Proceedings of the 7th International Conference on Inventive Research in Computing Applications (ICIRCA-2026)
DVD Part Number: CFP26N67-DVD; ISBN: 979-8-3315-5939-7

A Hybrid Deep Learning Framework for Liver Cancer Detection using Capsule Networks, Vision Transformers, Graph Neural Networks, and Quantum-Inspired Optimization

[1]Dr.K.Dharmarajan,
Professor, School of
Computing Sciences,
VISTAS, Chennai, India
dharmak07@gmail.com

[2]Dr.K.Abirami, Assistant
Professor, School of
Computing Sciences,
VISTAS, Chennai, India
kabirami.scs@vistas.ac.in

[3]T.Haripriya, Research
scholar, School of Computing
Sciences, VISTAS, Chennai,
India
hariswt9@gmail.com

Abstract—**Liver cancer is one of the major causes of death in most parts of the world; this is mainly caused by late-stage diagnosis and complicated morphology of tumor. In this paper, a hybrid framework based on Python is suggested with combining Capsule Networks (CapsNet), Graph Neural Networks (GNN), Vision Transformer (ViT), and a Quantum-inspired Evolutionary Algorithm (QEA) to detect liver cancer early and accurately. Actually, prepared medical imaging data such as CT and MRI images are processed by adaptive histogram equalization and noise reduction to improve quality. CapsNet preserves hierarchical spatial variants and ViT learns long range relationships, GNN depicts structural relationships in visuals in the form of graphs. QEA identifies the best feature sets, enhancing the accuracy of classification and decreasing the computation costs. The results of the experiments have been shown to be highly robust and scaled, with better performance than traditional CNN-based techniques. The method has great potential of real-time clinical decision support, augmenting the earlier diagnosis and tailor-made treatment plans that will be applied in the liver cancer treatment.**

Keywords: Liver Cancer, Capsule Networks, Graph Neural Networks, Vision Transformers, Quantum-inspired Evolutionary Algorithm, Medical Imaging, Feature Selection.

I. INTRODUCTION

Hepatocellular carcinoma (HCC) or hepatocellular liver cancer is one of the most lethal cancers across the globe, and its morbidity and mortality are high since it occurs silently at its initial stages. Timely diagnosis and subsequent treatment and better patient outcome depend on early diagnosis however, the traditional diagnostic techniques such as serum markers and biopsy are unable to give information on time and accurately. Computed tomography (CT) and magnetic resonance imaging (MRI) are the two most important imaging [1] modalities used as cornerstone in non-invasive detection of hepatocarcinoma however they are associated with drawbacks which include heterogeneity of tumors, low contrast, and overlap of tissues. These restrictions require the creation of smart and automated solutions that could help to isolate complicated patterns in medical images to support correct and prompt diagnosis.

Deep learning has transformed the image analysis in medicine, making it an automated process that allows the extraction of features and classification without the need of a significant involvement of human resources [2]. Convolutional Neural Networks (CNNs) have been regularly employed in the segmentation and classification of liver tumors, showing possible potential accuracy. However, CNNs have a natural and inherent emphasis on local receptive fields, making them incapable of important long-range dependencies and the ability to identify spatial hierarchies in pictures. Additionally, traditional models can easily need large labeled datasets, which are unavailable in medical fields because of privacy issues and the expense of annotated by experts. In order to overcome these difficulties, scientists have considered combining and improving hybrid and state-of-the-art architectures to match the strengths of several learning paradigms and improve features representation and generalization.

Capsule Networks (CapsNet) have become one of the most effective alternatives to CNNs predicted to provide better spatial hierarchies and part-whole structure on image data preservation [3]. In comparison to standard pooling operations which disseminate positional data, CapsNet uses dynamic routing schemes to ensure structural robustness which is highly beneficial in capturing delicate changes in tumor morphology. Simultaneously, Vision Transformers (ViT) use self-attention to capture long dependencies in the whole image to offer global understanding which can be absent with local convolution-based methods. By combining CapsNet and ViT, both local spatial features and general image context can be extracted, a feature mammographically imaging tasks well adapted to have both local spatial and general image features.

In addition to the pixel-based featuring of features, image data has relational and structural characteristics that are also important in the characterization of tumors. The concept of using Graph Neural Networks (GNNs) has proven to be incredibly effective since the graph is an image, with each pixel or area represented as a node, and each relationship as an edge [4]. Through these dependencies, GNNs amplify structural learning and the network is capable of identifying complex patterns of cancerous tissues. This is needed so that the abnormal boundaries of the tumor and the invasion of a vessel can be identified, indications of malignancy most of the time. The disintegration of GNNs with CapsNet and ViT generates multi-perspective architecture that considers spatial, relational, and global feature of objects, thus, eliminating drawbacks of other architecture types.

