Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 5, 1471-1486
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i5.7180
© 2025 by the authors; licensee Learning Gate

Vision transformer-based feature selection for liver tumor MRI analysis

Shilpa B[1*], Glan Devadhas G[2], T Y Satheesha[3]
[1]Dept. of ECE, SOET, CMR University, Bengaluru, India, 560043, Dept. of ECE, BIT, Bengaluru, India, 560004; shilpaprashanth2014@gmail.com (S.B.).
[2]SOET &Director-DORI, CMR University, Bengaluru, India, 560043; drglan.d@cmr.edu.in (G.D.G.).
[3]School of Computer Science Engineering, REVA UNIVERSITY, Bengaluru, Sathanur, India, 560064; ty.satish@gmail.com (T.Y.S.).

Abstract: Liver tumors, classified as benign or malignant, pose significant diagnostic challenges. In infancy, benign liver tumors may progress to malignancy, making early and accurate classification crucial. Traditional manual classification methods are inefficient, time-consuming, and prone to errors, necessitating advanced automated techniques. This study introduces a novel Vision Transformer with Learned Invariant Feature Transform-based statistical features (ViT+LIFT based Stat features) approach for liver tumor classification. Magnetic Resonance Imaging (MRI) liver tumor images from the ATLAS dataset serve as input. The preprocessing stage employs an Adaptive Wiener Filter (AWF) to enhance image quality. A Dynamic Context Encoder Network (DCE-Net) is then utilized to segment the liver and lesions. Feature extraction incorporates Shape Index Histogram (SIH), shape features, ResNet features, and LIFT with statistical features. Finally, the Vision Transformer (ViT) classifies liver tumors based on these extracted features. The proposed ViT+LIFT based Stat features model achieved superior classification performance, with an accuracy of 91.732%, sensitivity of 90.118%, and specificity of 90.710%. These results demonstrate the effectiveness of the proposed method in improving liver tumor classification accuracy, reducing diagnostic delays, and minimizing the need for invasive biopsies.

Keywords: *Adaptive wiener filter, Classification of liver tumor, Dynamic context encoder network, Magnetic resonance imaging, Vision transformer.*

1. Introduction

The image processing methods have become progressively significant in diverse applications owing to an emergence of computer technologies. It is specifically accurate for clinical imaging namely ultrasonography, MRI, nuclear medicine and Computed Tomography (CT) that can be employed for assisting doctors in research, diagnosing and treatment [1]. Liver is an organ accountable for blood purification as it is exposed to the impureness more than other organs. When a liver is affected with the diseases such as tumor, it is enormously significant to be diagnosed for malignancies at earlier. The diagnosing approaches mostly depend upon imaging methods, such as CT and MRI scans, exploiting Dynamic Contrast-Enhanced schemes. When compared to CT, dynamic contrast-enhanced MRI (DCEMRI) technique provides high specificity as well as sensitivity in diagnosing liver tumor, owing to its finest tissue contrasts and comprehensive blood supply classification. With the MRI, diverse contrast processes can be employed for increasing difference of several tissues, for example; T1-, T2-, and diffusion-weighted sequences. When the T2-weighted MRI with a fat saturation is not accessible, standard T2-weighted MRI is employed, similar to medical practices [2]. Developing and designing computer-aided image processing methods to assist doctors enhance their diagnosing has gained significant interest over last few years [1].

History: Received: 26 February 2025; Revised: 24 April 2025; Accepted: 28 April 2025; Published: 14 May 2025
* Correspondence: shilpaprashanth2014@gmail.com

Liver is an essential organ located in upper abdomen that supports digestion and eliminates waste products from blood. The tumor is classified into two types such as malignant and benign [3]. Liver having malignant tumor is referred to liver tumor [4]. Therefore, it is crucial to identify the type of tumor. For determining tumor type, liver should be segmented [4]. The segmentation is a complicated process owing to varying sizes of liver cancers as well as their shapes. Hence, an automatic procedure for segmentation of liver tumor assists specialists with exact and earlier diagnosing of liver cancers. The liver segmentation regarding CT images is a challengeable task owing to occurrence of same intensity objects in abdomen with no obvious delineation amongst liver and objects [3, 5, 6]. The classification of tumor is accomplished by Neural Network (NN) classifier. Various researchers employed techniques, like Deep Learning (DL), fuzzy-enabled approaches and Machine Learning (ML) for diagnosing liver tumor. DL-enabled models have been made machine vision tasks highly progressive and adaptable to support in clinical diagnosing [7]. Some of the examples of DL techniques are Deep Belief Network (DBN), Stacked Autoencoders (SAE) and Convolutional Neural Network (CNN). An accuracy of DL-enabled techniques is considerably higher than conventional ML-enabled approaches [8-14].

The research aims to design ViT+LIFT based Stat features for liver tumor classification using MRI liver tumor images from the ATLAS dataset [15]. AWF is used for pre-processing, followed by liver area and lesion segmentation using DCE-Net. SIH, shape, ResNet, and LIFT features are extracted, along with statistical features like mean, contrast, entropy, energy, variance, and homogeneity. ViT is then used for liver tumor classification.

1.1. Motivation

MRI is an imaging methodology for detection and classification of liver tumor. The conventional approaches utilized for classification are time-consumable and required skilled experts for analyzing tumors. Thus, an automated and incorporated approaches are necessary to classify the types of liver tumor. Motivated by this fact, liver tumor classification model is developed by reviewing traditional approaches. This section presents the existing liver tumor classification methods and their demerits.

The proposed ViT+LIFT-based statistical feature framework for liver tumor classification is designed to achieve the following objectives:

- Acquire MRI liver tumor images from the ATLAS dataset and enhance them using the Adaptive Wiener Filter (AWF) for noise reduction.
- Perform liver area and lesion segmentation using the Dynamic Context Encoder Network (DCE-Net).
- Extract features including Shape Index Histogram (SIH), shape features (circularity, irregularity, area, and perimeter), ResNet features, and Learned Invariant Feature Transform (LIFT) with statistical features (mean, contrast, entropy, energy, variance, and homogeneity).
- Utilize Vision Transformer (ViT) for liver tumor classification, integrating the extracted features for improved accuracy.

The layout of other sections is: Section 2 interprets literature overview of liver tumour classification methods and their disadvantages, section 3 describes ViT+LIFT based Stat features methodology, section 4 illustrates outcomes of ViT+LIFT based Stat features and section 5 mentions conclusion of ViT+LIFT based Stat features.

2. Literature Review

ResNet-10, designed for primary liver tumor classification [16] has been criticized for its complexity and lack of demonstrations. Other studies, such as Zheng's 3D convolution [17] and Patel's PocketNet+nnUNet [8] have shown high performance but lack multi-center external testing sets. Wang's UNet++ [18] improved detection of tumor margins and invasion but was expensive and impacted healthcare costs and accessibility.
Recent advancements in deep learning have improved liver tumor segmentation precision [19].

Researchers have combined an attention-augmented U-Net with a conditional GAN to produce synthetic annotated datasets, achieving a Dice score of 88.60% for liver MRI. RFiLM-Net [20] a two-stage U-Net-based model that integrates radiomic features for CT-based tumor segmentation, achieved a Dice similarity coefficient of 0.87, ensuring high segmentation accuracy and clinical relevance.

