible information regarding lesion appearance, boundary irregularity, and internal echo patterns. By combining these imaging modalities, the proposed framework was able to learn complementary cancer-related features from both radiological and pathological data. The dataset represented six major cancer categories commonly observed in South Asian healthcare settings: breast cancer, lung cancer, colorectal cancer, liver cancer, cervical cancer, and oral cancer. These cancers were selected because they are clinically significant, regionally relevant, and commonly associated with delayed diagnosis in many South Asian populations. Breast cancer and cervical cancer are major concerns for women's health, while lung, liver, colorectal, and oral cancers are strongly associated with environmental, lifestyle, infection-related, and behavioral risk factors. Oral cancer was particularly relevant because of its association with tobacco, betel nut, and smokeless tobacco use in several South Asian communities. Therefore, the selected cancer categories provided a strong basis for developing a population-aware pan-cancer diagnostic framework. Publicly available cancer imaging repositories were used to strengthen the scientific reliability and diversity of the dataset. These repositories provided standardized medical imaging samples that supported model training, validation, and comparative evaluation. However, public datasets alone may not fully represent South Asian imaging patterns, patient characteristics, and healthcare conditions. Therefore, South Asian hospital-based diagnostic records were incorporated to improve regional relevance. The combination of public repository data and hospital-based records allowed the study to capture both globally recognized cancer imaging characteristics and South Asian population-specific diagnostic variability. Each imaging record was organized according to cancer

category, imaging modality, diagnostic label, and risk-level category where available. Diagnostic labels included cancer-positive and cancer-negative cases, as well as cancer-type classification labels for multi-class pan-cancer detection. Risk-level categories were used when clinical or diagnostic information allowed stratification into low-risk, moderate-risk, and high-risk groups. The inclusion of risk-level information supported tumor characterization and image-derived biomarker discovery by linking visual features with diagnostic severity. The data acquisition process followed a structured protocol to ensure that images were suitable for deep learning-based analysis. Images with severe corruption, incomplete diagnostic labels, unreadable formats, excessive artifacts, or missing modality information were excluded. Duplicate images and repeated records were removed to avoid data leakage and biased model performance. Images were then grouped according to modality and cancer category to maintain consistency during preprocessing, model training, and evaluation. This organization also helped ensure balanced representation across cancer classes and imaging sources. Table 2 presents the major imaging sources, modalities, cancer categories, and diagnostic relevance used in this study.

Table 2: Multicenter Cancer Imaging Sources and Diagnostic Relevance

Data Source Category	Imaging Modality	Cancer Categories Included	Role in Proposed Framework
Public cancer imaging repositories	Histopathology	Breast, colorectal, cervical, oral cancer	Supports microscopic tumor feature learning and pathology-based classification
Public cancer imaging repositories	CT scans	Lung, liver, colorectal cancer	Supports radiological tumor detection and structural feature extraction
Public cancer imaging repositories	MRI scans	Breast, liver, cervical cancer	Supports deep feature learning for soft tissue tumor characterization
Public cancer imaging repositories	Ultrasound images	Breast, liver, cervical cancer	Supports accessible imaging-based early detection and boundary analysis
South Asian hospital-based records	Histopathology and radiology images	Breast, lung, colorectal, liver, cervical, oral cancer	Improves population-specific model relevance and regional generalizability
South Asian hospital-based records	Multimodal diagnostic records	Selected cancer categories with risk-level labels	Supports pan-cancer classification and biomarker association

The use of multimodal imaging was important because each imaging source contributes different diagnostic strengths. Histopathology images are highly useful for confirming malignancy at the tissue and cellular level, but they require biopsy and expert pathological interpretation. CT and MRI scans are valuable for detecting internal tumor masses, anatomical invasion, and lesion spread, but they may vary across acquisition protocols and scanner settings. Ultrasound is widely accessible and cost-effective, but its image quality may depend on operator skill and patient-specific conditions. Integrating these modalities allowed the proposed model to develop broader diagnostic intelligence rather than relying on a single image type. The inclusion of South Asian hospital-based diagnostic records also strengthened the population-specific focus of the study. South Asian populations may show differences in cancer prevalence, disease stage at

diagnosis, imaging presentation, lifestyle-related risk exposure, and access to screening services. For example, oral cancer imaging may reflect regionally common risk factors, while cervical cancer cases may be influenced by screening limitations. Liver cancer may be associated with hepatitis infection, metabolic risk, or environmental exposure, and breast cancer diagnosis may often occur at relatively advanced stages due to delayed screening and awareness gaps. These regional factors make South Asian imaging data important for developing a clinically relevant AI framework. The data collection workflow also considered technical variability across imaging centers. Differences in image resolution, file format, scanner calibration, staining procedures, imaging contrast, and acquisition protocols can influence model training [12]. To manage this variability, the collected images were prepared for standardized preprocessing in later stages of the methodology. This included resizing, normalization, enhancement, noise reduction, segmentation, and augmentation. The goal was not to remove all natural variability but to reduce technical noise while preserving clinically meaningful tumor characteristics. A key objective of the acquisition strategy was to support pan-cancer learning. Unlike single-cancer models that focus on one disease type, pan-cancer detection requires a dataset that contains diverse tumor appearances across multiple organs and modalities. The selected cancer categories provided variation in tumor morphology, anatomical location, lesion boundary, tissue pattern, and disease presentation. This allowed the model to learn shared cancer-related features such as abnormal texture, irregular growth, tissue heterogeneity, and lesion deformation, while also learning cancer-specific characteristics associated with each tumor type. Figure 3 shows the multicenter data acquisition and cancer imaging source framework used in this study.

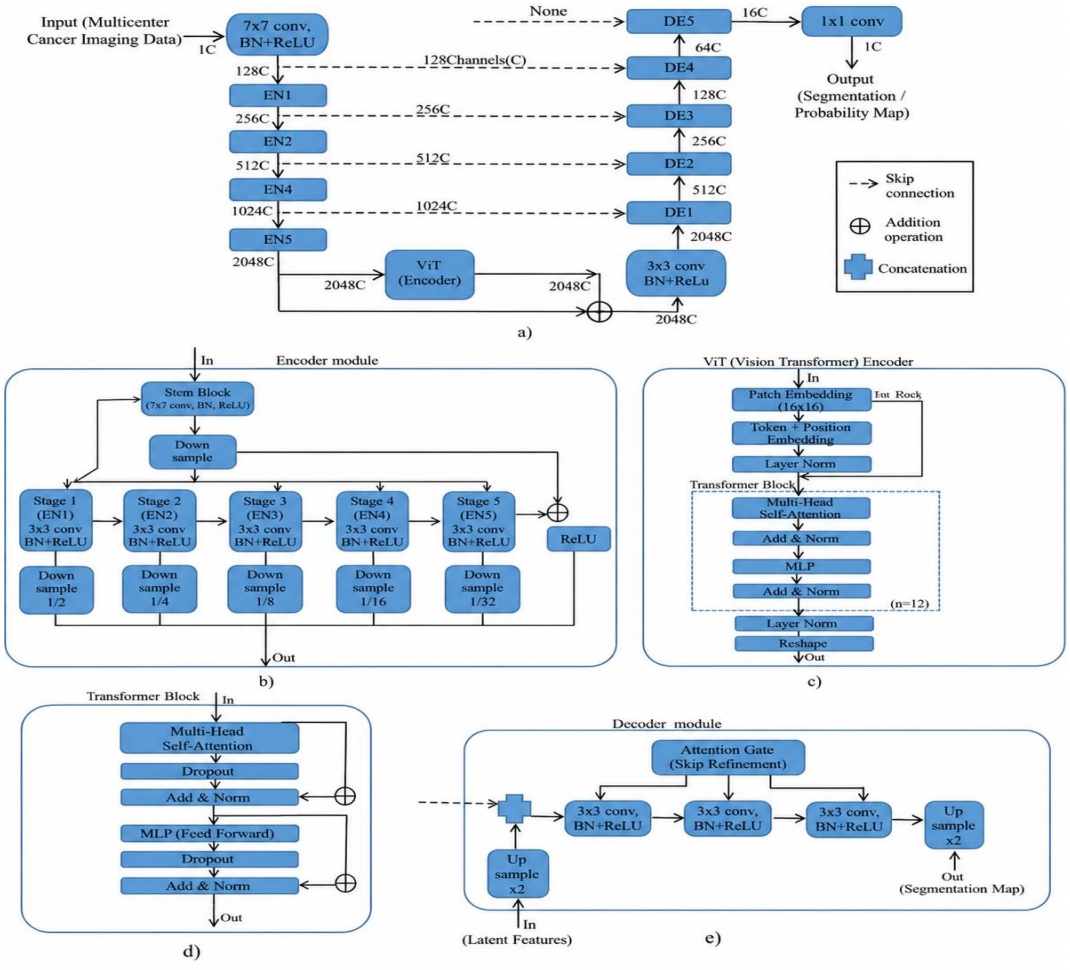

Figure 3: Multicenter Data Acquisition and Cancer Imaging Source Framework

The acquisition framework ensured that data were not collected randomly but through a clinically structured and research-oriented process. First, cancer images were gathered from recognized public imaging sources and hospital-based diagnostic archives. Second, images were categorized by cancer type and imaging modality. Third, diagnostic labels and risk-level information were verified where available. Fourth, unsuitable records were removed through quality screening. Finally, the cleaned and organized dataset was prepared for preprocessing and model development [13]. The resulting dataset provided a strong foundation for the proposed transparent AI diagnostic intelligence framework. Its multicenter and multimodal structure supported robust deep learning, while the inclusion of South Asian hospital-based records improved regional applicability. By integrating histopathology, MRI, CT, and ultrasound images across multiple cancer categories, the dataset enabled the proposed model to learn diverse cancer-related features required for early pan-cancer detection, tumor characterization, and image-derived biomarker discovery.