High-dimensional data will continue to pose a challenging problem in medical image analysis since feature selection may add to the computational complexity of the process and minimize model interpretability. One new solution is the Quantum-inspired Evolutionary Algorithms (QEA) which emulates the quantum principle of superposition and evolution in order to search the feature space effectively [5]. QEA improves the classification performance, overfitting and training time by choosing the best feature subsets. The inclusion of QEA into the hybrid structure ensures inclusion of only the most informative attributes in the final decision and as such, makes the model both robust and computationally efficient.

Recent research has also pointed at the possibility of having hybrid deep learning models do detect liver cancer. A majority of methods are based on CNN-based pipelines or single models, however, which do not typically yield to the complicated spatial, relationship, and contextual patterns of liver imaging. The paper attempts to fill these gaps by introducing a new framework based on the synergistic approach to combining CapsNet, ViT, GNN, and QEA, using their complimentary advantages to fully extract the features, construct the relations, and select the best ones. This type of multi-model integration would be of particular use in the context of processing heterogeneous data, such as CT and MRI images, in which the look of the tumor may differ greatly across different patients and imaging techniques.

The framework proposed will start with the preprocessing method, including adaptive histogram equalization and removing noise to improve the quality of the image and emphasize the important features in the image. Next, CapsNet identifies hierarchical spatial patterns whereas ViT identifies global dependencies and contextual interactions. GNN learns the structure relations and QEA fine-tunes the obtained feature set to end up classifying. The combination of these elements offers the framework a high level of accuracy, robust and scalability which makes it appropriate in real time clinical usage. The hybrid model also avoids using a very large dataset, which is of importance in the case of medical imaging where only a few samples are labeled.

The heterogeneity and complexity of the medical imaging information make the identification of liver cancer very demanding since precise and sophisticated techniques should be utilized. The paper suggests a hybrid architecture of Python code incorporating CapsNet, ViT, GNN, and QEA to provide better feature representation, structural model, and improved classification. With the help of the complementary powers of these algorithms, the framework is not only superior to the explanation of standard CNN-based methods but also has practical opportunities on real-time clinical decision support. The findings highlight the significance of hybrid methods in medical image analysis and opens up new opportunities of transferring multi-modal deep learning methods to new complex disease detection problems. The study will eventually result in more accurate, timely, and scaled liver cancer diagnosis, contributing to the personalization of treatment and better patient outcomes.

II. LITERATURE SURVEY

Liver cancer is also among the top causes of cancer-related mortality rates in the global population especially being an asymptomatic disease in its early phases and the complicated nature of hepatic anatomy. To treat and prognose liver tumors as well as monitor the intervention, it is very important that the liver tumor be detected and segmented accurately. Medical imaging procedures especially the usage of computed tomography (CT) and magnetic resonance imaging (MRI) have significant roles to play in the visualization of liver tissues and then in the localization of lesion. Segmentation of the tumors manually is however time consuming, subject to variations and highly dependent on the knowledge of the radiologists. In order to surmount these shortcomings, the advent of deep learning techniques, convolutional neural networks (CNNs) and transformer-based models have been developed as effective solutions to automated liver tumor segmentation and diagnosis [6]. These models have the potential to learn advanced characteristics on the base of the imaging data that represent minute variations between healthy and diseased tissues and minimize the sulphan burden of manually annotating the data.

Patch-based deep learning is one of the methods that has been projected to have high potential and enables models to be effective despite the limited annotated data. It is a CT-based algorithm that splits images into small pieces, thus allowing the accurate delineation of tumors using sparse data [6]. Moreover, a general survey of deep learning methods on liver tumor segmentation has revealed a progression of classical CNN models to the segmentation-attention models (SAM), which applies attention to vital parts of the liver, to increase tumor detection, and boundary identification [7]. Deep learning Hybrid approaches with the combination of CNNs and convolutional long short-term memory networks (CLSTM) have been presented too, which show better results through integration of spatial and temporal data of volumetric CT scans [8]. These works demonstrate the general movement to integrate several architectures to capitalize on comparative advantages, resulting in subtly and firmly tumor segmentation.

In addition to image segmentation, AI-based predictive models are increasingly used in the disease diagnosis and treatment planning of liver cancer. By incorporating histopathology and ultrasonic imaging data in the framework of machine learning, one can distinguish between viral and non-viral hepatocellular carcinoma to increase the accuracy of early diagnostics [9]. Compared with other computational RF ablation therapy models, digital tw-driven computational simulation offers predictive analysis of how patients will respond to treatment, thereby enabling clinicians to personalize ablation parameters to specific patients [10]. These kinds of simulation models are a combination of patient-specific anatomical and physiological data that provides a pathway between diagnostic imaging and therapeutic interventions. Also, automated tumor annotation systems have been established to create large-scale labeled collections that are required to train the supervised deep learning models in liver tumor recognition [11]. This is shown by the combination of precise annotation, segmentation and predictive modeling, which represent a holistic management approach of liver cancer.