Hybrid deep learning frameworks have also been explored, with ResUNet [21] and Inception v4 achieving a Dice score of 98.86% for CT-based liver tumor segmentation and a transformer-based segmentation model [22] reaching 97.28% Dice accuracy on the LiverHCC dataset. These studies demonstrate that deep learning-driven automated segmentation enhances diagnostic efficiency and patient outcomes, underscoring the need for further integration of transformers and hybrid architectures in medical imaging.

2.1. Challenges

The challenges experienced by traditional techniques are interpreted as follows.

- The method in Goedhart [16] focuses on improving MRI image computable features extraction but lacks multiphasic T1-weighted and T2-weighted MRI information for accurate liver tumor classification.
- The method successfully achieved tumor segmentation [17] but faced co-registration errors due to motion artifacts.
- Class imbalance in Deep Learning techniques hinders clinical diagnoses like liver cancer classification, affecting model performance and generalization. DL-based feature extraction methods are needed.

3. Proposed Vit+Lift Based Stat Features for Liver Tumor Classification

Liver cancer is one amongst common cancer over a globe and thus, automated classification techniques are important to assist doctors in tumor diagnosing process. Here, ViT+LIFT based Stat features is presented for liver tumor classification. Firstly, input MRI liver tumor image is obtained from the ATLAS dataset. Next, considered MRI liver tumor image is pre-processed by AWF. Thereafter, liver area segmentation and lesion segmentation are conducted utilizing DCE-Net. Afterwards, features such as SIH, shape features, ResNet features and LIFT with statistical features are extracted. At last, classification of liver tumor is performed employing ViT. Figure 2 reveals the pictorial presentation of ViT+LIFT based Stat features for liver tumor classification.

The study chose Vision Transformer (ViT) for liver tumor classification over traditional CNNs or hybrid methods due to its unique advantages in handling complex medical imaging tasks like MRI scans. ViT processes images by dividing them into patches, capturing global dependencies and long-range relationships, which is crucial in medical images where tumor structures may span large portions of the scan. It has demonstrated superior performance in medical imaging tasks, with better generalization and classification accuracy compared to CNNs.

3.1. Input MRI Liver Tumor Image Acquisition

An input MRI liver tumor image is acquired from the ATLAS dataset [15] which can be formulated by,

$$M = \{M_1, M_2, \dots, M_l, \dots, M_d\} \qquad (1)$$

Where, M_l represents l^{th} input MRI liver tumor image whereas M_d indicates overall MRI liver tumor images in dataset M. ATLAS dataset [15] is categorized into two sets like training set and testing set. A training set as well as testing set comprises the information of 60 patients and 30 patients in about 90 formats.

3.2. Pre-Processing Utilizing AWF

Pre-processing is essential to enhance quality of images and ensures that extracted features are appropriate for liver tumor classification. Here, AWF is utilized for pre-processing an input MRI liver tumor image M_l. AWF [23] modifies an output of filter in accordance to local variances of image. A significant goal of this filter is to lessen Mean Squared Error (MSE) amongst actual image and restored image. The pre-processed outcome of this technique is helpful to preserve the edges of an image. The beneath expression is utilized for processing the pixels to obtain final results.

$$W_l = \varsigma + (1 - v + \Delta) * (z(n, w) - \varsigma) \tag{2}$$

$$v = \frac{\sigma_\alpha}{\sigma_v + 1} \tag{3}$$

$$\Delta = \frac{\sigma_v}{\sigma_\alpha + \sigma_\mu + 1} \tag{4}$$

3.3. Liver Area Segmentation using DCE-Net

Liver area segmentation is a process to identify and segment liver areas from adjacent anatomical structures in clinical imaging like MRI. Here, DCE-Net is employed for segmenting liver area by considering pre-processed MRI liver tumor image $W_l$as an input. DCE-Net [24] adopts a structure of U-Net, comprising of decoder and encoder. DCE-Net is incorporated with newer elements such as involution layer, Context Extraction Module (CEM), Channel Attention Gate (CAG) and Dynamic Residual Module (DRM).

3.3.1. Involution Layer

An initial convolutional (conv) layer is replaced with involution layer for processing an input image and enhancing aggregation ability of parameter on a channel. An involution layer can uncouple the information communications considerately for balancing efficiency and accuracy.

3. 3.2. DRM

The conv is replaced with a dynamic convolution as well as residual blocks for enhancing feature extraction ability. DRM with shortcut process can lessen gradient vanishing caused by enhancing depth of network. Furthermore, it increases converging speed and decreases computational costs.

3. 3.3. CEM

CEM is included at a bottleneck of DCE-Net for compensating the spatial information loss after numerous down-sampling layers. Moreover, it comprises blocks such as Residual Multi-kernel Pooling (RMP) and Dense Atrous Convolution (DAC). CEM is capable to incorporate and extract context semantic information for generating higher-level feature maps. A receptive field's size in an individual branch is computed as follows.

$$\boldsymbol{r_i = r_{i-1} + (f_i - 1) \prod_{h=1}^{i-1} x_h} \tag{5}$$

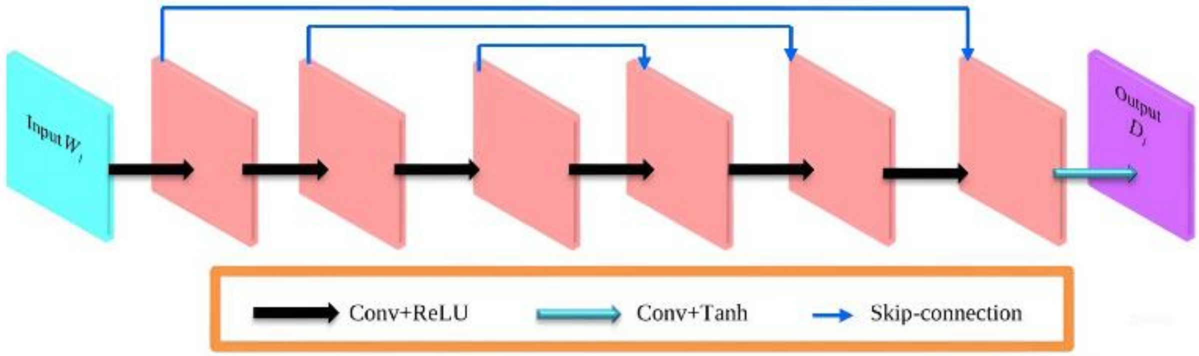


Figure 1.
Structure of DCE-Net.

