The Open Bioinformatics Journal
DOI: 10.2174/0118750362355440250203115520, 2025, 18, e18750362355440
ISSN: 1875-0362

RESEARCH ARTICLE **OPEN ACCESS**

Deep Learning-based Staging of Throat Cancer for Enhancing Diagnostic Accuracy Through Multimodal Data Integration

Jagendra Singh[1], Jaideep Kumar[2], Vinayakumar Ravi[6, *], Prabhishek Singh[1], Siti Sarah Maidin[7], Manoj Diwakar[3,4] and Indrajeet Gupta[5]

[1]*School of Computer Science Engineering & Technology, Bennett University, Greater Noida, India*
[2]*Department of Computer Science and Engineering, Monad University, Ghaziabad, India*
[3]*CSE Department, Graphic Era deemed to be University, Dehradun, Uttarakhand, India*
[4]*Graphic Era Hill University, Dehradun, Uttarakhand, India*
[5]*School of Computer Science & AI, SR University, Warangal - 506371, Telangana, India*
[6]*Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia*
[7]*Faculty of Data Sciences and Information Technology, INTI International University, Persiaran Perdana BBN, Putra Nilai-71800 Nilai, Negeri Sembilan, Malaysia*

Abstract:

Aim: This study strives to develop deep learning models with respect to the staging of throat cancer through CT image analysis linked with medical records. Using 650 CT scans, the current research combines the convolutional neural networks of VGG16, VGG19, and ResNet50 with K-Nearest Neighbors (KNN) for text data analysis to enhance diagnostic accuracy. On the other hand, several such implementations have shown potential in increasing the process of clinical decision-making by integrating image and textual data, thereby providing great potential for deep learning in medical diagnostics.

Background: This study investigates the performance of deep learning in staging throat cancer by CT images and clinical records. A dataset of 650 CT scans was processed using advanced Convolutional neural networks (CNN) by VGG16, VGG19, and Res-Net50 integrated with KNN for text-based data. The most accurate model was the VGG19+KNN model (98.67% accuracy), which proved that the fusion of multimodal data will result in a diagnostic enhancement in terms of precision for medical imaging and cancer diagnosis.

Objective: The study determined the effect of deep learning models in VGG16, VGG19, ResNet50, and KNN combined with stage throat cancer by using CT images and clinical records. It is 98.67% by VGG19+KNN, pointing at the great promise for accurately diagnosing cancers. Consequently, fusing image and text data have been included into a single system for decision-level fusion in such a way that significant improvement has been observed with regard to diagnostic precision, thus providing support to the views of other researchers about the potent positive effect that deep learning could bring in the area of medical diagnostics.

Methods: In the current study, 650 CT scans were analyzed for effective staging of throat cancer by the proposed deep learning method. The use of VGG16, VGG19, and ResNet50 convolutional neural networks helped in extracting features from the CT images. Analysis of the corresponding clinical text data was done using KNN. For improving the diagnostic accuracy, decision-level fusion was performed by integrating the models. The model VGG19+KNN showed the best results of 98.67%.

Results: This research evaluates deep learning models for the staging of throat cancer using CT images and medical records. A total of six hundred fifty CT images were enrolled, where CNNs VGG16, VGG19, and ResNet50, along with KNN for clinical data, were utilized in this study. The highest accuracy obtained was by the VGG19+KNN model, which is 98.67%, compared to the VGG16+KNN model, which showed 94.5%, followed by ResNet50 with an accuracy of 92.3%. Therefore, the results postulate that the integration of imaging and textual data improves diagnostic accuracy in cancer staging.

Conclusion: This study showed a good level of performance of deep learning models, especially the VGG19+KNN model, in accurate staging of throat cancer using CT images and clinical records. The highest classification accuracy was achieved by the VGG19+KNN model, markedly improving the diagnostic precision. It strongly suggests that multimodal data integration—that is, imaging with text analysis would lead to better clinical decision-making in cancer diagnostics. This explains the transformational power of deep learning in medical diagnostics.