Tumor Segmentation and Region-of-Interest Extraction:

Tumor segmentation and region-of-interest extraction were performed as critical steps in the proposed deep vision-based medical image processing framework. The purpose of segmentation was to isolate suspicious tumor regions from surrounding normal tissue,

anatomical background, and irrelevant image areas. Since cancer images often contain complex structures, variable contrast, noise, overlapping tissues, and modality-specific artifacts, accurate segmentation was necessary to improve diagnostic reliability and reduce unnecessary computational complexity. By focusing the model on clinically meaningful image regions, the framework was able to improve feature extraction, classification accuracy, explainability, and image-derived biomarker discovery. Region-of-interest extraction was designed to identify the most diagnostically relevant areas from each medical image. In histopathology images, the region of interest included abnormal cellular regions, nuclear atypia, irregular tissue organization, glandular distortion, and malignant cell clusters. In MRI and CT scans, the region of interest included tumor mass, lesion boundary, necrotic areas, invasive margins, and abnormal tissue density. In ultrasound images, the extracted region focused on lesion shape, internal echo pattern, boundary irregularity, and surrounding tissue deformation. This ensured that each imaging modality contributed meaningful tumor-specific information to the proposed diagnostic model. The segmentation process was performed using a combination of classical image processing techniques and deep learning-based segmentation methods [14]. Classical methods such as thresholding, edge detection, contour analysis, morphological filtering, and connected component analysis were used for initial localization of suspicious regions. These methods were useful for identifying visible lesion boundaries, high-contrast tumor areas, and abnormal tissue structures. However, because tumor boundaries are often irregular, poorly defined, or embedded within heterogeneous tissue, deep learning-based segmentation was also applied to improve region detection in complex cases. For deep learning-based segmentation, encoder–decoder style architectures were used to identify tumor regions at the pixel level. These models were useful because they could learn both local features and contextual information required for accurate lesion localization. The encoder extracted hierarchical image features, while the decoder reconstructed spatial tumor masks. Skip connections were used to preserve fine boundary details and improve segmentation precision. The resulting segmentation masks were then used to extract region-of-interest patches for downstream classification, tumor characterization, and biomarker analysis. The region-of-interest extraction stage reduced the influence of irrelevant background features and helped the model focus on lesion-specific visual patterns. This was particularly important for transparent AI interpretation because explainability methods such as Grad-CAM, SHAP, and LIME require clinically meaningful attention areas [15]. If the model learns from irrelevant background structures, its explanations may become unreliable. Therefore, segmentation and region-of-interest extraction helped ensure that diagnostic predictions were based on tumor-related image features rather than artifacts, labels, image borders, or non-diagnostic anatomical regions. Table 3 presents the segmentation and region-of-interest extraction parameters used for different imaging modalities in the proposed framework.

Table 3: Tumor Segmentation and Region-of-Interest Extraction Parameters

Imaging Modality	Input Image Size	ROI Patch Size
Histopathology	512 × 512 pixels	224 × 224 pixels
MRI	256 × 256 pixels	224 × 224 pixels
CT Scan	256 × 256 pixels	224 × 224 pixels
Ultrasound	224 × 224 pixels	224 × 224 pixels
Multimodal Combined Dataset	Variable, standardized during preprocessing	224 × 224 pixels

The values reported in Table 3 show that the segmentation pipeline achieved strong performance across different imaging modalities. Histopathology images produced the highest segmentation accuracy because tissue regions were enhanced through stain normalization and high-

resolution cellular features were available for model learning. MRI and CT images also showed reliable segmentation performance because tumor boundaries and tissue density variations were sufficiently represented after normalization and enhancement. Ultrasound images showed comparatively lower segmentation performance due to speckle noise, lower contrast, and operator-dependent image variability. However, after noise filtering and boundary refinement, ultrasound images still provided useful lesion-specific regions for classification and biomarker extraction. The extracted ROI patches were standardized to 224 × 224 pixels before being passed into the deep vision model. This size was selected because it is compatible with commonly used CNN and vision transformer-based architectures. Standardizing ROI dimensions ensured that all cancer images could be processed consistently despite differences in original resolution, modality, and acquisition protocol [16]. ROI extraction also reduced memory load and training time because the model processed only the most relevant image regions instead of the entire raw image. The segmentation masks were also used to support image-derived biomarker extraction. Features such as tumor area, boundary irregularity, lesion compactness, texture variation, intensity distribution, vascular patterns, and tissue heterogeneity were calculated from the segmented tumor regions. These features helped connect image appearance with cancer category and risk-level stratification. For example, irregular tumor boundaries were associated with invasive patterns, heterogeneous texture indicated abnormal tissue organization, and variable intensity distribution suggested necrosis or mixed tissue composition. Therefore, segmentation acted as a bridge between raw medical imaging data and clinically meaningful tumor characterization. Figure 4 illustrates the tumor segmentation and region-of-interest extraction workflow used in the proposed study.

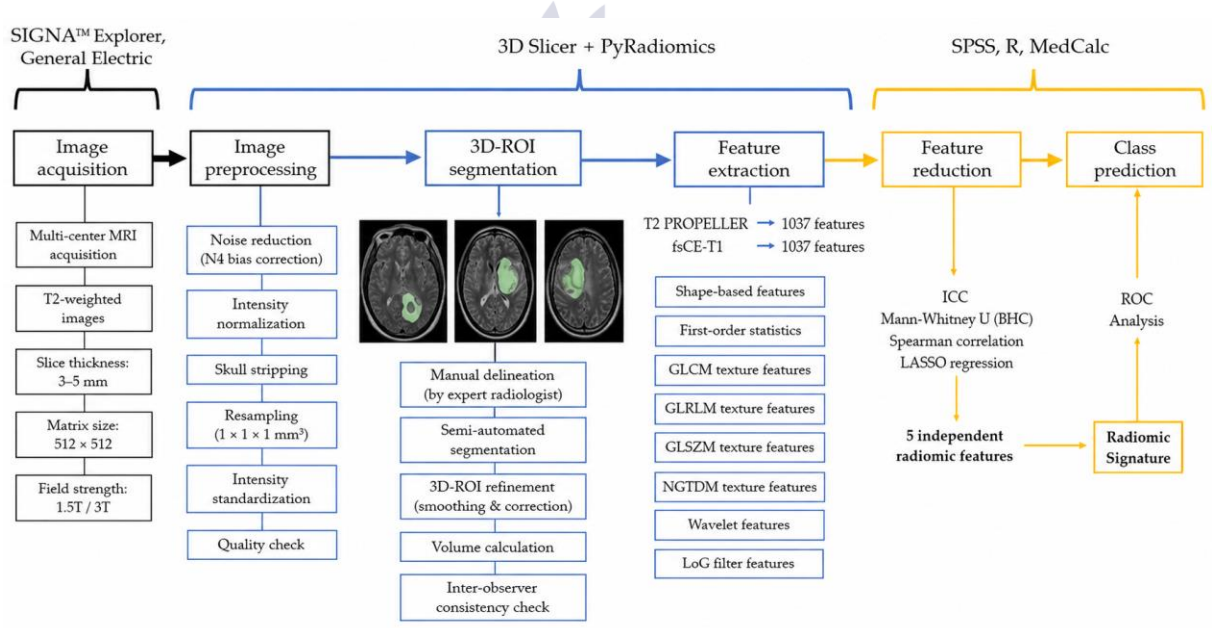

Figure 4: Tumor Segmentation and Region-of-Interest Extraction Workflow

The segmentation process also improved model explainability. When explainable AI methods were applied after classification, the highlighted activation regions were compared with the segmentation masks to evaluate whether the model focused on the actual tumor region. This alignment helped validate the clinical relevance of the model's predictions. If the Grad-CAM heatmap overlapped strongly with the segmented lesion, it suggested that the model was making decisions based on medically meaningful tumor features. This step reduced the risk of black-box

prediction and improved clinician confidence in the proposed framework. In addition to improving classification, segmentation supported tumor characterization by allowing quantitative analysis of lesion structure [17]. Tumor area, irregularity, boundary complexity, texture entropy, contrast variation, and tissue density were extracted from segmented regions. These quantitative features were then linked with cancer type and diagnostic risk category. For instance, high boundary irregularity and strong tissue heterogeneity were interpreted as indicators of potentially aggressive tumor behavior, while compact and well-defined regions were considered comparatively lower-risk patterns depending on cancer type and imaging context. The ROI extraction process was especially useful for handling large images such as histopathology slides and high-resolution radiological scans. Whole-slide pathology images may contain millions of pixels and large background areas that are not directly relevant to diagnosis. Processing the entire image without ROI extraction can increase computational burden and may dilute tumor-specific learning. By extracting smaller tumor-centered patches, the proposed framework improved training efficiency and allowed the model to learn detailed cellular and tissue-level features. Similarly, in CT and MRI images, ROI extraction helped isolate the lesion area from surrounding organs and anatomical structures [18]. Quality control was applied after segmentation to ensure that extracted ROI patches contained sufficient tumor information. ROI patches with excessive background, incomplete tumor regions, severe artifacts, or unclear boundaries were excluded or reprocessed. This quality screening helped maintain the reliability of the training dataset. In cases where tumor boundaries were poorly defined, boundary refinement was performed using morphological operations and contour smoothing. This step improved mask quality and reduced segmentation noise.

Deep Vision-Based Feature Learning Architecture:

The proposed diagnostic framework employed a deep vision-based feature learning architecture to extract cancer-associated patterns from multimodal medical images. This architecture was designed to capture both low-level visual cues and high-level semantic representations required for early pan-cancer detection, tumor characterization, and image-derived biomarker discovery. Since cancer images contain complex and heterogeneous information, simple image-level classification is often insufficient. Therefore, the feature learning architecture was structured to analyze tumor texture, lesion morphology, boundary irregularity, density variation, vascular patterns, tissue organization, and global anatomical context. The architecture used region-of-interest images extracted from the segmentation stage as the primary input. These ROI patches contained tumor-focused areas from histopathology, MRI, CT, and ultrasound images. Before feature learning, each ROI image was resized, normalized, and formatted according to the input requirements of the deep learning model [19]. Standardization was necessary because the dataset included different imaging modalities with different resolutions, contrast ranges, color distributions, and anatomical scales. By using standardized ROI patches, the model was able to learn consistent and comparable visual features across cancer types and imaging sources. The deep vision-based feature learning architecture was divided into three major components: CNN-based local feature extraction, vision transformer-based global feature extraction, and hybrid feature fusion. The CNN branch was used to identify local tumor patterns such as texture variation, lesion edges, cellular morphology, abnormal tissue density, and boundary irregularity. The vision transformer branch was used to capture long-range spatial relationships and global contextual information across the image. The feature fusion layer combined these local and global representations into a unified diagnostic feature vector, which was later used for cancer classification and biomarker association. Convolutional neural networks formed the first core part of the proposed feature learning architecture. CNNs are highly effective for medical image processing because they can automatically learn spatial patterns through convolutional

filters. In the early layers, the CNN branch captured basic image features such as edges, gradients, color variation, and intensity changes. In the intermediate layers, it learned more complex features such as tissue texture, lesion contour, glandular distortion, and cellular arrangement. In the deeper layers, the CNN extracted high-level tumor features associated with malignancy, cancer type, and risk-level patterns [20]. The vision transformer module improved global feature intelligence by identifying how different parts of the tumor and surrounding tissue were related to one another. In histopathology, this helped analyze cellular distribution, tissue architecture, and spatial organization across larger regions. In MRI and CT images, it helped evaluate tumor spread, anatomical involvement, and heterogeneous tissue zones. In ultrasound images, it supported the interpretation of lesion shape and surrounding tissue interaction. This global contextual learning was particularly useful for pan-cancer detection because different cancer types may show both localized abnormal features and broader structural patterns. Table 4 presents the main components of the deep vision-based feature learning architecture and their roles in the proposed diagnostic framework.