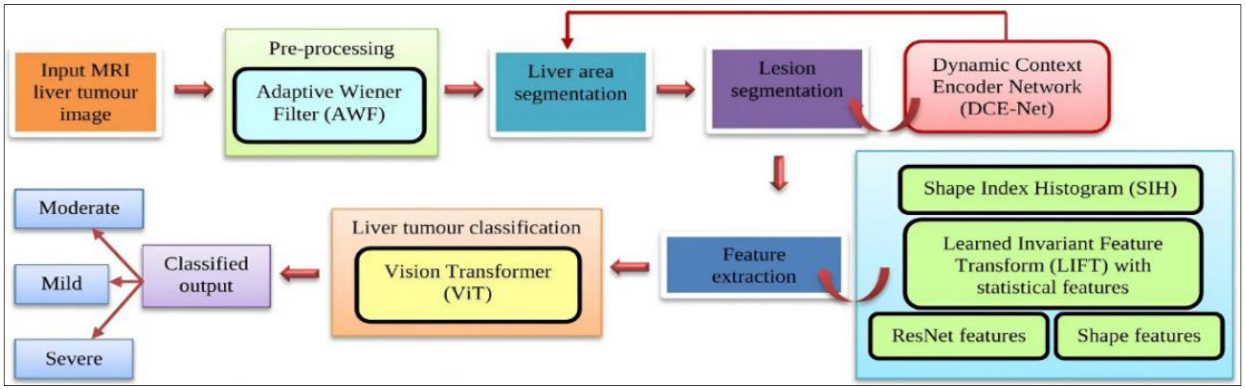

Input MRI liver tumour image
Pre-processing
Adaptive Wiener Filter (AWF)
Liver area segmentation
Lesion segmentation
Dynamic Context Encoder Network (DCE-Net)
Shape Index Histogram (SIH)
Learned Invariant Feature Transform (LIFT) with statistical features
ResNet features
Shape features
Feature extraction
Liver tumour classification
Vision Transformer (ViT)
Classified output
Moderate
Mild
Severe

Figure 2.
Pictorial presentation of ViT+LIFT based Stat features for liver tumor classification.

3.3.4. CAG

CAG is fixed at routes of skip connections in up-sampling layers of decoder for improving an extraction accuracy of contextual features. This element is introduced for suppressing inappropriate areas while focusing on salient features in feature map. In addition, CAG can be effortlessly incorporated into U-shaped CNN structure with minimum computational overhead while efficiently enhancing power and system's sensitivity. The liver area segmented image is specified as D_l and figure 1 delineates structure of DCE-Net.

3.4. Lesion Segmentation using DCE-Net

Lesion segmentation is a process to identify and delineate abnormal tissue areas within MRI liver images. The purpose of lesion segmentation is to accurately outline lesion or tumor boundaries for differentiating them from adjacent healthier liver tissues. Here, lesion segmentation is performed using DCE-Net by considering D_l as an input. The architecture of DCE-Net is already explained in section III.A.1 and a lesion segmented image is specified as C_l.

3.5. Feature Extraction based on Liver Area Segmented Image and Lesion Segmented Image

Feature extraction is a process to extract significant or appropriate features that can be efficiently classify several types of liver tumors. Here, feature extraction is accomplished based on liver area segmented image D_l and lesion segmented image C_l. The features considered for extraction are SIH, shape features, ResNet features and LIFT with statistical features.

3.5.1. SIH

An input C_l is applied with SIH to acquire texture image. SIH [25] is the second-order curvature measure modeled form eigen values of local image. It is developed by selecting the group of τ_u bin centers $\beta_{u1}, \dots, \beta_{u\tau}$ averagely distributed over shape index interval of $[-\pi/2, \pi/2]$. A shape bin contribution G_u at position a [26] can be illustrated as,

$$T_l = G_u(a; \sigma, \beta_u, b_u) = \exp\left(-\frac{(\beta_u - u)^2}{2b_u^2}\right) \tag{6}$$

3.5.2. Designed LIFT with Statistical Features

The model for liver tumor classification uses statistical features and LIFT designed in Figure 3, selected based on empirical studies and medical imaging research. Key features like mean, entropy, energy, and contrast capture texture-based characteristics. Automated methods like correlation analysis and feature importance techniques reinforce feature selection, retaining relevant features for training. This balances theoretical justification with automated methods for predictive performance. Future studies may expand the selection process with diverse methods.

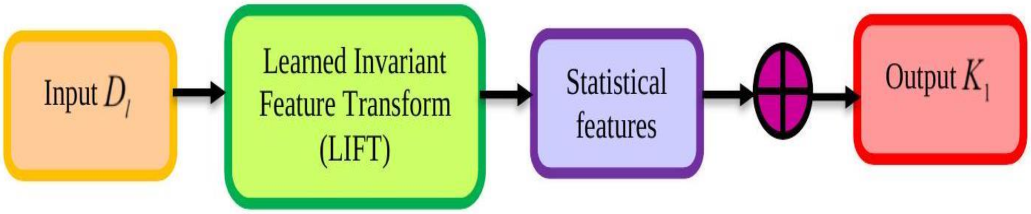


Figure 3.
Designed LIFT with statistical features.

3.5.3. LIFT

LIFT [27] is constructed based on conventional concepts of feature extraction by improving them with learned elements for achieving superior performance under diverse conditions. It contains three units namely descriptor, orientation estimator and detector. The LIFT feature can be depicted as L_l.

3.5.4. Statistical Features

The considered statistical features for extraction are mean, entropy, variance, energy, homogeneity and contrast [28] are applied to L_l.

(i) Mean: Mean refers to an average of pixel-wise brightness in a region, which is given by,

$$N_1 = \sum_{k=0}^{Z-1} kP(k) \tag{7}$$

(ii) Entropy: Entropy acts as the gauge for overall randomness of an image and it is computed as,

$$N_2 = \sum_{k=0}^{Z-1} P(k)\log_2 [P(k)] \tag{8}$$

(iii) Variance: Variance measures the deviation of image intensity from mean that is modelled as,

$$N_3 = \sigma^2 = \sum_{k=0}^{Z-1} (k - N_1)^2 P(k) \tag{9}$$

(iv) Energy: Energy is a measure of homogeneousness of an image, which can be evaluated by,

$$N_4 = \sum_{k,c} P(k,c)^2 \tag{10}$$

(v) Homogeneity: Homogeneity illustrates the closeness of matrix's component distribution resembling its diagonal matrix. The expression of homogeneity can be formulated as,

$$N_5 = \sum_{k,c} \frac{P(k,c)}{1+|k-c|} \tag{11}$$

(vi) Contrast: Contrast specifies to a variation among the brightness of the region or object and their neighbourhood regions or objected in similar view. The equation of contrast is given as,

$$N_6 = \sum_{k,c} |k-c|^2 P(k,c) \tag{12}$$

$$K_1 = \{N_1, N_2, \ldots, N_6\} \tag{13}$$

- K_1 = Feature vector after applying statistical features
- $N_1, N_2, \ldots, N_6$ = Statistical features like mean, entropy, variance, energy, homogeneity, and contrast.

1) *Shape features*

The lesion segmented image C_l is applied with shape features namely circularity, irregularity, area and perimeter [29] for obtaining feature vector.

(i) Area: Area specifies to the overall pixels in an image and it is manifested byQ_1.
(ii) Perimeter: Perimeter represents a distance across the boundary of an area, which is signified as Q_2.
(iii) Circularity: Circularity is specified as a measure for identifying roundness of an object, which can be represented as,

$$Q_3 = (Q_2^\wedge 2)/(4 * \pi * Q_1) \tag{14}$$

Here,

- Q_1 = Area; Q_2 = Perimeter;Q_3 = Circularity

(iv) **Irregularity:** Irregularity captures the non-standard or unusual aspects of an image and it is depicted as Q_4. After applying shape features to C_l, feature vector K_2 is obtained, such that,

$$K_2 = \{Q_1, Q_2, Q_3, Q_4\} \tag{15}$$

Here,

- K_2= Feature vector after applying shape features
- Q_1= Area; Q_2= Perimeter; Q_3= Circularity; Q_4= Irregularity

2) *ResNet features*

In addition, ResNet features are applied over liver area segmented image D_l for acquiring feature vector. ResNet [30] employs conv layer for feature extraction and also, this network consists of fully-connected (FC) classifiers for allocating labels of input image based on extracted features. The feature vector obtained after applying ResNet features is implied by a term K_3. An overall feature vector acquired from feature extraction stage is manifested as K_l, such that,

$$K_l = \{K_1, K_2, K_3\} \tag{16}$$

3.6. Liver Tumor Classification using ViT

Liver tumor classification is important for diagnosing and treatment planning of tumor. The understanding of tumor types and its attributes can assist to overcome the factors affecting patient's outcomes. Here, ViT is utilized for performing liver tumor classification into mild, severe and moderate by taking X_l as an input, such that,

$$X_l = \{T_l, K_l\} \tag{17}$$

Here,

T_l = Textural image, K_l = Feature vector.