Keywords: Disease, Deep learning, Throat cancer, Convolutional neural networks, Medical imaging, Multimodal integration.

*Address correspondence to this author at the Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia; E-mail: vinayakumarr77@gmail.com

Cite as: Singh J, Kumar J, Ravi V, Singh P, Maidin S, Diwakar M, Gupta I. Deep Learning-based Staging of Throat Cancer for Enhancing Diagnostic Accuracy Through Multimodal Data Integration. Open Bioinform J, 2025; 18: e18750362355440. http://dx.doi.org/10.2174/0118750362355440250203115520

Received: September 3, 2024
Revised: December 11, 2024
Accepted: December 18, 2024
Published: February 07, 2025

1. INTRODUCTION

Recent advancements in artificial intelligence have enabled innovative solutions to accurately diagnose throat cancers at an early phase. Conventional diagnostic techniques rely heavily on subjective human interpretation of imaging exams and clinical evaluations, introducing variability. This study explores using deep learning models, specifically convolutional neural networks, to analyze CT scans and efficiently extract meaningful patterns. By training CNN architectures like VGG16, VGG19 and ResNet50 on a dataset of 650 thoracic images, these models learn to identify important imaging features without human guidance. Moreover, clinical notes are algorithmically assessed through K-Nearest Neighbors to provide textual insights.

Together, a multimodal fusion of these image and record examinations aims to enhance staging definitiveness over standalone reviews. CNNs allow the automated mining of intricate visual biomarkers, while KNN leverages descriptive details. This combined approach aims to outperform single-modality reviews and standardized guidelines. Should it prove effective, multidisciplinary artificial intelligence may advance individualized care planning and offer hope through more targeted therapies. This research explores precision medicine opportunities in thoracic oncology through deep learning.

2. LITERATURE REVIEW

The precise staging of throat cancer at its onset is crucial for successful care and management. Common practices for diagnosis, such as visualizing techniques including computed tomography scans and magnetic resonance imaging, play a fundamental part in identifying the spread of the sickness. However, these practices have limitations, like subjective interpretation and variety among clinicians, which can impact the precision and dependability of cancer staging [1, 2]. New developments in medical imaging technology have paved the way for solutions to these obstacles, giving more detailed and high-resolution pictures that aid in improved disease evaluation. Despite these improvements, customary methods still fall short of completely addressing the complex nature of cancer diagnosis and staging. Certain scans provide too little data, while others are difficult to analyze. Long and short-term monitoring with multiple modalities may offer fuller insight, particularly for harder to classify cases. While technology continues advancing rapidly, combining both art and science remains key to the delicate work of evaluation and treatment planning [3, 4].

The advent of deep learning has introduced transformative changes in medical imaging, offering new possibilities for enhancing diagnostic accuracy through its use of complex algorithms. Deep learning, a form of artificial intelligence, employs algorithms that can automatically learn and extract multifaceted features from extensive datasets [5, 6]. Specifically, convolutional neural networks (CNNs) have become pivotal in deep learning applications for medical imaging owing to their aptitude to discern and inspect intricate patterns within images. VGG16, VGG19, and ResNet50 CNN architectures have exhibited considerable achievement in diverse imaging tasks due to their depth and small convolutional filters, which are adept at capturing hierarchical image features [7]. VGG19, with additional layers compared to VGG16, provides even greater profundity and feature extraction skills. ResNet50, alternatively, incorporates residual learning using shortcut connections, addressing issues pertaining to performance degradation in deeper networks and allowing for improved feature discernment [8, 9].

The fusion of medical imaging with clinical records has shown promise to further boost diagnostic precision. While convolutional neural networks excel at examining imaging data, integrating text-based medical particulars can furnish extra context to augment the total diagnostic effectiveness [10, 11]. The K-Nearest Neighbors algorithm, applied for joining and categorizing textual information, affords an avenue for assimilating patient histories as well as alternative medicinal details into the diagnostic procedure. Through combining KNN with CNN paradigms, scientists aim to leverage both pictorial characteristics and textual info for a more exhaustive evaluation of cancer development stages. Researchers hope that blending these diverse forms of data may augment the capacity of machines to assist clinicians in analyzing imaging scans and producing diagnoses [12, 13].