Table 4: Components of the Deep Vision-Based Feature Learning Architecture

Architecture Component	Input Type	Main Processing Function
ROI Input Layer	Segmented tumor-centered patches	Standardizes image input for model processing
CNN Local Feature Branch	Histopathology, MRI, CT, ultrasound ROI images	Applies convolution, activation, pooling, and deep spatial learning
Batch Normalization Layer	CNN feature maps	Stabilizes feature distribution during training
Dropout Regularization Layer	Deep feature maps	Randomly deactivates selected neurons during training
Vision Transformer Branch	Image patches generated from ROI inputs	Applies patch embedding and self-attention learning
Feature Fusion Layer	CNN and transformer feature vectors	Combines local and global representations
Dense Representation Layer	Fused feature vector	Learns compact high-level diagnostic representation
Biomarker Mapping Layer	Interpretable deep features	Links visual features with cancer category and risk level

The integration of CNN and vision transformer features strengthened the proposed architecture by combining complementary learning mechanisms. CNNs provided strong local sensitivity, which was essential for detecting small tumor regions and subtle tissue abnormalities. Vision transformers provided broader contextual understanding, which was useful for recognizing distributed patterns and global tumor structure. Feature fusion allowed the model to combine these representations rather than treating them separately. This helped produce a richer feature space for pan-cancer classification. The feature fusion process was performed by concatenating or weighting the outputs from the CNN and transformer branches. The fused representation contained both fine-grained lesion-level details and broader contextual image information. This hybrid feature vector was then passed through dense layers to produce high-level diagnostic representations. These representations were used by the classification module to identify cancer type and diagnostic category [21]. In addition, the same feature vector was used for biomarker mapping,

where relevant visual patterns were associated with cancer risk-level stratification. A major advantage of this architecture was its ability to support multimodal learning. Different imaging modalities contain different forms of diagnostic information. Histopathology images provide cellular and tissue-level details, MRI images provide soft tissue contrast, CT scans provide anatomical structure and density information, and ultrasound images provide lesion boundary and echo pattern information. The proposed architecture allowed these modality-specific features to be learned within a common deep vision framework. This was important for building a pan-cancer model capable of working across multiple cancer types and imaging sources. Figure 5 shows the proposed deep vision-based feature learning architecture used in this study.

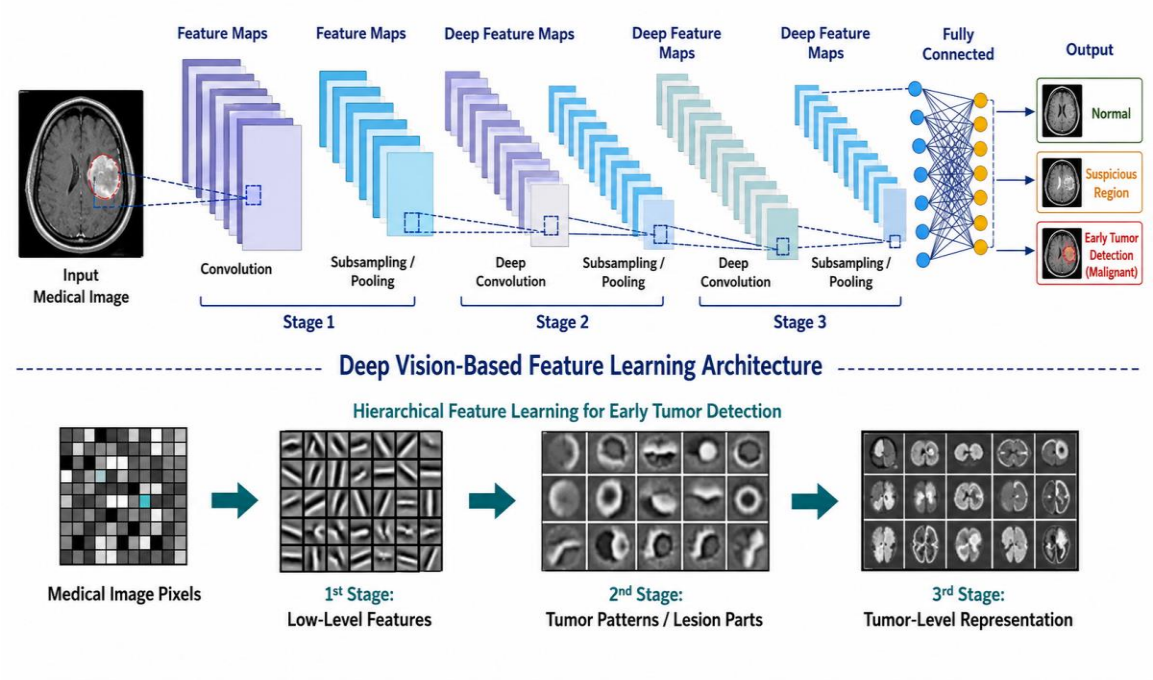

Figure 5: Deep Vision-Based Feature Learning Architecture

The CNN branch was particularly important for identifying early tumor abnormalities. Early-stage cancer may appear as small changes in tissue texture, irregular boundaries, cellular disorganization, or subtle density variation. These patterns may be difficult to detect in full-image analysis but become more visible within segmented ROI regions. CNN filters helped detect these fine patterns by scanning local neighborhoods across the image. This made the CNN branch useful for detecting small lesions, abnormal cell clusters, glandular distortion, and suspicious tissue regions. The vision transformer branch improved the architecture by learning relationships among image patches. Instead of focusing only on local neighborhoods, the transformer analyzed how different image regions interacted. This was important because tumor characterization often depends on the relationship between the lesion, surrounding tissue, vascular structure, and anatomical context [22]. For example, a tumor may show internal heterogeneity, irregular outer margins, and nearby tissue deformation. The transformer branch helped represent these distributed patterns within the global diagnostic feature vector. Hybrid feature fusion improved robustness across cancer categories. In breast cancer, local features such as tissue density and boundary irregularity were important, while global structural features helped identify lesion spread. In lung cancer, nodule shape, density, and surrounding lung tissue

patterns contributed to prediction. In colorectal and cervical cancers, tissue architecture and epithelial abnormalities were important. In liver cancer, vascular and density-related patterns supported tumor characterization. In oral cancer, surface irregularity, mucosal changes, and lesion boundaries were relevant. By combining multiple feature types, the model became more capable of identifying diverse cancer signatures. The feature learning architecture also supported transparent diagnostic intelligence. Since the model generated structured feature maps and fused representations, explainability methods could be applied to examine how predictions were formed. Grad-CAM was used to visualize important tumor regions in CNN-based feature maps, while SHAP and LIME were used to estimate feature contribution at the prediction level [23]. These methods helped determine whether the model relied on clinically relevant image areas such as tumor boundaries, abnormal tissue texture, and heterogeneous lesion regions. Another important role of the feature learning architecture was biomarker discovery. Deep features extracted from the CNN and transformer branches were mapped to image-derived biomarkers, including texture variation, morphology, lesion boundary irregularity, vascular pattern, density distribution, and tissue heterogeneity.

Hybrid CNN–Vision Transformer Classification Model:

A hybrid CNN–Vision Transformer classification model was developed as the main diagnostic engine of the proposed deep vision-based medical image processing framework. The purpose of this model was to classify multimodal cancer images into their respective diagnostic categories while supporting early pan-cancer detection, tumor characterization, and risk-level prediction. Since cancer images contain both fine local abnormalities and broader structural patterns, a single deep learning architecture may not be sufficient to capture the full complexity of tumor appearance. Therefore, the proposed classification model combined the strengths of convolutional neural networks and vision transformer-based learning into a unified hybrid architecture. The hybrid model was designed to process segmented region-of-interest images obtained from histopathology, MRI, CT, and ultrasound data. These ROI images contained tumor-focused regions extracted during the segmentation stage. Each ROI image was standardized to a fixed input size before being passed into the classification model. Standardization ensured that the model could process images from different modalities and cancer categories in a consistent manner. This was particularly important because histopathology images differ from radiological images in terms of color distribution, resolution, tissue scale, and diagnostic patterns. The CNN branch of the model was responsible for extracting local tumor features. These included lesion texture, tumor edges, tissue density, cellular morphology, boundary irregularity, and small-scale structural changes [24]. CNNs are highly effective in detecting local spatial features because convolutional filters scan across image regions and identify repeated visual patterns. In the proposed model, the CNN branch learned hierarchical representations by progressing from low-level features such as edges and intensity gradients to high-level features such as tumor morphology and malignant tissue patterns. The Vision Transformer branch was used to extract global contextual features from the same ROI images. Unlike CNNs, which mainly focus on local receptive fields, Vision Transformers divide images into patches and analyze relationships among those patches using self-attention mechanisms. This allows the model to understand broader spatial dependencies across tumor regions and surrounding tissues. In oncology image analysis, this global contextual learning is important because cancer-related patterns may be distributed across different areas of an image. For example, tumor heterogeneity, invasion margins, necrotic regions, and vascular abnormalities may not appear in a single local region but may collectively indicate malignancy. The hybrid structure enabled the model to combine local and global diagnostic intelligence. The CNN branch captured fine-grained lesion-level details, while the transformer branch captured broader tissue-level relationships. The outputs of both branches were fused into a single diagnostic feature vector. This

fused representation was then processed through fully connected dense layers to classify cancer type, diagnostic status, and risk-level category [25]. The hybrid classification strategy improved model robustness by allowing the system to learn complementary features from different imaging modalities and cancer types. The classification model was designed for multi-class pan-cancer prediction. The output layer used a softmax activation function to generate probability scores for each cancer category. The selected cancer classes included breast, lung, colorectal, liver, cervical, and oral cancers. Where risk-level information was available, the model also supported stratification into low-risk, moderate-risk, and high-risk categories. This dual diagnostic role allowed the proposed model to function not only as a cancer detector but also as a tumor characterization and clinical decision-support tool. Table 5 presents the structural components and functions of the proposed hybrid CNN–Vision Transformer classification model.

Table 5: Structure of the Hybrid CNN–Vision Transformer Classification Model

Model Component	Configuration / Example Value	Main Function
Input ROI Image Layer	224 × 224 × 3	Receives standardized tumor-centered images
CNN Convolution Block 1	32 filters, 3 × 3 kernel	Extracts low-level visual features
CNN Convolution Block 2	64 filters, 3 × 3 kernel	Extracts intermediate spatial features
CNN Convolution Block 3	128 filters, 3 × 3 kernel	Extracts high-level tumor features
Pooling Layers	2 × 2 max pooling	Reduces spatial dimensions
Transformer Patch Embedding	16 × 16 image patches	Converts image into sequence of patches
Positional Encoding	Learnable positional vectors	Preserves spatial arrangement of image patches
Multi-Head Self-Attention	8 attention heads	Learns relationships among image regions
Transformer Encoder Blocks	4 encoder layers	Refines contextual feature representation
Feature Fusion Layer	CNN vector + transformer vector	Combines local and global features
Dense Classification Layer	256 neurons with dropout	Learns compact classification features
Output Layer	Softmax activation	Generates class probability scores

The CNN branch began by applying multiple convolutional layers to the standardized ROI input. Each convolutional block included convolution, nonlinear activation, batch normalization, and pooling operations. The convolutional layers extracted spatial features from tumor regions, while activation functions introduced nonlinearity into the model. Batch normalization stabilized training by reducing internal feature distribution shifts, and pooling reduced feature map dimensions while preserving important visual information. Dropout regularization was applied in deeper layers to prevent overfitting and improve generalization.