3.7. Architecture of ViT

ViT [31, 32] is a kind of NN developed for image classification tasks that exploits transformer model. The original transformer acquires input as one-dimensional (1D) series of token embeddings. For handling two-dimensional (2D) images, an image $g \in \Re^{\mathrm{E}\times\mathrm{H}\times\mathrm{B}}$ is reshaped into series of flattened 2D patches $g_y \in \Re^{O\times Y^2 \cdot B}$, wherein (B,H) is a resolution of actual image, B denotes total channels, (Y, Y) depicts a resolution of individual image patch whereas $O = \mathrm{EH}/Y^2$ refers to a resultant count of patches that also acts as an effectual input series length for transformer. This transformer employs consistent latent vector dimension R throughout its layers and thus, patches are flattened and mapped to R sizes with learnable linearity projection as mentioned in Equation (18). An outcome of this linearity projection is specified as patch embeddings.

The trainable embedding is prepended to series of embedding patches $(t_0^0 = g_{\text{class}})$, whose condition at an output of transformer encoder (t_s^0) behaves as image depiction q. During fine-training and pre-training, the classification head is attached to t_s^0. A classification head is executed by multilayer perceptron (MLP) with single hidden layer at a pre-training time and by one linearity layer at fine-training time. The location embeddings are included to patch embeddings for retaining locational information. Therefore, resultant series of embedding vector acta as an input to encoder.

A transformer encoder comprises alternative layers of the multiheaded self-attention (MSA) as well as MLP blocks. The Layernorm (LN) is employed before each block and residual links after individual block. MLP consists of two layers with GELU non-linear function. The expressions of this structure are formulated as follows.

$$t_0 = \left[g_{\text{class}}\,; g_y^1 A; g_y^2 A; \ldots ; g_y^O A\right] + A_{pos}, A \in \Re^{(\mathrm{Y}^2 \cdot \mathrm{B})\times R}, A_{pos} \in \Re^{(O+1)\times R} \quad (18)$$

$$t'_{\,s} = \mathrm{MSA}\big(\mathrm{LN}(t_{s-1})\big) + t_{s-1},\ s = 1, \ldots, S \quad (19)$$

$$t_s = \mathrm{MLP}\big(LN(t'_s)\big) + t'_s,\ s = 1, \ldots, S \quad (20)$$

$$q = LN(t_s^0) \quad (21)$$

The overall computational complexity is:

$$O\big(mn(Lk^2 + L_r k_r^2 + d)\big) + O\left(\frac{m^2 n^2}{P^4} d\right) \quad (22)$$

This indicates that the Vision Transformer's self-attention mechanism significantly impacts complexity, particularly for large image resolutions.

4. Results and Discussions

The liver tumour classified output is symbolized as V_l and structure of ViT is represented in Figure 4.

4.1. Experiment Setup

ViT+LIFT based Stat features designed for classification of liver tumor is implemented in PYTHON tool.

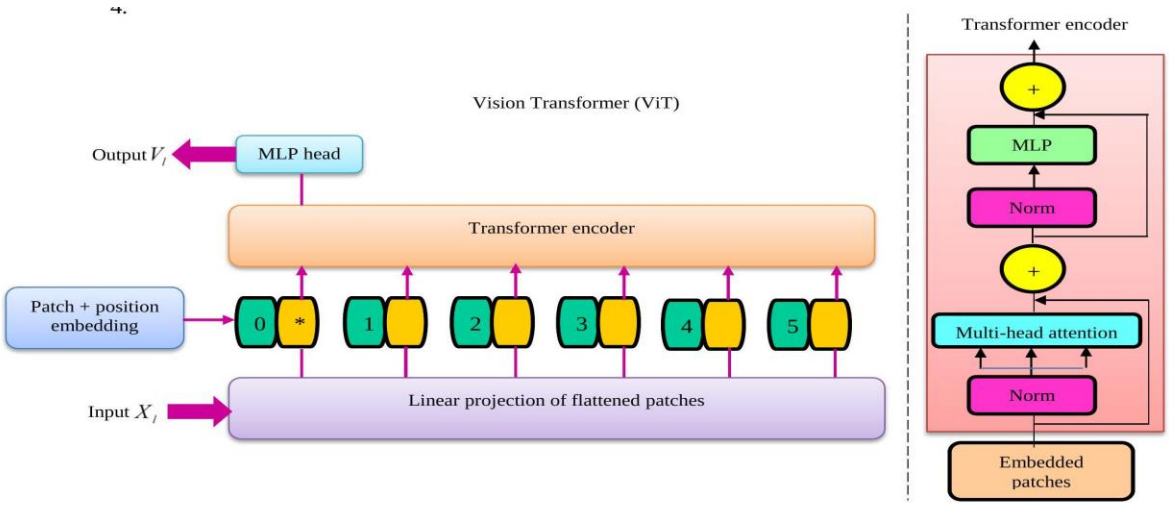


Figure 4.
Structure of ViT.

4.2. Dataset Description

The ATLAS Dataset [15] is divided into two sets: a training set and a testing set. The training set contains data from 60 patients from 2012 to 2020, including images and labels for liver tumors in 90 formats. The testing set contains 30 patients from 2020 to 2023, with the same data structure. The dataset is specific to the ATLAS dataset and no external datasets were used in testing. To enhance the model's generalizability and robustness, it is suggested to validate the method using additional external datasets.

4.3. Experimental Outcomes

Figure 5 interprets the experimentation results of ViT+LIFT based Stat features. Figure 5 a), b), c), d) and e) displays input image-1, pre-processed image-1, liver area segmented image-1, lesion segmented image-1 and classified image-1 whereas input image-2, pre-processed image-2, liver area segmented image-2, lesion segmented image-2 and classified image-2 are revealed in figure 5 f), g), h), i) and j). Figure 5 k), l), m), n) and o) describes input image-3, pre-processed image-3, liver area segmented image-3, lesion segmented image-3 and classified image-3.