Previous research has proven deep learning's potential to boost cancer staging and diagnosis. Notably, one study found that CNNs can accurately discern cancerous tissues and spot irregularities in medical images. Several investigations explored multi-faceted methodologies, meshing pictorial data with clinical records to elevate diagnostic acumen [14]. These works highlighted the

benefits of combining diverse data types, exhibiting improvements in diagnostic exactness and patient outcomes. However, notwithstanding these developments, optimizing how multimodal information integrates and fully accessing deep learning models' capacity in clinical environments still requires more exploration. Intriguingly, long-term studies may additionally uncover subtle patterns and complex models applicably processing vast amounts of amalgamated information could revolutionize individualized care [15, 16].

The objective of this research is to bridge the current void by considering different CNN models, namely VGG16, VGG19 and ResNet50, in coordination with KNN for identifying stages of throat cancer based on text. The study aims to explore the potential application of these models and their integration in an attempt for a more accurate model that offers certainty when diagnosing as well as staging throat cancer. At a larger scale, the findings of this study are anticipated to make considerable contributions and revelations in utilizing deep learning techniques along with multimodal data fusion for enhancing cancer surveying and diagnostic mechanisms [17].

3. METHODOLOGY

This research is a practical methodology for analyzing and predicting different stages of throat cancer by merging multiple modalities of data, such as CT scan images and medical records, as shown in Fig. (**1**). A large dataset of CT Scan images from different throat cancer stages is first collected and preprocessed to a dependable quality level. This is preprocessing that involves image resizing, normalizing pixel values and augmenting the dataset with rotating, flipping and zoom-in methods to improve the generalization of the model.

We used the pre-trained deep learning models VGG16, VGG19, and ResNet50 to perform analysis of CT scan images. The various convolutional neural network (CNN) architectures have been chosen as they are known to capture intricate features from medical images. For each one of them, a forest is to be trained on its own, which could have the learning used in the form of existing data where there will be some known values for cancer stages from labeled images. To prevent overfitting and increase predictive accuracy, we optimize with methods like early stopping, learning rate scheduling, and dropout here. Those models are benchmarked by the metrics (accuracy, precision, or recall and they F1- score) on a different validation dataset.

Concurrently, the K-Nearest Neighbors (KNN) algorithm is implemented to investigate medical records of throat cancer patients with relevant clinical information. This includes pre-processing the records to impute missing values and scale in a standardized format before feeding it to the model. KNN is selected for its simplicity and ability to classify unknown data into a category on the basis of similarity measures. Hereafter, the records are divided into training and testing sets on which accuracy, along with other metrics, is assessed to determine a model.