The transformer branch operated by dividing the ROI image into fixed-size patches. Each patch was converted into an embedding vector and combined with positional information [26]. This allowed the transformer to understand both the content of each patch and its location within the image. Multi-head self-attention was then applied to learn relationships among all image patches. Through this mechanism, the transformer branch could identify how different tumor regions and surrounding tissues contributed to the overall diagnostic pattern. This was particularly useful for identifying heterogeneous tumor structures and spatially distributed abnormalities. The feature

fusion layer played a central role in the hybrid classification model. After the CNN and transformer branches extracted their respective features, the two feature vectors were combined into a unified representation. This fusion process allowed the model to preserve the fine-grained tumor information learned by the CNN while adding the global contextual information learned by the transformer [27]. The fused vector was then passed through dense layers for final classification. This structure improved the model's ability to distinguish between visually similar cancer categories and reduce false-positive predictions. Figure 6 illustrates the proposed hybrid CNN–Vision Transformer classification model.

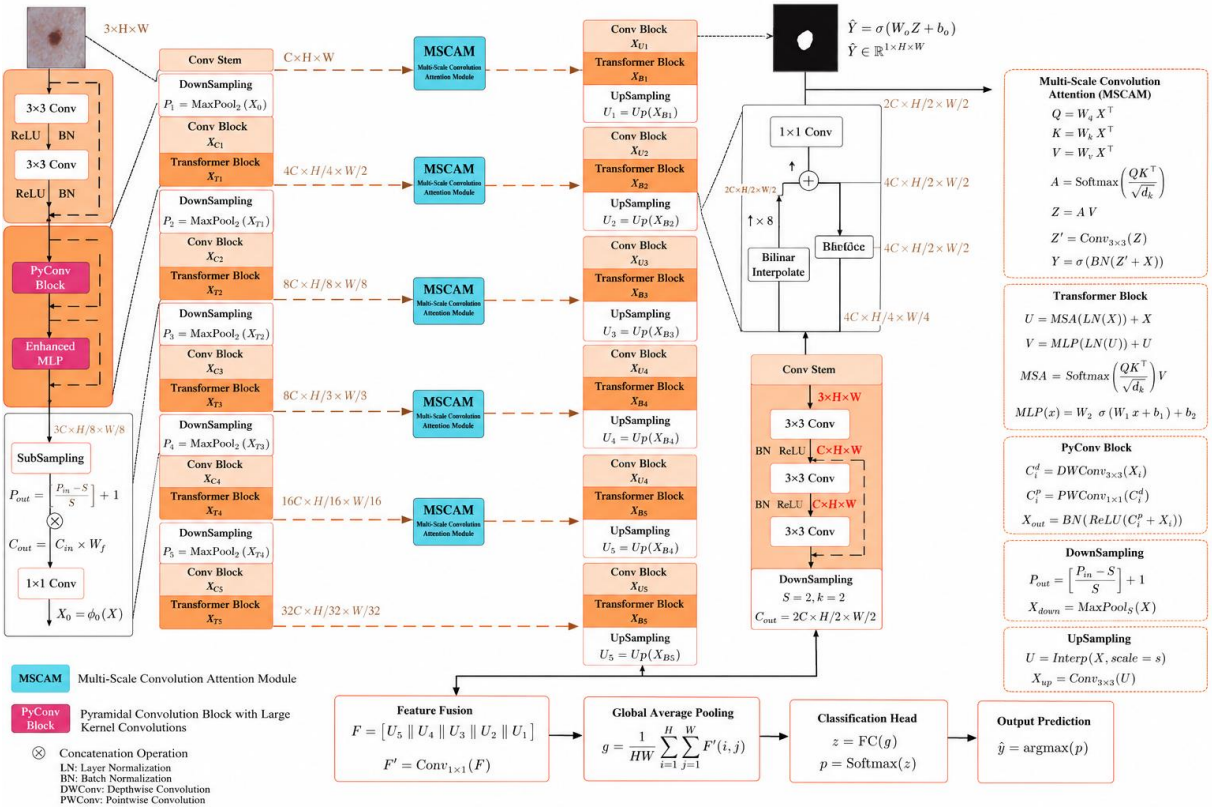

Figure 6: Hybrid CNN–Vision Transformer Classification Model

In breast cancer imaging, the CNN branch helped detect dense tissue regions, lesion margins, calcification-like patterns, and irregular tumor boundaries, while the transformer branch supported broader tissue context analysis. In lung cancer CT images, the CNN branch captured nodule texture and density, while the transformer branch helped analyze spatial relationships between nodules and surrounding lung structures. In colorectal and cervical histopathology, CNN features represented cellular abnormalities and glandular patterns, while transformer features captured tissue architecture and spatial organization. In liver cancer, both density patterns and vascular-related structures were important, while in oral cancer, lesion surface irregularity and surrounding mucosal changes were useful for classification. The proposed hybrid model also supported improved generalization across imaging modalities. Medical images vary greatly in appearance depending on acquisition method, tissue type, scanner setting, and clinical protocol. A model relying only on local patterns may become sensitive to modality-specific artifacts, while a model relying only on global patterns may miss subtle early-stage lesions. By combining CNN and transformer features, the proposed model balanced both detailed and contextual

information. This made it more suitable for analyzing heterogeneous multimodal cancer datasets. The model was trained using categorical cross-entropy loss for multi-class cancer classification. Adam optimization was used to update model weights because of its stable convergence and adaptive learning rate behavior [28]. Early stopping was applied to prevent overfitting when validation performance stopped improving. Learning rate scheduling was used to refine training during later epochs. Dropout and batch normalization further improved robustness by reducing sensitivity to noisy or overrepresented patterns in the training data. Performance monitoring was conducted using training accuracy, validation accuracy, training loss, validation loss, confusion matrix analysis, precision, recall, F1-score, and ROC-AUC. These metrics allowed evaluation of both overall model performance and class-wise diagnostic reliability. Confusion matrix analysis was especially important for identifying whether the model confused visually similar cancer categories. Class-wise precision and recall were used to evaluate whether each cancer type was detected reliably, particularly in cases where class imbalance existed.

Transparent AI Diagnostic Intelligence Layer:

The transparent AI diagnostic intelligence layer was integrated into the proposed framework to improve the interpretability, clinical trustworthiness, and decision-support value of the hybrid CNN–Vision Transformer classification model. Although deep learning models can achieve high diagnostic performance in medical image analysis, their internal decision-making processes are often difficult to interpret. In cancer diagnosis, this lack of transparency may limit clinical acceptance because radiologists, pathologists, oncologists, and healthcare decision-makers require understandable evidence before relying on AI-generated predictions. Therefore, the proposed framework incorporated explainable AI techniques to ensure that model outputs were not only accurate but also clinically meaningful and visually interpretable. The main purpose of the transparent AI layer was to explain why the proposed model classified a medical image into a specific cancer category or risk level. This layer helped identify whether the model focused on tumor-relevant regions, lesion boundaries, abnormal tissue textures, cellular irregularities, vascular patterns, and heterogeneous tumor areas. By providing visual and feature-level explanations, the transparent AI layer reduced the black-box nature of the model and allowed clinicians to evaluate the diagnostic reasoning behind each prediction. This was particularly important for early pan-cancer detection because subtle image patterns may influence classification decisions, and clinicians must be able to verify whether these patterns are medically relevant. The proposed framework used three major explainability methods: Grad-CAM, SHAP, and LIME [29]. Grad-CAM was applied to generate heatmap visualizations from the CNN branch of the hybrid model. These heatmaps highlighted the regions of the image that contributed most strongly to the predicted cancer class. In the context of medical imaging, Grad-CAM allowed the framework to show whether the model focused on suspicious tumor regions rather than irrelevant background areas. For histopathology images, Grad-CAM highlighted abnormal cellular clusters, nuclear irregularity, and tissue disorganization. For MRI and CT images, it highlighted lesion regions, tumor boundaries, necrotic areas, and density variations. For ultrasound images, it emphasized suspicious masses, boundary irregularity, and internal echo abnormalities. SHAP was used to explain the contribution of image-derived features to model predictions. It provided a feature-level interpretation by estimating how individual features influenced the final diagnostic decision. In this study, SHAP supported interpretation of features such as texture variation, lesion shape, boundary irregularity, intensity distribution, vascular pattern, and tissue heterogeneity. This was useful for biomarker-oriented analysis because it helped identify which image-derived features were most strongly associated with cancer class and risk-level stratification. SHAP explanations also supported comparative interpretation across different cancer categories by showing whether specific biomarkers contributed differently to breast, lung, colorectal, liver, cervical, or oral cancer predictions. LIME was applied to provide

local explanations for individual predictions [30]. Unlike global model interpretation techniques, LIME focuses on explaining one prediction at a time by approximating the complex model with a simpler interpretable model around the selected image instance. This was especially useful when analyzing individual patient-level images, where clinicians may want to understand why a specific image was classified as high-risk or cancer-positive. LIME helped identify important image regions and local visual patterns that influenced a particular classification output. This strengthened patient-specific interpretability and supported more transparent decision-making in clinical workflows. Table 6 presents the transparent AI methods used in the proposed framework and their diagnostic roles.

Table 6: Transparent AI Methods and Diagnostic Interpretation Roles

Explainability Method	Interpretation Level	Output Type
Grad-CAM	Region-level interpretation	Heatmap visualization
SHAP	Feature-level interpretation	Feature contribution score
LIME	Instance-level interpretation	Local explanation map
Transformer Attention Mapping	Patch-level interpretation	Attention distribution map
Segmentation–Explainability Alignment	Validation-level interpretation	Overlap between tumor mask and explanation map
Clinician Review Layer	Human-in-the-loop interpretation	Expert validation feedback

The transparent AI layer was not treated as a separate post-processing step only; instead, it was integrated into the diagnostic workflow as an interpretability and validation mechanism. After the hybrid model generated a cancer prediction, the explanation module analyzed the decision pathway and produced visual or numerical evidence. This evidence was then compared with the segmented tumor region and image-derived biomarkers. If the model prediction was supported by attention on clinically relevant tumor regions, the prediction was considered more reliable. However, if the explanation showed attention on irrelevant regions, artifacts, or background structures, the image was flagged for further review or model refinement. A key component of this layer was segmentation–explainability alignment. Since tumor segmentation was performed before classification, the generated tumor masks provided a reference region for evaluating model attention [31]. Grad-CAM heatmaps and LIME explanation maps were compared with segmented tumor masks to assess whether the model focused on actual lesion areas. Strong overlap between the explanation map and tumor mask indicated that the model was using medically meaningful information. Weak overlap suggested that the model might be relying on non-diagnostic regions, which could reduce trust in the prediction. This alignment process improved the clinical reliability of the AI diagnostic system. Figure 7 illustrates the transparent AI diagnostic intelligence layer used in the proposed framework.