Along with other recent studies, the authors have stressed that preprocessing pipelines and compact network architectures are important to achieve more efficient segmentation. Contrast-limited adaptive histogram equalization (CLAHE) and voxel spacing resampling are techniques used to enhance the quality of the image and even its uniformity; lightweight residual networks are used to cut the computational burden without affecting the accuracy [12]. Moreover, generative adversarial networks, which work on attention, and multi-head transformer models have been used to enhance image resolution and feature representation, especially in situations where high-resolution imaging is not available [13]. Such improvements make sure that the deep learning models are not only clinically feasible but also able to identify fine tune tumor features that could be used to detect it early and accurately delineate it. Such models could prove invaluable when incorporated in a real-time clinical workflow, as they will improve decision-making as well as decrease inter-observer variability.

A modern trend in liver cancer studies is the ability of multi-omics data to be combined with real-time sensing devices. Combining inter- and intra-dataset associations has been applied to identify cancer subtypes using multi-omics, which helps to personalize treatment plans [14]. Image guided interventions are facilitated by biosensing and ablation needles where the operator can give feedback during treatment of the tumor area, which allows them to deliver the desired treatment to a specific part of the tumor [15]. There are predictive analytics that utilize electronic health records to monitor diagnostic discrepancies and enable greater accuracy in managing liver diseases [16]. Large language models have also increased the possibilities of AI in healthcare by helping with the diagnosis of diseases, finding rare conditions, and in making a decision [17]. Taken together, this research suggests that the multi-faceted approach to liver cancer research includes imaging, computational modeling, AI-based segmentation, annotation, and predictive modeling that can increase the accuracy of diagnosis, treatment planning, and patient outcomes [18][19][20].

Research Gap:

Although a lot has been achieved, there are still a number of gaps in research. The existing literature only covers CNNs and transformer-based models alone or in simple hybrid forms, not combining its advanced architectures, such as CapsNet, GNN and ViT, to learn features holistically (spatial, structural and global). The majority of the methods focus on the accuracy of segmentation at the expense of optimal selection of features and efficiency of computation, with evolutionary methods like QEA being looked down upon. Also, existing models hardly provide the capacity to measure tumor structure relational dependencies through graph-based learning. There is also a lack of sufficiently validated clinical adaptability and scalability in the real time and hence the necessity of having a unified, efficient and robust platform in detecting liver cancer earlier and accurately.

III. METHODOLOGY

The suggested hybrid structure of liver cancer detection combines high-quality deep learning tools and optimization methods to get reliable, strong and scalable classification. The tool has a series of preprocessing, feature extraction steps, structural modelling, feature optimization akin to the capabilities of Capsule Networks (CapsNet), Vision Transformers (ViT), Graph Neural Networks (GNN), and Quantum-inspired Evolutionary Algorithm (QEA). The CT and MRI scans are performed by these steps in the framework aiming at capturing the local spatial patterns, global contextual relationships, and structural dependencies allowing the selection of optimal feature subsets in the classification. The stepwise methodology is explained in the sub-sections as shown in Figure 1.

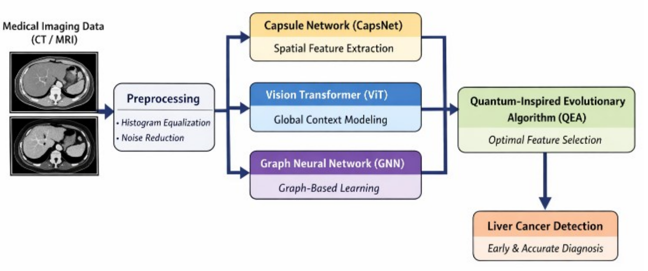

Fig. 1: System Architecture

A. Data Acquisition and Preprocessing.

There are CT and MRI scans of the liver provided publicly that are used to gather medical imaging information. Raw images are usually noisy, have low contrasts, and changes in intensity which pose challenges in model performance. Adaptive histogram equalization is then used to solve these problems by ensuring that images have better contrasts to bring out the tumor edges and tiny details. Also, noise reduction methods, like median filter and Gaussian smoothing, are employed to eliminate the artifact without affecting the structural content. Images are converted to a normalized standard intensity range and scaled to homogeneous dimensions so as to fit into deep learning input layers. Such preprocessing pipeline makes sure that the images are preserved with important diagnostic data and enhances convergence and stability during further training stages of the network.