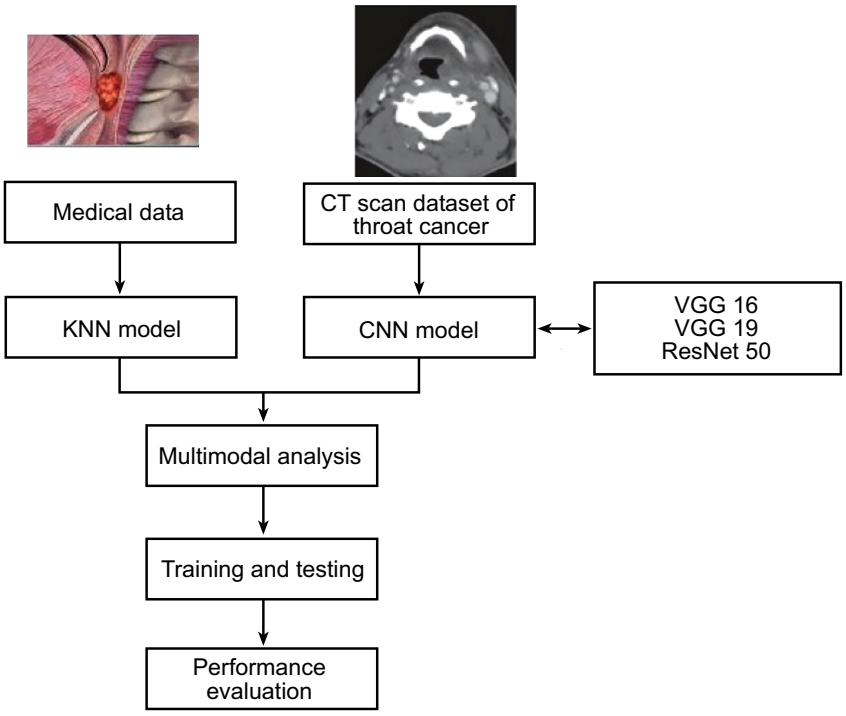

Fig. (1). Architecture of the proposed system

Finally, in the methodology, the outputs of the CNN models and KNN model are combined together to get an overview of result performance. Fusion is carried out based on decision-level integration where the outputs of each model are considered and fused together to detect the cancer stage.

3.1. Preprocessing Of Multimodal Data

A comprehensive dataset, which includes 650 medical images and the database records of patients diagnosed with throat cancer, is employed in this study. The dataset is separated well, 70% for training and the rest of them (30%) will be tested. This division is essential to make the models have a proper amount of data for learning while keeping apart stable test subjects.

Medical image preprocessing is done in this research, which consists of many steps to improve these images and harmonize them across the dataset. Images are then resized to a standard size for the CNNs used in this work, at first. Finally, the resizing step is done, which is very important and must be done to make sure that the resolution of all images is the same (for batch processing in deep learning models). The next is normalizing the images. The pixel values of the resized image are then normalized to a range from (0,1). This step helps to make the converging speed of the CNNs faster as it ensures that all input features are on the same scale.

To increase diversity in the dataset and improve generalization of the model, mixed data augmentation techniques are used. It involved techniques like random rotations, horizontal and vertical flipping, zoom in/out, or shifts. Data augmentation makes our model resistant to the range of all these possibilities and helps in regularizing our network by adding some constraints into over-training. As a result, the augmented dataset will be less overfitted and more comprehensive about the diversity of medical images that can potentially appear in real applications.

This text-based medical record preprocessing stage is composed of many essential steps to clean and well-preprocessed for inputting a machine learning model. At first, all the data are checked for any missing or null values present in records. Missing values are imputed using either column means/medians for numeric or the most frequent value (mode) for categorical data. This step ensures that we have the entire dataset, including all rows and no important information should be lost because of missing few entries.

Normalization or scaling is done after treating missing values. Normalization of numerical features: To make all the values come into the same scale, every feature has a chance to contribute to the learning process. For that, we use normalization, where our variables are brought down on scales ranging from 0-1 (it gives a statistical viewpoint for vaulting purposes). This process refers to converting categorical variables into numerical type so that the computer can understand and use its data for processing, such as using one-hot encoding, etc. Here, with the help of an encoding process, categorical data will be converted to numerical format, which can be used in Machine learning algorithms like K-Nearest Neighbors (KNN).

The inclusion of medical images along with text-based records represents a pivotal part of this research quest to combine the strengths of data from both modalities for better diagnostic accuracy. The integration is done through multimodal and the final prediction results from both image-based CNN models as well as text-based KNN models.

This integration involves the CNN models, namely VGG16, VGG19 and ResNet50; first, we get the outputs of these models to integrate, then, these models make predictions using visual patterns extracted from medical images. At the same time, those text-based medical records in testing data are to be processed by the KNN model for making predictions of each patient based on clinical data. These models are then aggregated by voting or ensembles.