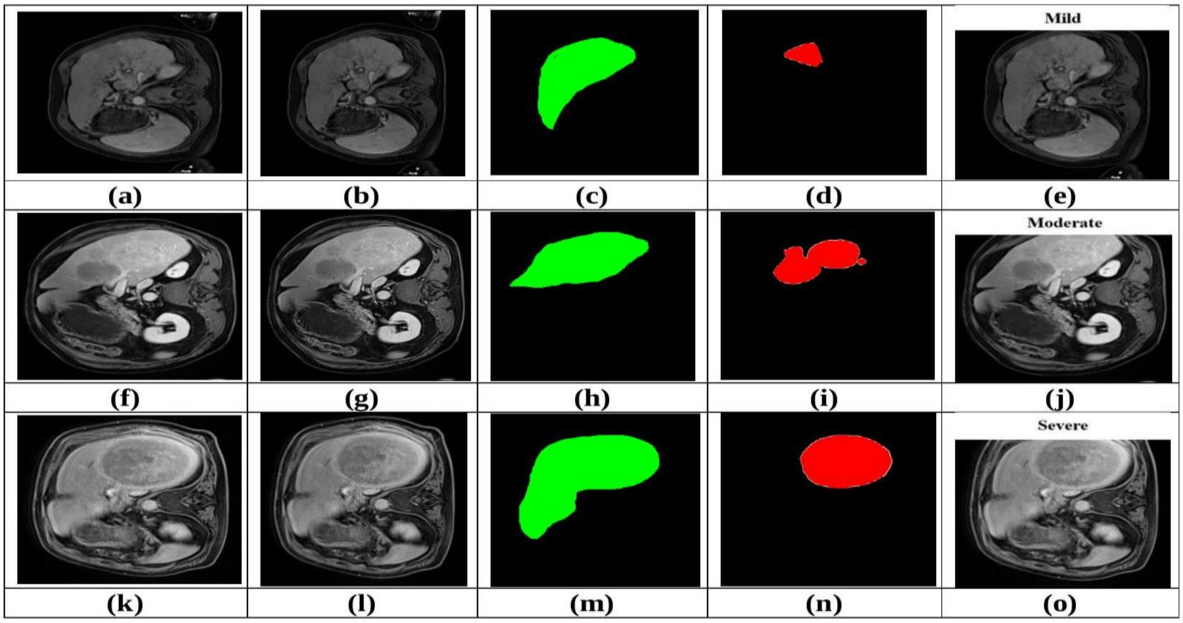


Figure 5.
Experimental results of ViT+LIFT based Stat features, a) input image-1, b) preprocessed image-1, c) liver area segmented image-1, d) lesion segmented image-1, e) classified image-1, f) input image-2, g) pre-processed image-2, h) liver area segmented image-2, i) lesion segmented image-2, j) classified image-2, k) input image-3, l) pre-processed image-3, m) liver area segmented image-3, n) lesion segmented image-3, o) classified image-3.

4.4. Evaluation Metrics

Accuracy, sensitivity and specificity are taken into concern as metrics to evaluate ViT+LIFT based Stat features. The dataset used has a class imbalance between benign and malignant cases, potentially affecting the model's performance evaluation. This can lead to misleading performance metrics like accuracy. Future work will include F1-score and precision-recall curves to provide a more accurate and balanced assessment of the model's effectiveness, ensuring a more accurate and balanced evaluation of its effectiveness.

4.4.1. Accuracy

Accuracy [3] refers to correctness of model to classify liver tumor and it is formulated by,

$$Acc = \frac{T_{pos} + T_{neg}}{T_{pos} + F_{pos} + F_{neg} + T_{\mathrm{neg}}} \tag{22}$$

Here,
Acc = Accuracy of the model; T_{pos} = True positive count; T_{neg} = True negative count F_{pos} = False positive count; F_{neg}= False negative count

4.4.2. Sensitivity

Sensitivity [3] specifies to an ability of model for classifying individual with liver tumor, which is modeled as,

$$Sen = \frac{T_{pos}}{T_{pos} + F_{neg}} \tag{23}$$

4.4.3. Specificity

Sensitivity [3] illustrates to a competence of model for detecting individuals without liver tumor that can be represented as,

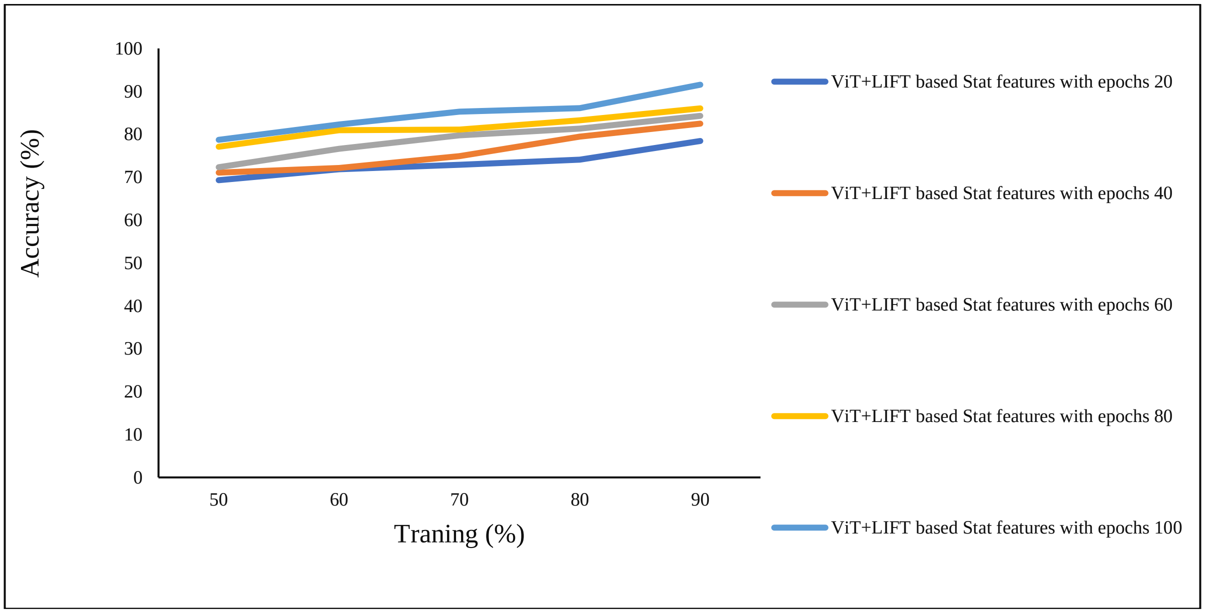
$$\text{Spe} = \frac{T_{\text{neg}}}{T_{\text{neg}} + F_{\text{pos}}} \tag{24}$$

4.5. Performance Analysis

The performance of the ViT+LIFT-based statistical features model was assessed by varying the number of training epochs (20, 40, 60, 80, and 100) while utilizing 90% of the training data. Figure 6 illustrates the variations in key performance metrics:

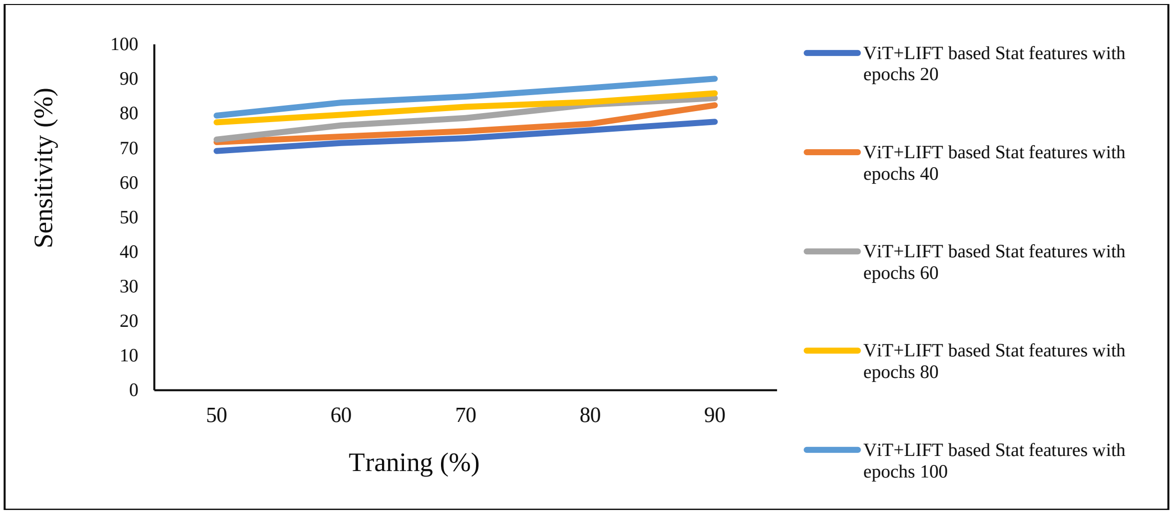
- Accuracy improved from 78.697% at 20 epochs to 91.535% at 100 epochs, demonstrating progressive learning and improved feature representation.
- Sensitivity values increased from 79.395% to 90.043%, indicating the model's enhanced capability to detect liver tumors over time.
- Specificity rose from 80.947% to 90.564%, confirming the model's reliability in distinguishing between tumor and non-tumor cases.

To validate the efficacy of ViT+LIFT-based statistical features, the model was benchmarked against ResNet-10 [16] 3D Convolution+C-LSTM [17] PocketNet+nnUNet [8] and UNet++ [18] using the ATLAS dataset under identical training conditions (90% training, 10% testing). Preprocessing steps such as normalization, augmentation, and resizing were consistently applied across all models for a fair comparison.

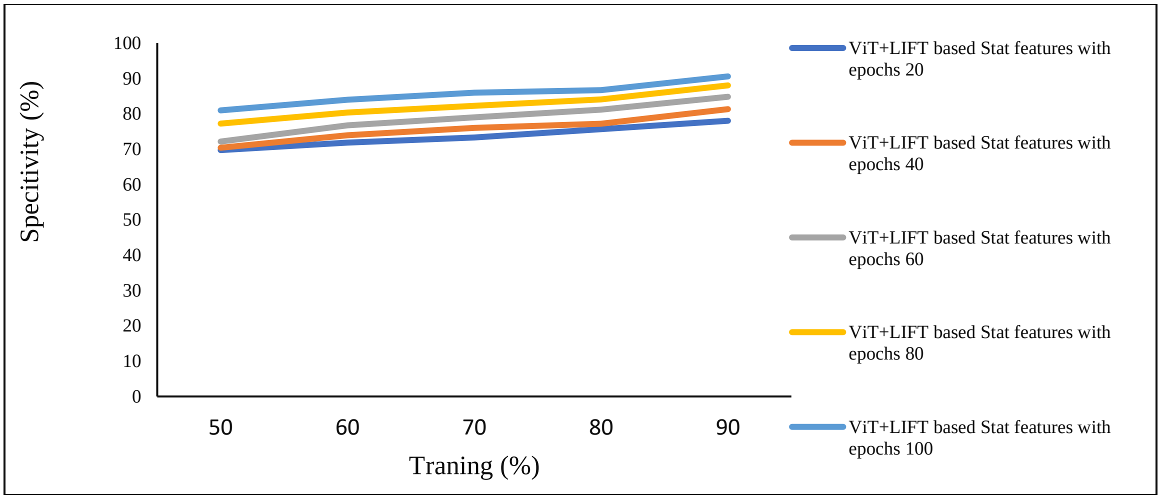


(a)

DOI: 10.55214/25768484.v9i5.7180



(b)



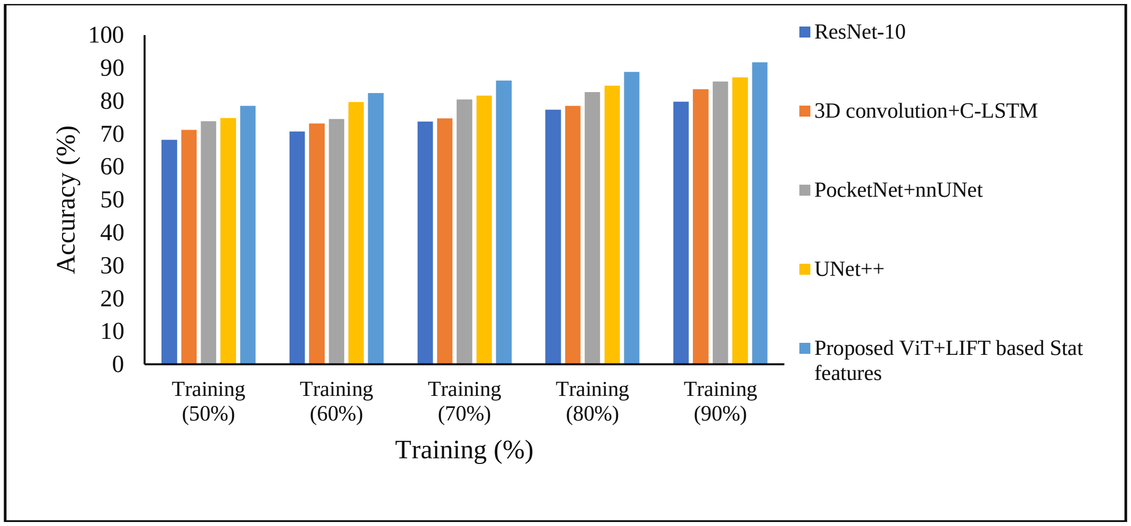
(c)

Figure 6.
Performance estimation of ViT+LIFT-based statistical features: (a) Accuracy, (b) Sensitivity, and (c) Specificity, illustrating the effectiveness of the proposed method across key performance metrics.

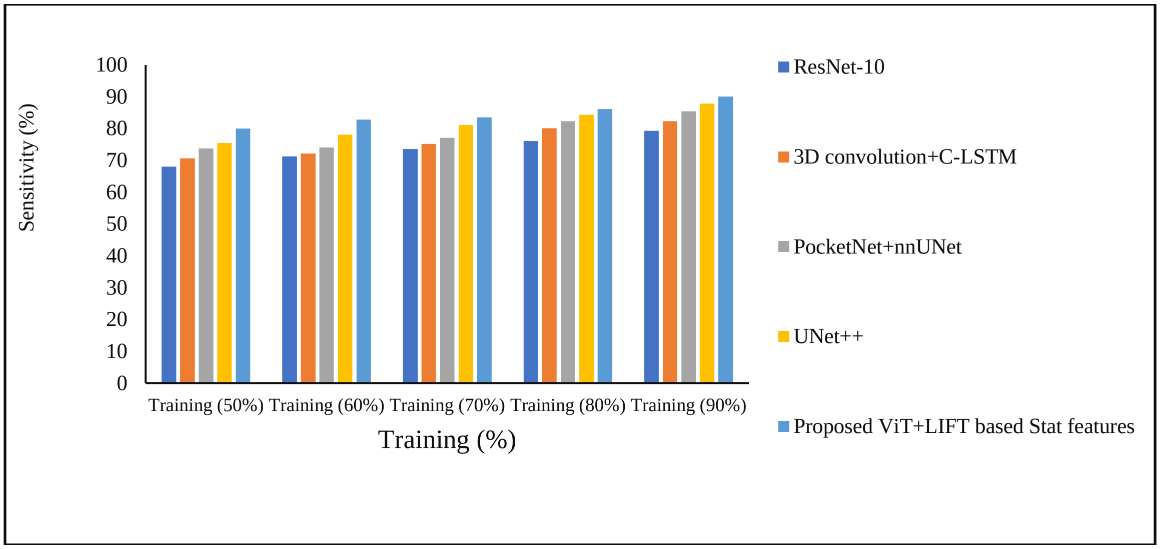
As depicted in Figure 7 and Table 1, the ViT+LIFT-based statistical features method consistently outperformed baseline models across key metrics:

- Accuracy: 91.732%, exceeding ResNet-10 (79.776%), 3D Convolution+C-LSTM (83.533%), PocketNet+nnUNet (85.868%), and UNet++ (87.138%).
- Sensitivity: 90.118%, surpassing ResNet-10 (79.282%), 3D Convolution+C-LSTM (82.319%), PocketNet+nnUNet (85.430%), and UNet++ (87.819%).
- Specificity: 90.710%, outperforming ResNet-10 (80.445%), 3D Convolution+C-LSTM (82.062%), PocketNet+nnUNet (85.671%), and UNet++ (86.433%).

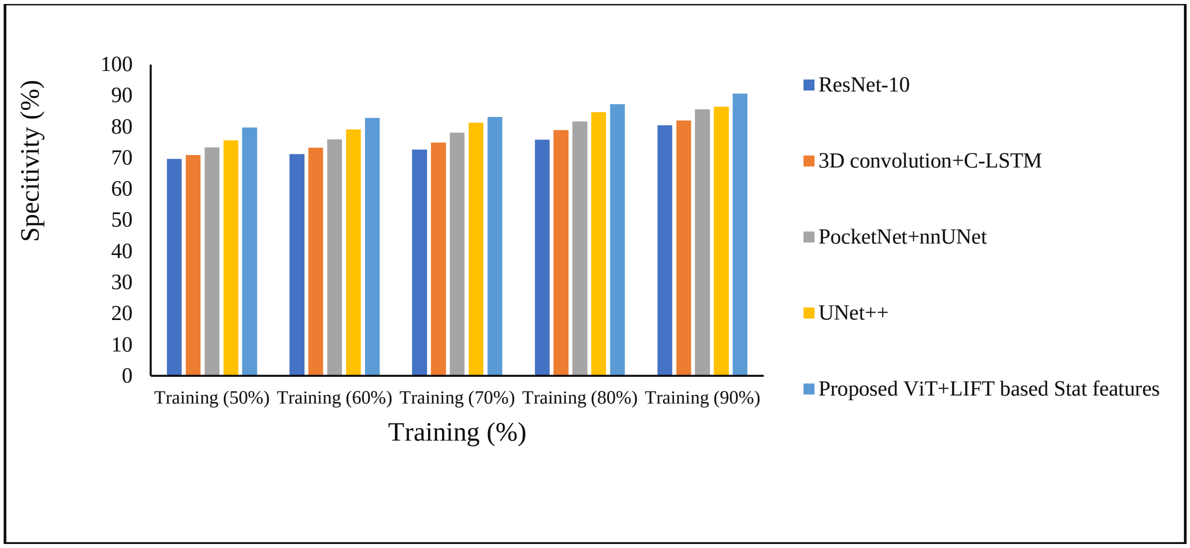
A similar trend was observed for 50% training data, reinforcing the robustness of the ViT+LIFT-based statistical features model in both low-data and high-data scenarios.


Accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
Training (50%)
Training (60%)
Training (70%)
Training (80%)
Training (90%)
Training (%)
ResNet-10
3D convolution+C-LSTM
PocketNet+nnUNet
UNet++
Proposed ViT+LIFT based Stat features

(a)


Sensitivity (%)
100
90
80
70
60
50
40
30
20
10
0
Training (50%)
Training (60%)
Training (70%)
Training (80%)
Training (90%)
Training (%)
ResNet-10
3D convolution+C-LSTM
PocketNet+nnUNet
UNet++
Proposed ViT+LIFT based Stat features

(b)



(c)

Figure 7.
Comparative evaluation of ViT+LIFT-based statistical features: (a) Accuracy, (b) Sensitivity, and (c) Specificity, demonstrating the performance of the proposed method against baseline models for varying training data sizes (50% and 90%).

Table 1.
Comparative Discussion of Vit+Lift Based Stat Features.

Setups	**Metrics/ Methods**	**ResNet-10**	**3D convolution+C-LSTM**	**PocketNet+nn UNet**	**UNet++**	**ProposedViT+LIFT based Stat features**
Training data=50%	Accuracy (%)	68.158	71.238	73.810	74.823	78.493
	Sensitivity (%)	68.016	70.705	73.756	75.464	79.982
	Specificity (%)	69.654	70.898	73.436	75.699	79.724
Training data=90%	Accuracy (%)	79.776	83.533	85.868	87.138	91.732
	Sensitivity (%)	79.282	82.319	85.430	87.819	90.118
	Specificity (%)	80.445	82.062	85.671	86.433	90.710

The ViT+LIFT-based statistical features model demonstrated superior classification performance, effectively distinguishing various liver tumor types. Notably, its high sensitivity highlights its ability to detect even subtle tumor characteristics, while its specificity ensures fewer false positives. The model's ability to outperform established deep learning architectures underscores its potential in clinical diagnostics, reducing diagnostic delays and minimizing the need for invasive procedures. In conclusion, the experimental results confirm that the ViT+LIFT-based statistical features approach is a highly effective and reliable method for liver tumor classification, offering substantial improvements over existing techniques.

5. Conclusion

An exact diagnosing of liver tumor is significant for avoiding redundant liver biopsy. The widely utilized imaging method like MRI has led to constant increase in detection and diagnosing of liver tumor. As Artificial Intelligence (AI) develops, the effective classification methods that are capable to adjust various real-time applications are becoming available. Owing to noises, appearance variabilities and non-homogeneity seen in tumor tissues, the classification of liver tumor is complicated. In this research, ViT+LIFT based Stat features is designed for liver tumor classification. Initially, an input MRI liver tumor image is acquired from the ATLAS dataset. Then, pre-processing of considered MRI liver tumor image is done by AWF. Then, liver area segmentation and lesion segmentation are performed utilizing DCE-Net. After that, features such as SIH, shape features like circularity, perimeter, area and irregularity, ResNet features as well as LIFT with statistical features namely mean, contrast, entropy, energy, variance and homogeneity are extracted. Lastly, liver tumor classification is accomplished employing ViT. Additionally, ViT+LIFT based Stat features achieved accuracy, sensitivity and specificity of 91.732%, 90.118% and 90.710% while training data is considered as 90%. As a future task, types of features essential for liver tumor classification will be selected by removing unnecessary features.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgment:

Authors acknowledge the support from CMR University, REVA University and BIT, Bengaluru for the facilities provided to carry out the research.