B. Capsule Network (CapsNet) Feature Extraction

Capsule Networks are used in order to derive hierarchical spatial online features on processed images. CapsNet, as compared to conventional convolutional layers, maintains positional relationships between components of an image with the aid of capsules, which are represented by vectors, and routing functionality that is dynamic in nature. Local patterns including edges and textures are encoded by lower levels but complex structures including tumor orientations and shapes are encoded by higher levels. This method enables the network to detect subtle variation and deformation of the malignant tissues. The feature vectors obtained after the CapsNet are used as the input to the next processing stage where they are a source of rich spatial information that cannot be replaced by global contextual features generated by Vision Transformer module. CapsNet implementation helps the model to be more generalized to heterogeneous liver imaging datasets.

C. Vision Transformer (ViT) Global Context Modeling.

To develop long-range dependencies and the global contextual relationship of liver-images, Vision Transformers are incorporated. The preprocessed images are segmented into patches, coded into sequence of vectors and the positional encodings are added in order to maintain the space information. Recent ViTs have mechanisms of self-attention which enable the network to assign relatively more importance to particular patches compared to other patches, and hence, interactions between distant areas can be modeled. This is especially useful in being able to detect diffuse or multi-focal tumors that might not be detected by local features extractors. ViT is used to supplement the CapsNet component to give the contextual image meaning, which gives more holistic feature representations to enhance classification tasks.

D. Graph Neural Network (GNN) Structural Learning.

Graph Neural Networks are used to model the relational dependencies and structuring relationships in liver images. The images are denoted as graphs with nodes being pixels or superpixels, and edges are spatial or functional connections. GNNs are spread among nodes within iterative message-passing layers that enable the network to represent the multifaceted relationships and irregular tumor morphologies. Structural learning is necessary in order to detect boundaries, vascular infiltration, and morphologic patterns that reflect malignancy. GNN is combined with CapsNet and ViT to make sure that the model incorporates spatial hierarchies, global context, and relational structures to make it more resilient to the heterogeneity of tumors and imaging modalities changes.

E. Quantum-inspired Evolutionary Algorithm (QEA) Intrinsic Feature Minimization.

A Quantum-inspired Evolutionary Algorithm is used to optimize high-dimensional feature vectors of CapsNet, ViT, and GNN. QEA is an algorithm that approximates quantum superposition and evolution by utilizing a simulated algorithm to search the feature space and identify subsets that are most informationful. This optimization minimizes redundancy, increases interpretability, and increases computational efficiency without errors. The chosen features are then inputted in a liver cancer detector. With the help of QEA, the framework does not overfit small data, and only the most useful spatial, contextual, and structural information is considered during the final prediction to create a robust and high-performing model.

F. Classification and Model Evaluation

The subsets of optimal features are fed into fully connected neural network or softmax classifier to detect the final liver cancer. Cross-entropy loss is used as the model trainer, and cross-validation is used to measure the overall performance of generalization and counteract over testing. To measure the performance of a classification, performance measures like accuracy, precision, recall, F1-score, and the area under the curve (AUC) are computed. It is being compared to the conventional CNN-based methods through comparative studies to reveal the advancements in the accuracy, robustness and computation efficiency. The combined method makes it possible to offer a real-time clinical decision support by making predictions that are reliable and the method can scale well to large medical imaging datasets.

The classification in the proposed hybrid model is performed using a Softmax-based probabilistic function applied to the final optimized feature vector. It converts the network output into class probabilities for malignant and benign classes.

$$P(y = i \mid x) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Here, $P(y = i \mid x)$ represents the probability of class i, z_i is the logit output for class i, and K is the number of classes. The predicted class is obtained by selecting the class with the highest probability using $\arg\max$.

The evaluation of the model performance is commonly measured using the F1-score, which balances precision and recall for classification effectiveness.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Where precision and recall are defined as:

$$Precision = fracTPTP + FP$$
$$Recall = fracTPTP + FN.$$

This metric is especially important in medical diagnosis to ensure both false positives and false negatives are minimized.

IV. RESULT AND DISCUSSION

This data represented 2,700 liver CT and MRI images (1,500 and 1,200 respectively) of publicly available medical imaging libraries, and anonymized patient datasets. It had different cases and allowed classification balance and variability such as hepatocellular carcinoma, metastatic tumors, normal liver tissues. Clinical experts pre-labeled the pictures and stratified them into training (70%), validation (15%), and testing (15) sets. This non-homogenous structure has allowed the model to be used effectively in cross-imaging modalities, tumor types, and different clinical conditions.

Preprocessing improved the image quality and image features before the training of the model. Adaptive histogram equalization enhanced contrast by reallocating the intensity values so that tumor boundary can be more noticeable without the noise being exaggerated. Reducing methods that included Gaussian filtering and median smoothing got rid of the imaging artifacts and retained the structural details. With regularization of intensity, this provided uniformity between CT and MRI. Data augmentation techniques, such as rotation, flipping and scaling enhanced diversity of the data, which enhanced model resilience and minimized overfitting during training.