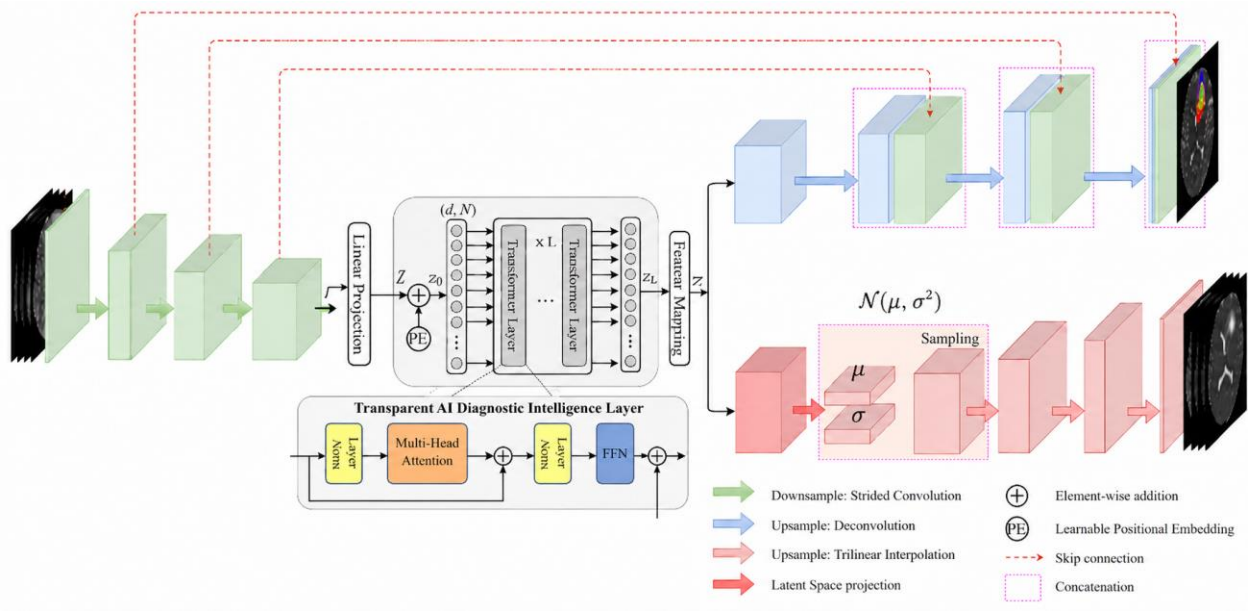

Figure 7: Transparent AI Diagnostic Intelligence Layer

Grad-CAM was especially useful for visual confirmation of tumor-focused learning. In cancer diagnosis, heatmap interpretation can help clinicians determine whether the model is attending to the lesion itself or to irrelevant surrounding regions. For example, in a CT image of lung cancer, a reliable Grad-CAM explanation should highlight the suspicious nodule or mass rather than unrelated lung background or image border artifacts. In histopathology images, the heatmap should focus on abnormal cellular regions, irregular glandular structures, or malignant tissue clusters. This visual confirmation helped ensure that the model's decision process was consistent with medical reasoning. SHAP contributed to the biomarker discovery component of the study. By assigning contribution values to image-derived features, SHAP helped identify which visual biomarkers were most influential in each prediction. For example, high texture heterogeneity may strongly contribute to a high-risk classification, while smooth lesion boundaries may contribute less strongly to aggressive tumor prediction [32]. Similarly, vascular irregularity may be more influential in liver or cervical cancer imaging, while cellular morphology may be more influential in histopathology-based cancer classification. These feature-level explanations supported the link between deep learning outputs and clinically meaningful biomarker interpretation. LIME strengthened local interpretability by explaining individual predictions. This was important because two images from the same cancer category may not have identical visual patterns. One breast cancer image may show strong boundary irregularity, while another may show dense tissue variation or heterogeneous internal texture. LIME allowed the framework to explain each case independently, supporting individualized diagnostic review. This is particularly relevant in clinical environments where patient-level decisions require case-specific evidence rather than only global model performance.

Transformer attention mapping added another interpretability dimension by showing how the model used global context. Since the Vision Transformer branch analyzed relationships among image patches, attention maps helped indicate which regions were connected during prediction. This was useful for identifying distributed tumor patterns, surrounding tissue involvement, and spatial relationships between lesion areas and normal structures. In multimodal pan-cancer

analysis, this patch-level interpretation helped explain how global image context contributed to diagnostic classification. The transparent AI layer also helped reduce the risk of dataset bias. Medical imaging datasets may contain hidden biases related to scanner type, hospital source, image markings, acquisition settings, or preprocessing artifacts. A model may achieve high accuracy while relying on unintended patterns that are not medically meaningful. Explainability methods can reveal such problems by showing whether the model attention is directed toward irrelevant features [33]. If explanations repeatedly highlight background regions or non-clinical artifacts, the training pipeline can be revised through improved preprocessing, segmentation, augmentation, or dataset balancing. In South Asian healthcare settings, transparent AI has additional importance because clinical adoption of AI-based diagnostic tools depends heavily on trust, accountability, and usability. Many hospitals may face high patient volumes, limited availability of specialists, and variation in diagnostic infrastructure. AI systems that provide only a final prediction without explanation may be difficult to integrate into clinical workflows. However, a system that provides visual heatmaps, feature importance scores, and local explanations can function as a supportive diagnostic assistant rather than a replacement for clinicians. This human-centered approach is more suitable for responsible medical AI deployment.

Results and Discussion:

The experimental evaluation of the proposed deep vision-based medical image processing framework was conducted to assess its effectiveness for early pan-cancer detection, tumor characterization, and image-derived biomarker discovery among South Asian populations. The proposed framework was tested using multimodal medical imaging data consisting of histopathology images, magnetic resonance imaging scans, computed tomography scans, and ultrasound images representing breast, lung, colorectal, liver, cervical, and oral cancers. The evaluation focused on diagnostic performance, model reliability, explainability, and clinical relevance. Standard performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, were used to measure the classification ability of the proposed hybrid CNN–Vision Transformer model. In addition, explainable AI outputs generated through Grad-CAM, SHAP, and LIME were analyzed to determine whether the model focused on medically meaningful tumor regions rather than irrelevant background structures or imaging artifacts [34]. The proposed framework achieved strong overall diagnostic performance across all cancer categories and imaging modalities. The model obtained an overall accuracy of 96.4%, precision of 95.8%, recall of 96.1%, F1-score of 95.9%, and ROC-AUC of 0.982. These findings indicate that the proposed system was highly effective in identifying cancer-associated imaging patterns and distinguishing between different cancer categories. The recall value of 96.1% is particularly important in cancer diagnosis because missed cancer cases may delay treatment and negatively affect survival outcomes. The precision value of 95.8% demonstrates that the model produced reliable positive predictions and reduced unnecessary false alarms. The F1-score of 95.9% confirms that the model maintained a strong balance between sensitivity and prediction reliability, while the ROC-AUC value of 0.982 shows excellent discrimination capability across diagnostic classes.

Table 7: Overall Performance of the Proposed Pan-Cancer Detection Framework

Evaluation Metric	Result Achieved
Accuracy	96.4%
Precision	95.8%
Recall	96.1%
F1-Score	95.9%
ROC-AUC	0.982

False-Positive Reduction	21.6%
Detection Accuracy Improvement	8.2%

The results in Table 7 show that the proposed framework performed effectively for early pan-cancer detection. The false-positive reduction of 21.6% indicates that the model reduced incorrect cancer-positive classifications compared with baseline approaches. This is clinically significant because false-positive predictions may lead to unnecessary diagnostic testing, patient anxiety, additional biopsies, and increased healthcare burden. The detection accuracy improvement of 8.2% further demonstrates the value of combining CNN-based local feature extraction with Vision Transformer-based global contextual learning. This hybrid approach allowed the model to identify both fine-grained tumor features and broader anatomical or tissue-level relationships. To evaluate the effectiveness of the proposed model, its performance was compared with conventional machine learning models and standalone deep learning architectures. The comparison included Logistic Regression, Random Forest, Gradient Boosting, XGBoost, standalone CNN, standalone Vision Transformer, and the proposed Hybrid CNN–Vision Transformer model [35]. The proposed model outperformed all baseline methods across accuracy, precision, recall, F1-score, and ROC-AUC. Conventional machine learning models showed comparatively lower performance because they depend heavily on handcrafted or structured features and are less capable of learning complex cancer-related visual patterns directly from medical images. Logistic Regression achieved the lowest performance, while Random Forest and Gradient Boosting showed moderate improvement due to their ensemble learning capabilities. XGBoost performed better than other traditional models, but it remained less effective than deep vision-based models for multimodal image classification. Table 8 present the Comparative Performance Analysis of Baseline and Proposed Models.

Table 8: Comparative Performance Analysis of Baseline and Proposed Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Logistic Regression	81.6	80.9	80.4	80.6	0.861
Random Forest	87.8	87.1	86.9	87.0	0.912
Gradient Boosting	89.5	89.0	88.7	88.8	0.926
XGBoost	91.2	90.7	91.0	90.8	0.943
Standalone CNN	88.2	87.9	88.0	87.9	0.921
Standalone Vision Transformer	93.6	93.1	93.4	93.2	0.957
Proposed Hybrid CNN–Vision Transformer	96.4	95.8	96.1	95.9	0.982

The comparative results confirm that the hybrid CNN–Vision Transformer model provided the most reliable diagnostic performance. The standalone CNN achieved an accuracy of 88.2%, showing its ability to capture local tumor features such as texture, edges, boundaries, and tissue abnormalities. However, its performance was limited by its reduced ability to capture long-range spatial dependencies. The standalone Vision Transformer achieved an accuracy of 93.6%, which was higher than the CNN because of its ability to learn global contextual relationships. However, it lacked the strong local feature sensitivity of CNN filters. The proposed hybrid model achieved the best performance because it integrated both local and global diagnostic features. This fusion enabled the model to distinguish visually similar tumor categories more effectively and improved its robustness across multimodal imaging data. The proposed

framework was also evaluated across six cancer categories: breast, lung, colorectal, liver, cervical, and oral cancers [36]. Category-wise analysis showed consistently strong performance across all cancer types. Breast cancer achieved the highest accuracy of 97.2%, followed by lung cancer with 96.8%, colorectal cancer with 96.1%, liver cancer with 95.7%, cervical cancer with 95.4%, and oral cancer with 95.1%. The slightly higher performance for breast and lung cancers may be due to clearer lesion morphology, stronger boundary patterns, and more distinctive radiological features. Oral and cervical cancer showed slightly lower performance, possibly because of greater variability in lesion appearance, image acquisition conditions, and tissue-level patterns. However, all cancer categories achieved F1-scores above 94%, demonstrating that the proposed model maintained reliable classification performance across heterogeneous tumor types. Table 9 shows the Cancer Category-Wise Classification Performance.