Copyright:

References

[1] E.-L. Chen, P.-C. Chung, C.-L. Chen, H.-M. Tsai, and C.-I. Chang, "An automatic diagnostic system for CT liver image classification," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 6, pp. 783-794, 1998. https://doi.org/10.1109/10.678613

[2] A. Hänsch *et al.*, "Improving automatic liver tumor segmentation in late-phase MRI using multi-model training and 3D convolutional neural networks," *Scientific Reports*, vol. 12, no. 1, p. 12262, 2022. https://doi.org/10.1038/s41598-022-16388-9

[3] A. Mohanapriya and S. Malathi, "Comparison and evaluation of BPN and SVM classifier to diagnose liver lesion using CT image," *International Journal of Latest Technology in Engineering, Management & Applied Science*, vol. 3, no. 12, pp. 94-98, 2014.

[4] V. Priya and V. Biju, "SVM based liver tumor classification from computerized tomography images," *International Journal of Advanced Engineering and Nano Technology*, vol. 2, no. 6, pp. 31-36, 2015.

[5] N. P. Nelaturi, V. Rajesh, and I. Syed, "Real-time liver tumor detection with a multi-class ensemble deep learning framework," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16103-16108, 2024. https://doi.org/10.48084/etasr.8106

[6] S. Subha and U. Kumaran, "Efficient liver segmentation using advanced 3D-DCNN algorithm on CT images," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19324-19330, 2025. https://doi.org/10.48084/etasr.9157

[7] S. BJ, S. Seema, and S. Rohith, "A visual computing unified application using deep learning and computer vision techniques," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 1, pp. 59–74, 2024. https://doi.org/10.3991/ijim.v18i01.42673

[8] N. Patel *et al.*, "Training robust T1-weighted magnetic resonance imaging liver segmentation models using ensembles of datasets with different contrast protocols and liver disease etiologies," *Scientific Reports*, vol. 14, no. 1, p. 20988, 2024. https://doi.org/10.1038/s41598-024-71674-y

[9] B. Thejaswini, T. Satheesha, and S. Bhairannawar, "EEG classification using modified KNN algorithm," presented at the 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC), 2023.

[10] A. Kumar and T. Satheesha, "Highly robust and efficient random feature coordination schema using DNN for melanoma skin cancer detection," presented at the 2022 International Conference on Industry 4.0 Technology (I4Tech), 2022.

[11] T. Satheesha, D. Sathyanarayana, and M. G. Prasad, "Proposed threshold algorithm for accurate segmentation for skin lesion," *International Journal of Biomedical and Clinical Engineering*, vol. 4, no. 2, pp. 40-47, 2015. https://doi.org/10.4018/978-1-5225-0549-5.ch009

[12] K. A. Kumar and T. Satheesha, "An efficient method to minimize the depth estimation error in melanoma skin cancer classification," in *2022 4th International Conference on Circuits, Control, Communication and Computing (I4C)*, 2022: IEEE, pp. 25-29.

[13] R. Hamsalekha and T. Satheesh, "Design and implementation of convolutional neural network model for melanoma classification," presented at the 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), 2023.

[14] T. Satheesha, D. Satyanarayana, and M. Giriprasad, "SVMs classification based on Insitu melanoma," presented at the International Conference on Circuits, Communication, Control and Computing, 2014.

[15] ATLAS Dataset, "ATLAS challenge dataset," Retrieved: https://atlas-challenge.u-bourgogne.fr/dataset, 2024.

[16] A. A. Goedhart, "Classification of primary liver tumors with radiomics and deep learning based on multiphasic MRI," Master's Thesis, Delft University of Technology. TU Delft Repository, 2023.

[17] R. Zheng *et al.*, "Automatic liver tumor segmentation on dynamic contrast enhanced MRI using 4D information: deep learning model based on 3D convolution and convolutional LSTM," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2965-2976, 2022. https://doi.org/10.1109/TMI.2022.3184471

[18] J. Wang, Y. Peng, S. Jing, L. Han, T. Li, and J. Luo, "A deep-learning approach for segmentation of liver tumors in magnetic resonance imaging using UNet++," *BMC Cancer*, vol. 23, no. 1, p. 1060, 2023. https://doi.org/10.1186/s12885-023-11432-x

[19] M. Ali *et al.*, "Segmentation of MRI tumors and pelvic anatomy via cGAN-synthesized data and attention-enhanced U-Net," *Pattern Recognition Letters*, vol. 187, pp. 100-106, 2025. https://doi.org/10.1016/j.patrec.2024.11.003

[20] L.-W. Tsai *et al.*, "RFiLM U-Net: Radiomic feature-integrated linear modulation network for precise liver tumor segmentation," *Journal of Medical and Biological Engineering*, pp. 1-10, 2025. https://doi.org/10.1007/s40846-025-00938-3

[21] H. Rahman *et al.*, "Automatic liver tumor segmentation of CT and MRI volumes using ensemble ResUNet-InceptionV4 model," *Information Sciences*, p. 121966, 2025. https://doi.org/10.1016/j.ins.2025.121966

[22] P. Sivanagaraju, S. V. Ramana, and P. P. Reddy, "An advanced transformer framework for liver tumor segmentation using MRI images," *Biomedical Signal Processing and Control*, vol. 107, p. 107808, 2025. https://doi.org/10.1016/j.bspc.2025.107808

[23] F. Wu, W. Yang, L. Xiao, and J. Zhu, "Adaptive wiener filter and natural noise to eliminate adversarial perturbation," *Electronics*, vol. 9, no. 10, p. 1634, 2020. https://doi.org/10.3390/electronics9101634

[24] J. Liu, L. Shao, C. Zhou, Z. Yan, Y. Han, and Y. Song, "DCE-Net: A dynamic context encoder network for liver tumor segmentation," *Research Square*, 2023. https://doi.org/10.21203/rs.3.rs-2272616/v1

[25] A. B. L. Larsen, J. S. Vestergaard, and R. Larsen, "HEp-2 cell classification using shape index histograms with donut-shaped spatial pooling," *IEEE Transactions on Medical Imaging*, vol. 33, no. 7, pp. 1573-1580, 2014. https://doi.org/10.1109/TMI.2014.2318434

[26] A. B. L. Larsen, A. B. Dahl, and R. Larsen, "Oriented shape index histograms for cell classification," in *Image Analysis: 19th Scandinavian Conference, SCIA 2015, Copenhagen, Denmark, June 15-17, 2015. Proceedings 19*, 2015: Springer, pp. 16-25.

[27] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "Lift: Learned invariant feature transform," presented at the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14, 2016.

[28] S. P. Deshmukh and D. Choudhari, "Novel dual neural network for the classification of liver cancer," *NeuroQuantology*, vol. 20, no. 9, pp. 303–309, 2022. https://doi.org/10.14704/nq.2022.20.9.NQ22027

[29] P. R. Kumar and S. Dhenakaran, "Structural (Shape) feature extraction for ear biometric system," in *Proceedings of the International Conference on Signal, Networks, Computing, and Systems: ICSNCS 2016*, 2017, vol. 1: Springer, pp. 161-168.

[30] D. McNeely-White, J. R. Beveridge, and B. A. Draper, "Inception and ResNet features are (almost) equivalent," *Cognitive Systems Research*, vol. 59, pp. 312-318, 2020. https://doi.org/10.1016/j.cogsys.2019.10.004

[31] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020. https://doi.org/10.48550/arXiv.2010.11929

[32] H. Yin, A. Vahdat, J. M. Alvarez, A. Mallya, J. Kautz, and P. Molchanov, "A-vit: Adaptive tokens for efficient vision transformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 10809-10818.

DOI: 10.55214/25768484.v9i5.7180