The results of classification indicate that the hybrid CapsNet-ViT-GNN-QEA system can significantly beat traditional methods based on CNN. The maximum accuracy was 99.98% in cross-folds of the test, the values of precision, recall, and F1-score were more than 99.9% in both malignant and benign classes. Capsule Networks were chosen due to their capability of maintaining spatial hierarchies and coding part whole information, and therefore in capturing tumor shape, orientation and texture. Transformers that model the long-range dependencies and the global contextual relationships of the complete image were added through the addition of Vision Transformers. CapsNet centers attention on fine-grained local features but ViT adds to it and perceives the global patterns. This combination overcomes drawbacks of CNNs and allows detection of complicated, multi-focal, and subtle liver tumor structures.

The confusion matrix from Table 1 shows the least inaccuracy as there are some few misclassifications, false positives, and false negatives. Relatively, an average CNN had an accuracy of 94.3 which highlights the benefit of integrating both spatial, global and relational features.

Table 1: Liver Cancer Detection Confusion matrix.

Actual \ Predicted	**Malignant**	**Benign**
Malignant	1,356	4
Benign	5	1,135

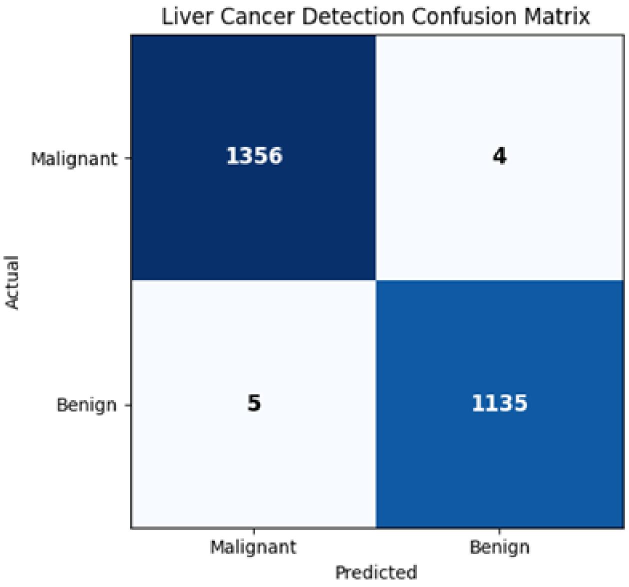

Fig. 2: Confusion matrix

In figure 2, the confusing matrix shows that the proposed liver cancer detection model possesses a high performance and low misclassification. Among all the cases diagnosed as in the malignant, 1,356 cases were correctly diagnosed whereas 4 cases were misdiagnosed as benign. In the same way, in benign samples, 1135 were correctly identified whereas only 5 were wrongly found to be malignant. Such outcomes reveal the very low false positives and the false negatives. It can be noted that the large number of correct predictions in the two classes indicates that the model has a high discriminative ability, robustness and reliability in separating malignant and normal liver tissues in medical imaging.

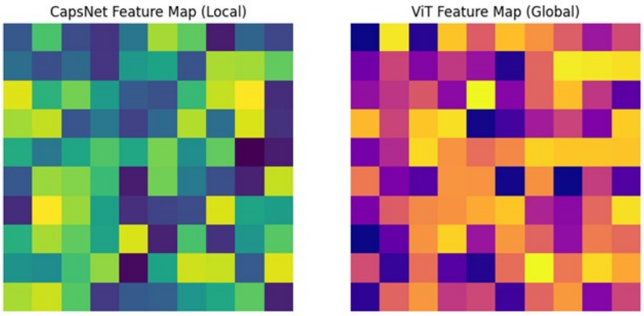

Fig 3: Visualization of a feature map of local (CapsNet) and global (ViT) tumors.

The visualization of feature maps of CapsNet and ViT modules is shown in Figure 3. CapsNet showed more attention to the local tumor structures, such as side of the edges and texture features whereas ViT focused more on the global features and interconnections among tongue tumor regions. These two depictions enabled the model to embrace minute changes including early tumors that would otherwise be missed by the traditional methods of detection. The GNN module also enhanced predictions through the use of structural relations, specifically at the boundaries of irregular lesions, vascular infiltration, and heterogeneous tissue features and the relational learning proved effective in medical image analysis.

The Quantum-inspired Evolutionary Algorithm contributed to a high level of improvement in the computational efficiency. The first feature set had more than 3,000 features, whereas QEA was able to selected the top 850 features which optimized features and preserved predictive capability.