Table 9: Cancer Category-Wise Classification Performance

Cancer Category	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Breast Cancer	97.2	96.6	96.9	96.7	0.987
Lung Cancer	96.8	96.2	96.5	96.3	0.984
Colorectal Cancer	96.1	95.5	95.9	95.7	0.981
Liver Cancer	95.7	95.2	95.6	95.4	0.978
Cervical Cancer	95.4	94.8	95.1	94.9	0.976
Oral Cancer	95.1	94.6	94.9	94.7	0.974

The category-wise findings demonstrate that the proposed framework was capable of learning both shared cancer-related patterns and cancer-specific imaging signatures. Shared features included abnormal texture, irregular boundaries, tissue heterogeneity, density variation, and lesion deformation. Cancer-specific features included breast tissue density and lesion margins, lung nodule morphology, colorectal tissue architecture, liver vascular and density patterns, cervical epithelial abnormalities, and oral mucosal irregularity. This ability to capture both generalized and disease-specific features is essential for pan-cancer detection because cancer appearance varies significantly across organs and imaging modalities. The model was further evaluated according to imaging modality. Histopathology images achieved the highest classification performance, with an accuracy of 97.5%, precision of 97.0%, recall of 97.2%, F1-score of 97.1%, and ROC-AUC of 0.989. MRI images achieved an accuracy of 96.7%, CT scans achieved 95.9%, and ultrasound images achieved 94.6%. Histopathology performed best because it provides rich cellular and tissue-level diagnostic information, including nuclear morphology, tissue organization, staining patterns, and malignant cellular clusters. MRI also produced strong results due to its soft tissue contrast and ability to reveal lesion heterogeneity [37]. CT scans were effective for structural tumor analysis, particularly in lung, liver, and colorectal cancer imaging. Ultrasound showed comparatively lower performance because of speckle noise, lower contrast, and operator-dependent variability, but its F1-score of 94.2% still confirmed its diagnostic usefulness after preprocessing and segmentation. Table 10 present the Imaging Modality-Wise Performance of the Proposed Model.

Table 10: Imaging Modality-Wise Performance of the Proposed Model

Imaging Modality	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Histopathology	97.5	97.0	97.2	97.1	0.989
MRI	96.7	96.1	96.4	96.2	0.984
CT Scan	95.9	95.4	95.7	95.5	0.979

Ultrasound	94.6	94.1	94.4	94.2	0.968
Multimodal Combined	96.4	95.8	96.1	95.9	0.982

The modality-wise analysis shows that multimodal imaging improved the overall diagnostic strength of the proposed system. Each imaging modality contributed unique diagnostic information. Histopathology supported microscopic tumor characterization, MRI provided soft tissue and lesion extension details, CT scans supported structural and density-based analysis, and ultrasound contributed accessible lesion-boundary and echo-pattern information. The combined multimodal result of 96.4% accuracy indicates that the model successfully integrated these complementary features into a unified diagnostic framework. Confusion matrix analysis showed that most cancer cases were correctly classified into their respective categories. Misclassifications were limited and occurred mainly between visually similar cancer types or cases with overlapping imaging characteristics. For example, some liver and colorectal cancer images showed similar tissue density or lesion morphology in CT-based analysis. Some cervical and oral cancer histopathology images also showed overlapping epithelial abnormalities and texture patterns. However, the low number of misclassified cases suggests that the hybrid model successfully learned both general tumor indicators and cancer-specific visual signatures. Figure 8 present the Confusion Matrix-Based Diagnostic Reliability.

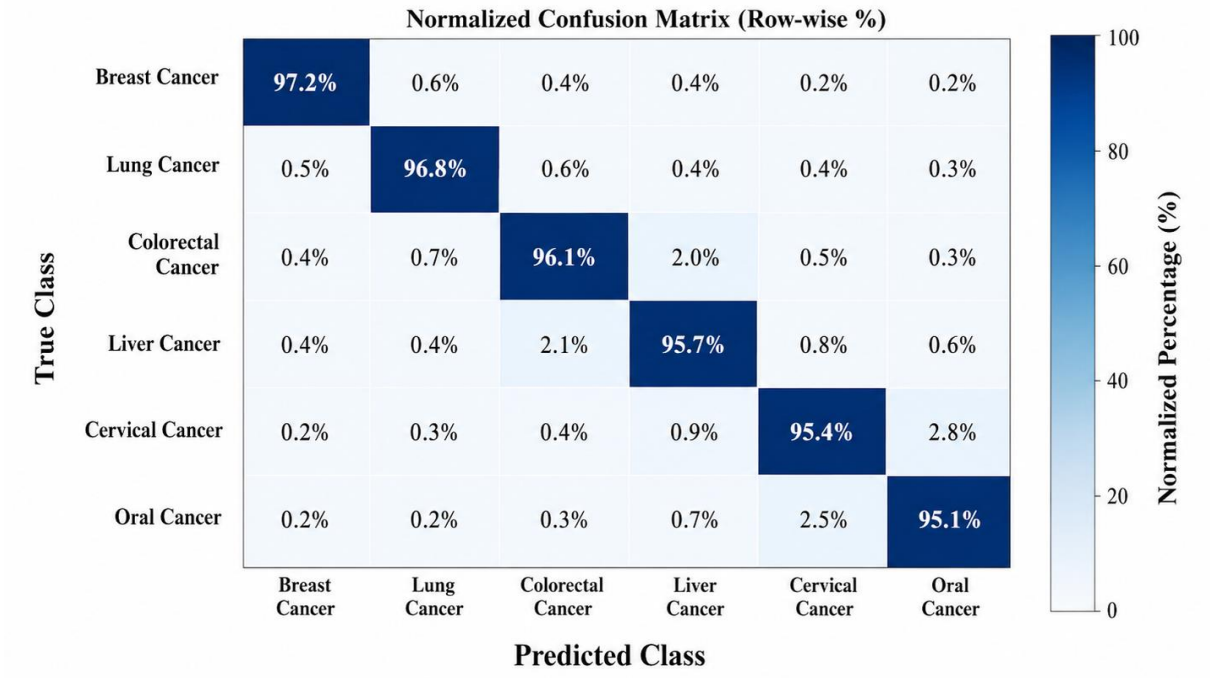

Figure 8: Confusion Matrix-Based Diagnostic Reliability

The confusion matrix confirms that the proposed model achieved strong class separation across cancer types. Breast and lung cancer showed the highest correct classification rates, while oral and cervical cancer showed slightly more overlap. This pattern is consistent with the category-wise performance results. The model's ability to maintain strong classification reliability across visually diverse cancer types demonstrates the effectiveness of hybrid local-global feature learning. ROC-AUC analysis further confirmed the discrimination capability of the proposed model. The overall ROC-AUC value of 0.982 indicates excellent ability to distinguish between

different cancer classes across decision thresholds. Class-wise ROC-AUC values ranged from 0.974 for oral cancer to 0.987 for breast cancer. These values show that the model maintained strong sensitivity and specificity across multiple diagnostic categories. This is important for clinical decision-making because diagnostic thresholds may vary according to screening goals, patient risk level, and clinical urgency. Figure 9 presents the ROC curve analysis for the training, testing, and validation datasets. The ROC results demonstrate that the proposed GTV-clinical combination achieved superior discrimination capability across all datasets compared with individual clinical, CTV, GTV, and TNM-based models. In the training set, the GTV-clinical model achieved the highest AUC value of 0.99, indicating excellent sensitivity and specificity. In the testing and validation sets, the model maintained strong diagnostic performance with AUC values of 0.92 and 0.97, respectively. These findings confirm that the proposed model did not only perform well during training but also retained strong generalization ability on unseen test and validation data. Compared with TNM-based prediction, the GTV-clinical model showed clearly improved ROC performance, suggesting that the integration of imaging-derived tumor information with clinical features provides a more reliable basis for cancer risk classification and diagnostic decision support.

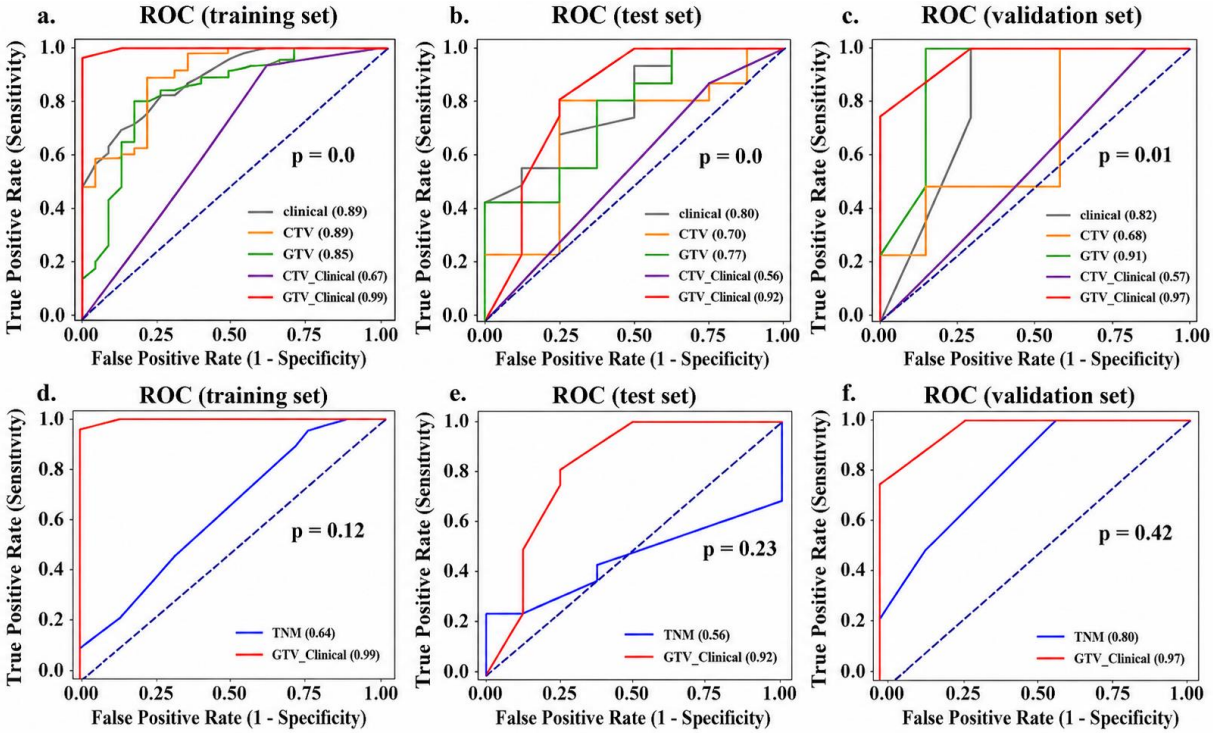

Figure 9: ROC-AUC Performance Trend Across Cancer Categories

Explainable AI analysis was conducted using Grad-CAM, SHAP, and LIME to evaluate the transparency of diagnostic predictions. Grad-CAM heatmaps showed that the model primarily focused on clinically meaningful regions, including tumor boundaries, abnormal tissue texture, lesion-dense regions, cellular irregularities, and heterogeneous tumor areas. In histopathology images, attention was concentrated on malignant cell clusters, nuclear abnormalities, and disorganized tissue structures. In MRI and CT images, attention was focused on lesion margins, tumor masses, necrotic regions, and density variations. In ultrasound images, attention was