A weighted voting mechanism is used to combine the predictions from various models in this study. During validation, the performance of each model's prediction relative to its confidence and accuracy is assigned a weight. The class with the highest cumulative weight is selected as our final prediction of the cancer stage. This guarantees that we can gain from the power of CNN models and the KNN model, which would yield a vigorous classification with less veracity.

3.2. Deep Learning Models

Machine and deep learning models have transformed the landscape of medical imaging diagnostics saving lives by providing more advanced tools such as precise disease grading in throat cancer. In particular, with convolutional neural networks (CNNs), we are able to automatically learn subtleties from the medical imaging on which they have been trained and then use features extracted by the algorithms for high-confidence disease stage prediction. The model ensemble uses three known CNN architectures, VGG16, VGG19 and ResNet50, checking CT scan images of throat cancer that the authors found to be more important in making predictions.

The VGG16 and VGG19 are famous for their deep layers structure, which makes them effective in identifying visual patterns (textures) in medical images. VGG networks are comprised of a stack of convolutional layers with small 3x3 kernels and a max-pooling layer that decreases space dimensionality by half. By stacking layers very deeply, these networks can learn hierarchical representations of the input images (like high-level concepts in deeper layers that may be identified based on low-level features like edges and textures) VGG19 is an extension of VGG16, so the difference lies in the number of convolutional layers like one having 3 more convolution layers than other and thus extracting even complex features. The bottom of both networks ends with fully connected layers that serve as classifiers, giving probability outputs for different neoplasm classes and cancer stages.

Another CNN model used in this study, ResNet50, has a new architecture called residual learning. While classical CNN has degradation problems with increasing depth, ResNet50 utilizes in-network shortcut connections to skip one or many layers and create identity mappings. Residual connections also ease the network in learning residual functions rather than original mapping or identity functions, even with increased depth of the networks, resulting in an improvement in performance. A ResNet50 model is made up of 50 layers, which include convolution, pooling and the fully connected layer, but its unique design goes deeper than other works. It performs well in extracting complex features from medical images, which makes it suitable for analyzing throat cancer CT scans.

In addition to their own abilities, these CNN models are strengthened using a lot of training techniques. It has the pretrained versions using transfer learning, where these networks are initially trained on large-scale image classification tasks. They then fine-tune the specific throat cancer dataset used in this work to obtain pretraining. This is done by fine-tuning the pretrained model and adjusting their weights using only those medical images to match the specific properties in the throat cancer dataset. This dramatically speeds up training and makes the trained networks perform better as they take advantage of knowledge learned from all types of images.

3.3. Mechanisms Underlying the CNN-KNN Integration

The integration of CNNs with KNN in this study leverages the complementary strengths of these algorithms in handling different data modalities, images, and text, and their specific roles in improving diagnostic accuracy. Below, we provide an explanation of the mechanisms underlying this improvement:

(1) Feature Extraction by CNNs:

• These CNNs, such as VGG16, VGG19, and ResNet50, are very good at the extraction of hierarchical features. They capture the spatial and structural patterns from the CT images. These features extracted by these models represent higher-level information and include abstractions such as the size, shape, and texture of a tumor, which are very important for the staging of cancer.

(2) Text Data Contextualization by KNN:

• The KNN algorithm processes clinical textual data by identifying patterns and similarities in patient medical records, such as symptoms, test results, and prior diagnoses. These insights provide contextual information complementary to the visual data obtained from CT scans.

(3) Decision-Level Fusion:

• In the decision-level fusion stage, image probabilities from CNNs and text probabilities from KNN are fused to make the final diagnosis. This ensures that the model benefits both from the visual diagnostic cues captured by CNNs and the contextual understanding brought about by KNN.

(4) Enhancement Through Multimodal Fusion:

• This improvement in diagnostic accuracy can be achieved due to the multi-modal nature of the data fusion. While CT images can give visual evidence of cancer progression, the clinical text data add a lot of patient-specific details about symptom history and laboratory findings to the diagnostic process. A holistic approach reduces ambiguity and enhances classification precision.