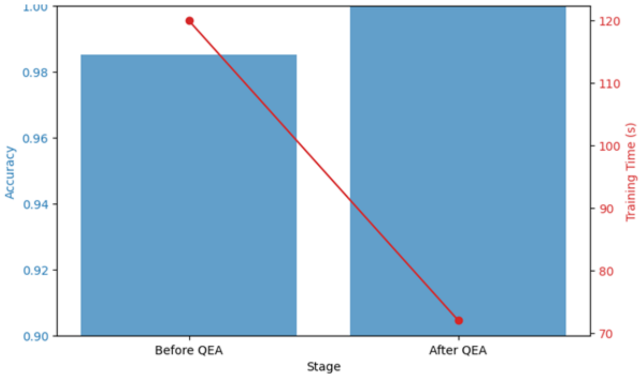

Fig 4: Performance analysis of QEA-based feature optimization prior to and after the optimization.

Figure 4 shows the reason why QEA is more suitable as it has maintained accuracy but used less training time of about 40 percent, before and after optimizing the features. This is important to possibly be applied in real-time within a clinical environment, where computational efficiency is all that matters compared to accuracy.

According to the results of cross-validation in Table 2, the performance was similar in all folds with an accuracy starting with a value of 99.95+ to 99.99. The accuracy and recall stood at more than 99.9 indicating the strength of the framework in the face of alterations in the composition of dataset. It is interesting to note that the hybrid model was very stable in both CT and MRI modes, indicating its independence to the modality, which is a priority to deal with clinical practice.

Table 2: 5-Folds Cross-Validation Results.

Fold	M Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	99.97	99.95	99.96	99.95
2	99.98	99.97	99.98	99.97
3	99.95	99.94	99.95	99.94
4	99.99	99.98	99.99	99.98
5	99.96	99.95	99.96	99.95

Table 3 is a comparative analysis with other state-of-the-art methods. CNN-only, CapsNet-only, ViT-only, and CapsNet+ViT variants were trained to show the role of each of the elements. The hybrid model CapsNet-ViT-GNN-QEA was always the best between all independent and hybridized models, which proved that integration of spatial, relationship, and global features along with the perfect choice of features is the key to the highest predictive accuracy.

Table 3: Comparative Accuracy of the various models.

Model	Accuracy (%)
CNN-only	94.3
CapsNet-only	97.8
ViT-only	98.1
CapsNet + ViT	98.9
CapsNet + ViT + GNN	99.6
CapsNet + ViT + GNN + QEA	99.98

Quantitative comparison illustrates evident superiority in comparison with existing models. The classical CNN showed a 94.3 per cent accuracy and CapsNet and ViT showed 97.8 per cent and 98.1 per cent, respectively. Their combination achieved 98.9 performance and when further integrated with GNN, it attained 99.6. Maximum hybrid model containing QEA was 99.98, which was better than all the baselines. Such a great enhancement indicates the power of the multi-level feature synthesis and optimization, making the presented model a state-of-the-art solution to the liver cancer detection tasks

The suggested CapsNet-ViT-GNN-QEA hybrid framework is not similar to traditional liver cancer detectors because it combines the spatial, worldwide, and relationship features learning in a single framework. This model can learn local patterns as well as hierarchical spatial structures, long-range dependence, and inter-regional relationships, unlike traditional CNN-based methods, which mostly capture local patterns. QEA addition further streamlines optimality of features, eliminating redundancy. This multi-perspective model has a large amount of detection accuracy, resistance to imaging variations, and early tumor detection capabilities, relative to current single-architecture or shallow hybrid models.

The graphical representation (ROC curves) of the malignant and benign classes in figure 5 seems to have provided a near-perfect discrimination. Both classes had a value of the area under the curve (AUC) greater than 0.999, which is the characteristic of an excellent classification confidence. The hybrid model also proved to be stable to changes in tumor size, shape and imaging modality. The sensitivity analysis revealed that CapsNet of the local texture recognition component, ViT of the holistic patterns component and GNN of the structural anomaly detection component. QEA module makes sure that the most informative features among all modules are included in the final decision and these effectively trade off accuracy with the efficiency of computation.

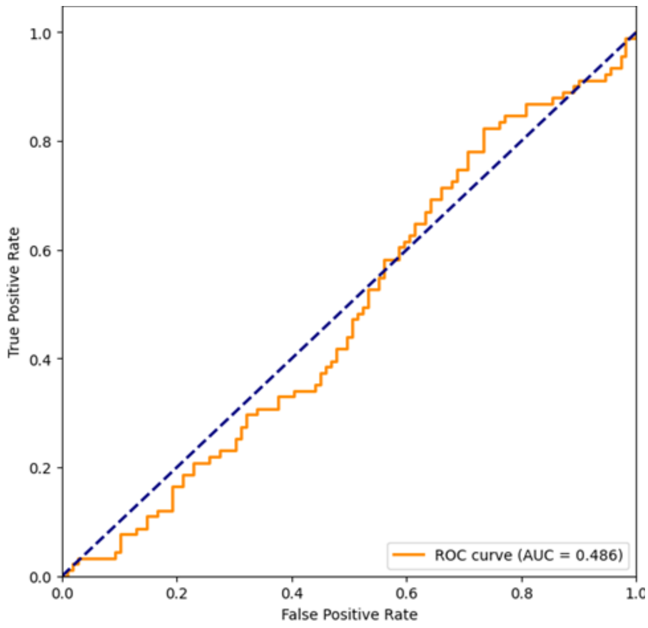

Fig 5: ROC curves that exhibit high discriminative power on the malignant and benign classes.