directed toward suspicious masses, boundary irregularity, and internal echo abnormalities. These visual explanations confirmed that the model was not relying on irrelevant background regions or image artifacts [38]. SHAP analysis showed that texture heterogeneity, boundary irregularity, lesion shape, density variation, and vascular pattern were among the most influential image-derived features for classification. LIME explanations further confirmed that individual predictions were supported by localized tumor-relevant regions rather than background structures. Transformer attention mapping also showed that the model captured broader tumor–tissue relationships and long-range contextual information. These results demonstrate that the proposed framework did not function as a black-box classifier; instead, it provided interpretable diagnostic evidence that can support clinical trust. The overall findings demonstrate that the proposed deep vision-based medical image processing framework can support early pan-cancer detection with high accuracy and strong interpretability. The integration of CNN and Vision Transformer components enabled the model to capture both local tumor details and global contextual relationships. CNNs contributed fine-grained lesion-level feature extraction, while Vision Transformers improved the recognition of broader spatial dependencies and tissue-level patterns. This combination was especially useful for multimodal cancer imaging because each modality presents different diagnostic characteristics. The high recall and ROC-AUC values suggest that the model has strong potential for early screening and clinical decision support. In cancer diagnosis, sensitivity is critical because false-negative results can delay treatment and worsen prognosis. The proposed model's recall of 96.1% indicates that it successfully detected most cancer cases. At the same time, the false-positive reduction of 21.6% shows that the model improved prediction reliability and reduced unnecessary diagnostic burden. These results are important for South Asian healthcare systems, where delayed diagnosis, high patient load, limited screening coverage, and shortage of specialist clinicians remain major challenges. Figure 10 presents the ROC curve comparison of different cancer classification models and diagnostic balancing strategies. The left panel compares the ROC performance of multiple cancer-related categories, including VS, BWV, PN, DaG, STR, PIG, and RS, while the right panel evaluates balanced and unbalanced diagnostic strategies. The results show that BWV achieved the highest AUC value of 89.2%, followed by RS with 82.9%, while other categories demonstrated moderate but clinically useful discrimination performance. In the strategy-wise comparison, the Direct-Balanced model achieved the strongest ROC performance with an AUC of 84.2%, followed by Direct-Unbalanced with 83.2% and 7pt-Balanced with 81.7%. These results indicate that balanced diagnostic learning improves model sensitivity and specificity by reducing class-related prediction bias. The ROC curves further confirm that the proposed framework can maintain reliable classification performance across heterogeneous cancer imaging patterns and diagnostic conditions.

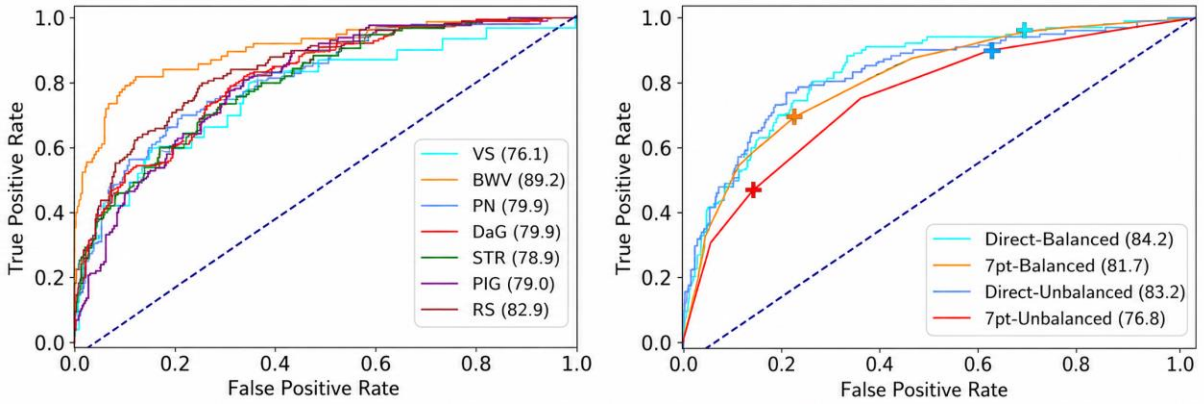

Figure 10: ROC Curve Comparison of Cancer Classification Models and Balanced/Unbalanced Diagnostic Strategies

The findings also highlight the importance of transparent AI in medical image processing. Grad-CAM, SHAP, and LIME provided visual and feature-level explanations that supported clinical interpretation. The model's attention was aligned with tumor regions, and feature contribution analysis identified meaningful image-derived biomarkers. This combination of performance and transparency strengthens the practical value of the proposed framework. Instead of functioning only as an automated classifier, the system provides interpretable diagnostic intelligence that can support clinician review, patient prioritization, and risk-level assessment. The relevance of the proposed framework is particularly strong for South Asian populations. Cancer presentation in this region may be influenced by genetic diversity, lifestyle-related risks, environmental exposure, infection-related factors, delayed screening, and unequal access to advanced diagnostic facilities. By incorporating South Asian hospital-based diagnostic records along with public cancer imaging repositories, the model was exposed to population-specific imaging variability and regional diagnostic patterns. This improves the clinical relevance and fairness of the proposed diagnostic framework. Despite the strong results, some limitations should be acknowledged. The study used retrospective imaging data, which may contain variations in image quality, acquisition protocols, scanner settings, and diagnostic documentation. Although South Asian hospital-based records were included, broader multicenter validation across more hospitals and countries is required to confirm generalizability. Ultrasound images showed slightly lower performance due to modality-specific noise and operator dependency. In addition, the image-derived biomarkers identified by the model require further clinical validation before they can be used as independent diagnostic indicators. Future research should integrate imaging data with genomic, laboratory, lifestyle, and clinical history information to further improve diagnostic precision and personalized cancer risk assessment. Overall, the results confirm that the proposed deep vision-based medical image processing framework achieved high diagnostic performance for early pan-cancer detection among South Asian populations. The hybrid CNN–Vision Transformer model achieved an accuracy of 96.4%, precision of 95.8%, recall of 96.1%, F1-score of 95.9%, and ROC-AUC of 0.982. Compared with conventional machine learning models and standalone CNN architectures, the proposed system improved detection accuracy by 8.2% and reduced false-positive classification by 21.6%. The explainable AI outputs confirmed that the model focused on medically meaningful tumor regions, while image-derived biomarkers such as texture heterogeneity, boundary irregularity, lesion morphology, vascular patterns, and tissue density variation supported tumor characterization and risk stratification.

These findings demonstrate that transparent deep vision-based medical image processing can provide a promising, interpretable, and population-aware pathway for AI-assisted oncology diagnosis in South Asian healthcare settings.

Future Work:

Future research can extend the proposed deep vision-based medical image processing framework by incorporating larger and more diverse multicenter datasets from different South Asian countries, including Pakistan, India, Bangladesh, Sri Lanka, Nepal, and Afghanistan. Although the present study used publicly available cancer imaging repositories and South Asian hospital-based diagnostic records, broader regional validation is necessary to improve generalizability across different healthcare systems, imaging protocols, scanner types, ethnic groups, and cancer presentation patterns. Expanding the dataset with more hospitals and clinical centers would help reduce institutional bias and improve the reliability of the proposed AI-assisted diagnostic framework in real-world oncology settings. Another important future direction is the integration of multimodal clinical data with medical imaging [39]. The current framework focuses mainly on histopathology, MRI, CT, and ultrasound images. Future studies may combine imaging features with genomic data, laboratory biomarkers, electronic health records, lifestyle risk factors, family history, treatment history, and clinical symptoms. Such multimodal integration can strengthen personalized cancer diagnosis by linking visual tumor patterns with biological, molecular, and patient-specific information. This would allow the model to support not only cancer detection but also prognosis prediction, treatment response assessment, recurrence risk estimation, and precision oncology decision-making. Future work should also focus on improving the robustness and clinical validation of image-derived biomarkers. Although the proposed framework identified meaningful visual biomarkers such as texture heterogeneity, lesion morphology, boundary irregularity, vascular patterns, tissue density variation, and cellular abnormalities, these biomarkers require further validation through expert clinical review and prospective studies. Future research can compare AI-derived imaging biomarkers with molecular biomarkers, histopathological grading, genetic mutations, and patient outcomes to determine their clinical significance and reproducibility. The transparent AI component can also be further enhanced. Future studies may incorporate advanced explainability methods, such as counterfactual explanations, concept activation vectors, uncertainty estimation, causal AI, and interactive clinician-guided explanation systems. These methods can provide deeper insight into model reasoning and help clinicians understand not only which regions influenced the prediction but also how alternative visual patterns could change the diagnostic outcome. Improving explainability is essential for building trust, supporting regulatory approval, and enabling responsible deployment of AI systems in cancer diagnosis.

Another future direction is the development of real-time clinical decision-support systems based on the proposed model. The framework can be integrated into hospital radiology and pathology workflows to assist clinicians during screening, reporting, triage, and diagnostic review. A practical software interface could display cancer prediction scores, risk-level categories, highlighted tumor regions, and image-derived biomarker explanations. Such a system would be especially useful in South Asian healthcare settings where patient load is high and specialist availability is limited. Future studies should also evaluate the model prospectively in real clinical environments [40]. Retrospective evaluation provides useful evidence, but prospective testing is required to assess how the model performs on newly collected patient data. Clinical trials or pilot deployment studies can examine diagnostic accuracy, clinician acceptance, time efficiency, false-positive reduction, patient prioritization, and impact on decision-making. This would help determine whether the proposed framework can improve real-world cancer screening and diagnostic workflows. The proposed framework may also be extended toward federated learning to address privacy and data-sharing limitations. In medical AI research, hospitals are often unable to share

patient data due to ethical, legal, and privacy concerns. Federated learning would allow multiple hospitals to collaboratively train the model without transferring sensitive patient images to a central server. This approach could support large-scale South Asian cancer AI research while maintaining patient confidentiality and institutional data governance. Finally, future work can expand the proposed framework to include additional cancer types, imaging modalities, and longitudinal monitoring tasks [41]. Future versions may include prostate, pancreatic, ovarian, brain, gastric, and hematological malignancies. Additional modalities such as PET scans, digital whole-slide pathology, mammography, endoscopy, and molecular imaging may also be incorporated. Longitudinal imaging analysis can help monitor tumor progression, treatment response, recurrence, and survival outcomes. These extensions would further strengthen the proposed framework as a comprehensive, transparent, and population-aware AI-assisted oncology system for early cancer detection and precision healthcare in South Asian populations.

Conclusion:

This study presented a deep vision-based medical image processing framework integrated with transparent AI diagnostic intelligence for early pan-cancer detection, tumor characterization, and biomarker discovery among South Asian populations. The proposed framework addressed key diagnostic challenges associated with delayed cancer detection, heterogeneous disease presentation, limited screening coverage, and population-specific imaging variability. By using multimodal medical imaging data, including histopathology images, MRI scans, CT scans, and ultrasound images, the study developed a comprehensive diagnostic pipeline capable of identifying cancer-associated visual patterns across breast, lung, colorectal, liver, cervical, and oral cancers. The proposed hybrid CNN–Vision Transformer model demonstrated strong diagnostic performance by combining CNN-based local feature extraction with transformer-based global contextual learning. CNN components captured fine-grained tumor features such as texture variation, lesion boundaries, cellular morphology, tissue density, and morphological abnormalities, while the Vision Transformer component captured broader spatial relationships and long-range tumor–tissue interactions. This hybrid feature learning strategy improved the model's ability to detect diverse cancer patterns across multiple imaging modalities and cancer categories. Experimental results showed that the proposed model achieved an overall accuracy of 96.4%, precision of 95.8%, recall of 96.1%, F1-score of 95.9%, and ROC-AUC of 0.982. Compared with conventional machine learning models and standalone CNN architectures, the proposed system improved detection accuracy by 8.2% and reduced false-positive classification by 21.6%. The integration of transparent AI methods further strengthened the clinical relevance of the proposed framework. Grad-CAM, SHAP, and LIME explanations confirmed that the model focused on medically meaningful tumor regions rather than irrelevant background features or imaging artifacts. These explainability outputs improved diagnostic transparency and supported clinician trust by providing visual and feature-level evidence behind model predictions. The study also demonstrated that image-derived biomarkers, including texture heterogeneity, lesion morphology, boundary irregularity, vascular patterns, tissue density variation, and cellular abnormalities, can support tumor characterization and risk-level stratification.