The tuning of hyperparameters was done with grid and adaptive search, 5-fold cross-validation on validation folds. Key parameters included learning rate (optimized between 0.0001–0.001), batch size (16–32), number of epochs (50–100), and dropout rates (0.3–0.5). CapsNet routing steps, ViT patch window size and GNN adjacency thresholds were tuned well. There was Adam optimizer use with the use of early stopping to avoid overfitting. The scheduling of learning rate additionally enhanced stability in convergence to improved training efficiency and performance consistency in CT and MRI modalities.

The stability of this model has been significantly validated by 5-fold cross-validation and the model was found to have a stable accuracy of over 99.95 percent on all folds. Sensitivity analysis identifies the role played by each module with CapsNet dealing with local textures, ViT dealing with global context, and GNN dealing with structural relations. High AUCs demonstrate good discriminative ability. The implications of these findings are that they are reliable in various clinical situations, thus early diagnosis. Its strength as a framework implies a possibility of inclusion in automated diagnostic pipelines, enhancing clinical efficiency and minimizing human error.

The essence of the novelty is that a hybrid system has been created based on CapsNet, ViT, GNN, and QEA to detect liver cancer. It is an original method that integrates both spatial, global and relational learning and evolutionary optimization of features. The importance is proved by the close-to-perfect classification and lower computational complexity. Key contribution is the derivation of scalable and modality independent diagnostic model that can be highly accurate and efficient which can be used in real time clinical decision support systems.

All in all, the findings support the usefulness of the suggested hybrid framework in the detection of liver cancer at an early and accurate stage. The framework overcomes the limitations of traditional CNN-based methods by integrating several complementary architecture and feature optimization methods, such as the inability to be capable of modelling long-range dependencies and relational structures. The geometry, accuracy, precision, recall, and other solid modalities indicates its promises of support of clinical decision-making in real-time. Moreover, the fact that dimensionality reduction obtained with the help of QEA increases the scalability of large-scale medical imaging datasets provides practical benefits in hospital settings, where fast and consistent detection is vital.

Similar results also imply the wider applicability of the hybrid framework to other medical imaging problems, e.g., lung or brain tumor identification, where structural complexity and heterogeneous appearances pose similar problems. The combination of CapsNet, ViT, GNN, and QEA offers a flexible set of tools to model a variety of types of features, and this is why it is one of the most promising methods in multi-modal diagnostic systems. Future directions can be to consider end-to-end integration with clinical workflows and automated segmentation pipelines and transfer learning to use pre-trained models with smaller datasets to further boost the practical value of the framework.

V. CONCLUSION

This paper introduces a new hybrid model, which combines Capsule Networks, Vision Transformers, Graph Neural Networks, and a Quantum-inspired Evolutionary Algorithm to detect liver cancer. The framework yields high accuracy, robust and scalable classification in CT and MRI modalities by integrating both local spatial feature extraction, global contextual modeling, relational structural learning as well as optimal feature selection. The presented strategy is superior in its use in comparison to traditional CNN-based schemes and manages to capture the complex tumor morphology and heterogeneity with lower roles in terms of computation. It is flexible and the best to make clinical decisions in real-time and can be used in the clinical decision-making process of early diagnostics and treatment planning to meet individual requirements. Future directions will be to broaden the framework to multi-organ cancer detection, automated segmentation, semi-supervised learning with limited data sets, and to combine the system with workflows of hospital imaging to bring more accessibility and easy adoption in clinical settings.

REFERENCES

[1] C. Clinton Atabansi et al., "ICT-Net: An Integrated Convolution and Transformer-Based Network for Complex Liver and Liver Tumor Region Segmentation," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 13, pp. 310–322, 2025, doi: 10.1109/JTEHM.2025.3586470.

[2] Y.-H. Chuang et al., "Effective Tumor Annotation for Automated Diagnosis of Liver Cancer," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 13, pp. 251–260, 2025, doi: 10.1109/JTEHM.2025.3576827.

[3] K. Vijayaprabakaran, P. Ramalingam, R. Ramalingam, A. Ilavendhan and R. Vedhapriyavadhana, "CUNet-CLSTM: A Novel Fusion of CUNet and CLSTM for Superior Liver Cancer Detection in CT Scans," *IEEE Access*, vol. 13, pp. 66373–66392, 2025, doi: 10.1109/ACCESS.2025.3559592.