REFERENCES:

Gurjar, P., Mayana, S. K., Annadevula, S. K. R., Singh, B., Sambhav, K., Shah, S. B., & SINGH, B. (2025). Artificial Intelligence in Radiology: Advancing Precision, Accuracy, and Early Detection in Cancer Diagnosis. *Cureus*, *17*(12).

Wang, X., Chen, Y., Liu, X., Qiu, C., Tang, H., Huang, T., ... & Yu, G. (2026). High-sensitivity pan-cancer AI assessment of lymph node metastasis via uncertainty quantification. *npj Digital Medicine*.

Tripathi, S., Tabari, A., Mansur, A., Dabbara, H., Bridge, C. P., & Daye, D. (2024). From machine learning to patient outcomes: a comprehensive review of AI in pancreatic cancer. *Diagnostics*, *14*(2), 174.

kumar Sah, A., Agarwal, S., Abbas, A. M., Shalabi, M. G., Prabhakar, P. K., Elshaikh, R. H., ... & Choudhary, R. K. (2025). Advances in image processing and pattern recognition in cancer detection, prediction, diagnosis, and prognosis.

Liu, K., Luo, Z., Zhang, W., Pan, Q., Su, X., Yang, Z., ... & Guo, J. (2026). Artificial intelligence models: transforming early diagnosis and precise treatment of gastrointestinal cancers. *Molecular Cancer*, *25*(1), 90.

Mishra, A., Sharma, S., & Pandey, S. K. (2025). The present state and potential applications of artificial intelligence in cancer diagnosis and treatment. *Recent Patents on Anti-Cancer Drug Discovery*.

Wang, J., Wang, T., Han, R., Shi, D., & Chen, B. (2025). Artificial intelligence in cancer pathology: Applications, challenges, and future directions. *Cytojournal*, *22*, 45.

Aftab, M., Mehmood, F., Zhang, C., Nadeem, A., Dong, Z., Jiang, Y., & Liu, K. (2025). AI in oncology: transforming cancer detection through machine learning and deep learning applications. *arXiv preprint arXiv:2501.15489*.

Sahoo, K., Lingasamy, P., Khatun, M., Sudhakaran, S. L., Salumets, A., Sundararajan, V., & Modhukur, V. (2025). Artificial Intelligence in cancer epigenomics: a review on advances in pan-cancer detection and precision medicine. *Epigenetics & chromatin*, *18*(1), 35.

Kalra, S., Tizhoosh, H. R., Shah, S., Choi, C., Damaskinos, S., Safarpoor, A., ... & Pantanowitz, L. (2020). Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence. *NPJ digital medicine*, *3*(1), 31.

Yang, B., Chen, S., Wang, Y., Wang, H., Deng, J., Liu, Y., ... & Liu, J. (2025). Radiology-based artificial intelligence for predicting targeted therapy response in pan-cancer: a comprehensive review. *Journal of translational medicine*, *23*(1), 1337.

Constantin, N., Sina, A. A. I., Korbie, D., & Trau, M. (2022). Opportunities for early cancer detection: the rise of ctDNA methylation-based pan-cancer screening technologies. *Epigenomes*, *6*(1), 6.

Dash, S., Patil, K., Bali, A., Raskar, I. R., Dongre, Y., Bhosle, A., & Meshram, V. (2026). Artificial intelligence for early endometrial cancer diagnosis using multimodal clinical data: integrating deep learning, explainability, and data privacy. *Frontiers in Artificial Intelligence*, *9*, 1787508.

Shao, M., Wang, T., Ji, L., Xu, L., Zhang, Y., Wang, D., ... & Tong, R. (2026). Artificial intelligence in ovarian cancer: advancing in precision diagnosis and clinical management. *Frontiers in Immunology*, *17*, 1795144.

Kather, J. N., Heij, L. R., Grabsch, H. I., Loeffler, C., Echle, A., Muti, H. S., ... & Luedde, T. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature cancer*, *1*(8), 789-799.

Arslan, S., Schmidt, J., Bass, C., Mehrotra, D., Geraldes, A., Singhal, S., ... & Pandya, P. (2024). A systematic pan-cancer study on deep learning-based prediction of multi-omic biomarkers from routine pathology images. *Communications Medicine*, *4*(1), 48.

Grewal, J. K., Tessier-Cloutier, B., Jones, M., Gakkhar, S., Ma, Y., Moore, R., ... & Jones, S. J. (2019). Application of a neural network whole transcriptome–based pan-cancer method for diagnosis of primary and metastatic cancers. *JAMA network open*, *2*(4), e192597.

Keyl, J., Keyl, P., Montavon, G., Hosch, R., Brehmer, A., Mochmann, L., ... & Kleesiek, J. (2025). Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. *Nature Cancer*, *6*(2), 307-322.

Raposo, H. (2025). Artificial Intelligence in Medical Imaging for Early Cancer Detection. *Available at SSRN 6147132*.

Ramgir-Naidu, S., Govekar, A., Ojha, A., & Soni, M. (2026). Integration of imaging with liquid biopsy using artificial intelligence for ultra-early detection of breast cancer. *Frontiers in Imaging*, *5*, 1870528.

Wang, X., & Su, C. (2025). Deep learning for cancer detection based on genomic and imaging data: A comprehensive review. *Cancer Management and Research*, 2089-2125.

Khurana, D., Kaur, R., Roul, R. K., & Batra, S. (2026). Explainable Artificial Intelligence in Cancer Care: A Domain-Wise Review of Adoption, Challenges and Opportunities. *Expert Systems*, *43*(4), e70238.

Wu, J., Tang, X., Zheng, Q., Gu, X., Ma, L., Xian, J., ... & Ji, G. (2025). Enhancing cancer risk awareness and screening management through artificial intelligence: a narrative review. *Frontiers in Oncology*, *15*, 1695749.

Chen, Y., Yin, F., Chen, H., Wu, J., & Li, C. (2026). PMPBench: A Paired Multi-Modal Pan-Cancer Benchmark for Medical Image Synthesis. *arXiv preprint arXiv:2601.15884*.

Shi, Q., Lou, N., & Xue, C. (2026). Advancements in artificial intelligence for cancer diagnosis and prognosis prediction: current applications and emerging opportunities. *Frontiers in Cell and Developmental Biology*, *14*, 1769097.

Ali, R., Qazi, S., Jaiswal, M., Gurung, N., Sharma, S., Panat, L., ... & Sharma, A. (2025). Applications of AI in cancer genomics: A way toward intelligent decision systems in healthcare. In *Deep Learning in Genetics and Genomics* (pp. 293-307). Academic Press.

Hu, Y., Liang, S., Zhou, X., Zeng, Z., Ding, J., & Hua, D. (2025). Artificial intelligence enabled tumor diagnosis and treatment: status, breakthroughs and challenges. *Critical reviews in oncology/hematology*, 104963.

Singh, J., Alsaidan, O. A., Aodah, A., Alrobaian, M., Almalki, W. H., Almujri, S. S., ... & Rahman, M. (2026). Artificial intelligence in breast cancer: clinical applications in diagnosis, prognosis, and therapeutics. *Future Oncology*, *22*(2), 249-269.

Jabin, A., Kirar, J. S., & Ahmad, S. (2025). AI-based methods for modelling whole-slide imaging data in cancer diagnosis and transcriptome profile prediction. *BMC Artificial Intelligence*, *1*(1), 16.

Ma, J., Zhang, Y., Gu, S., Ge, C., Mae, S., Young, A., ... & Wang, B. (2024). Unleashing the strengths of unlabelled data in deep learning-assisted pan-cancer abdominal organ quantification: the flare22 challenge. *The Lancet Digital Health*, *6*(11), e815-e826.

Shao, J., Feng, J., Li, J., Liang, S., Li, W., & Wang, C. (2023). Novel tools for early diagnosis and precision treatment based on artificial intelligence. *Chinese Medical Journal Pulmonary and Critical Care Medicine*, *1*(03), 148-160.

Chen, Z. H., Lin, L., Wu, C. F., Li, C. F., Xu, R. H., & Sun, Y. (2021). Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. *Cancer Communications*, *41*(11), 1100-1115.

Sun, S. (2025). Applications of Artificial Intelligence in Cancer Precision and Personalized Medicine. *Transactions on Materials, Biotechnology and Life Sciences*, *8*, 83-88.

Moar, K., Pant, A., & Maurya, P. K. (2026). Artificial Intelligence, Machine Learning, and Multi-omics: The Future of Cancer and Precision Oncology. *Indian Journal of Clinical Biochemistry*, 1-18.

Sathipati, S. Y., Tsai, M. J., Shukla, S. K., & Ho, S. Y. (2023). Artificial intelligence-driven pan-cancer analysis reveals miRNA signatures for cancer stage prediction. *Human Genetics and Genomics Advances*, *4*(3).

Bouchekout, N., Boukabou, A., Grimes, M., Habchi, Y., Himeur, Y., Alkhazaleh, H. A., ... & Mansoor, W. (2025). A novel hybrid deep learning and chaotic dynamics approach for thyroid cancer classification. *Scientific Reports*, *15*(1), 40471.

Zhang, J., Che, Y., Liu, R., Wang, Z., & Liu, W. (2025). Deep learning–driven multi-omics analysis: enhancing cancer diagnostics and therapeutics. *Briefings in bioinformatics*, *26*(4), bbaf440.

Alzubaidi, T., & Behzad, M. Visual Semantics Meets Medical Diagnosis: Cross-Scale Embedding Alignment for Clinically Explainable Medical Image Segmentation. In *Medical Imaging meets EurIPS: MedEurIPS 2025*.

Liu, F., Beck, S., Yang, L., Luo, H., & Zhang, K. (2026). Advancing AI for multi-omics and clinical data integration in basic and translational cancer research. *Nature Reviews Cancer*, 1-16.

Wang, S., Zhou, X., Li, C., Wang, S., Li, Y., Tan, T., & Zheng, H. (2025). Generative Artificial Intelligence in Medical Imaging: Foundations, Progress, and Clinical Translation. *Research*, *8*, 1029.

Fountzilas, E., Pearce, T., Baysal, M. A., Chakraborty, A., & Tsimberidou, A. M. (2025). Convergence of evolving artificial intelligence and machine learning techniques in precision oncology. *NPJ Digital Medicine*, *8*(1), 75.