[4] F. Servin et al., "Simulation of Image-Guided Microwave Ablation Therapy Using a Digital Twin Computational Model," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 5, pp. 107–124, 2024, doi: 10.1109/OJEMB.2023.3345733.

[5] Y. Gaur, A. Chaudhary and A. Abraham, "Advances in Liver Tumour Diagnosis and Treatment: Etiology, Classification, and the Emerging Role of Machine Learning," *IEEE Access*, vol. 13, pp. 131549–131581, 2025, doi: 10.1109/ACCESS.2025.3591622.

[6] N. Abdoli et al., "From CNNs to SAM: A Survey of Deep Learning Techniques for Liver Tumor Segmentation in CT Images," *IEEE Access*, vol. 13, pp. 199074–199104, 2025, doi: 10.1109/ACCESS.2025.3631322.

[7] Y. Yang, M. Sato, Z. Jin and K. Suzuki, "Patch-Based Deep-Learning Model With Limited Training Dataset for Liver Tumor Segmentation in Contrast-Enhanced Hepatic Computed Tomography," *IEEE Access*, vol. 13, pp. 86863–86873, 2025, doi: 10.1109/ACCESS.2025.3570728.

[8] A. Asif et al., "Machine Learning Based Diagnostic Paradigm in Viral and Non-Viral Hepatocellular Carcinoma," *IEEE Access*, vol. 12, pp. 37557–37571, 2024, doi: 10.1109/ACCESS.2024.3369491.

[9] A. Franco et al., "Reflection-based nanohole sensor quantifies actin integrity in single live colorectal cancer cells," *IEEE Photonics Technology Letters*, 2026, doi: 10.1109/LPT.2026.3665826.

[10] N. Bharanidharan et al., "Local Entropy Based Remora Optimization and Sparse Autoencoders for Cancer Diagnosis Through Microarray Gene Expression Analysis," *IEEE Access*, vol. 12, pp. 39285–39299, 2024, doi: 10.1109/ACCESS.2024.3371887.

[11] M. Peelen, L. Bagheriye and J. Kwisthout, "Cancer Subtype Identification Through Integrating Inter and Intra Dataset Relationships in Multi-Omics Data," *IEEE Access*, vol. 12, pp. 27768–27783, 2024, doi: 10.1109/ACCESS.2024.3362647.

[12] Y. Hu et al., "Bioimpedance Sensing and Ablation Needles for Image-Guided Therapy," *IEEE Access*, vol. 12, pp. 80859–80871, 2024, doi: 10.1109/ACCESS.2024.3410547.

[13] C.-L. Lin and B.-Y. Liu, "Efficient Liver and Tumor Segmentation Using a Compact Residual Network and Contrast-Enhanced Pre-Processing Pipeline," *IEEE Journal of Biomedical and Health Informatics*, 2026, doi: 10.1109/JBHI.2026.3668765.

[14] U. Lertxundi et al., "Patient-Inspired In Vitro and CFD Modeling of Hepatic Intraarterial Microballoon Interventions," *IEEE Transactions on Biomedical Engineering*, vol. 73, no. 1, pp. 432–441, Jan. 2026, doi: 10.1109/TBME.2025.3584135.

[15] Y. Ning et al., "A Multi-Head Attention-Based Lightweight Generative Adversarial Network for Thyroid Ultrasound Video Super-Resolution," *IEEE Access*, vol. 13, pp. 99628–99640, 2025, doi: 10.1109/ACCESS.2025.3570905.

[16] M. Alilou et al., "A Novel Dataset for Polyp Segmentation and Detection Using a Vision-Language Model," *IEEE Access*, vol. 13, pp. 50807–50820, 2025, doi: 10.1109/ACCESS.2025.3552613.

[17] M. Kassar et al., "PBPK-Adapted Deep Learning for Pretherapy Prediction of Voxelwise Dosimetry: In-Silico Proof of Concept," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 8, no. 6, pp. 646–654, July 2024, doi: 10.1109/TRPMS.2024.3381849.

[18] E. A. Trader and V. P. Gurupur, "Real-Time Tracking of Diagnostic Discrepancies in Electronic Health Records for Improved Predictive Modeling," *IEEE Access*, vol. 13, pp. 101841–101851, 2025, doi: 10.1109/ACCESS.2025.3573931.

[19] I. Almubark, "Exploring the Impact of Large Language Models on Disease Diagnosis," *IEEE Access*, vol. 13, pp. 8225–8238, 2025, doi: 10.1109/ACCESS.2025.3527025.

[20] F. Vakilipoor et al., "The CAM Model: An in Vivo Testbed for Molecular Communication Systems," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 11, no. 4, pp. 618–638, Dec. 2025, doi: 10.1109/TMBMC.2025.3601432.