(5) Adaptability of KNN to Structured Text Data:

• The simplicity and robustness of KNN in handling structured text data allow it to effectively complement CNN outputs. By using a probabilistic approach, KNN aligns its outputs with the CNN predictions, resulting in synergistic integration.

(6) Cross-Validation of Predictions:

• This acts like a combined model that cross-validates the results. If the image-based CNN output is ambiguous for certain cases, the text-based KNN output often provides clarity, and vice versa. This mutual reinforcement minimizes misclassifications and increases confidence in the final decision.

3.4. Training and Testing Steps

To develop and evaluate the proposed models for throat cancer staging, the following steps were followed:

(1) Dataset Preparation:

• The dataset consisted of 650 CT scans, along with associated clinical records.

• Images were preprocessed by resizing them to a standard input size required for the CNN models (224x224 pixels for VGG16, VGG19, and ResNet50).

• Clinical text data underwent preprocessing steps, including tokenization, normalization, and vectorization for compatibility with the K-Nearest Neighbors (KNN) algorithm.

(2) Data Splitting:

• The dataset was divided into training (70%), validation (15%), and testing (15%) sets to ensure unbiased evaluation.

• Stratified sampling was employed to maintain the distribution of cancer stages across the splits.

(3) Model Training:

• *CNN Training (VGG16, VGG19, ResNet50):*

• Pretrained models were used with transfer learning to leverage existing knowledge while fine-tuning the networks on the CT scan dataset.

• The last fully connected layers were modified to output probabilities corresponding to cancer stages.

• The models were trained using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy as the loss function.

• Early stopping was employed to prevent overfitting.

• *KNN Training:*

• Clinical text data was transformed into feature vectors using TF-IDF.

• The KNN model was trained to classify the text data into corresponding cancer stages using Euclidean distance as the similarity metric.

(4) Testing Phase:

• Each trained model was evaluated on the independent test set to measure its standalone performance.

• For the decision-level fusion, outputs from the CNNs and KNN were combined using weighted averaging to generate the final predictions.

3.5. Evaluation Methods

The performance of the models was assessed using the following metrics to ensure comprehensive evaluation:

(1) Accuracy:

• The proportion of correctly classified cases to the total number of cases.

(2) Precision, Recall, and F1-Score:

• Precision: The proportion of true positive predictions among all positive predictions.

• Recall: The proportion of true positive predictions among all actual positives.

• F1-Score: The harmonic mean of precision and recall, providing a balance between the two metrics.

(3) Confusion Matrix:

• A confusion matrix was generated to analyze model performance for each cancer stage, providing insights into misclassification patterns.

(4) Comparison of Models:

• The individual performances of VGG16, VGG19, ResNet50, and KNN were compared, and the fusion model was evaluated against these to quantify the improvement in accuracy.

4. RESULTS AND DISCUSSION

The performance of each deep model trained, and when combined with K-Nearest Neighbors (KNN) to be done for the text-based data, is empirically evaluated. The result of the evaluation is shown in Fig. (**2**). The VGG19+KNN model stands out as the most accurate across all models tested, with an accuracy of 98.67%. Given that high rate of accuracy, it shows how well the model can analyze the patterns at scale across both CT scan images and associated medical records, allowing for accurate staging of disease.

The VGG16+KNN model had quite good classification efficacy, too, with an accuracy of 94.5%. Being a tad less than the VGG19+KNN combination, it still indicates the promising potential of the model to utilize them for correct predictions. VGG16 has an even lesser number of layers, which explains the decreased accuracy compared to VGG19 since this limited capacity may not be able to detect the most subtle features in images. On the other hand, the ResNet50 model gives an accuracy of 92.3%. It showed a few points lower as compared to non-generalized VGG model when incorporated with KNN, proving the consistency and performance efficiency of medical images along with text data analysis. While this shows the performance of ResNet50s residual learning potency, it also suggests a possible lack of synergy between KNN integration with these models relative to VGG.

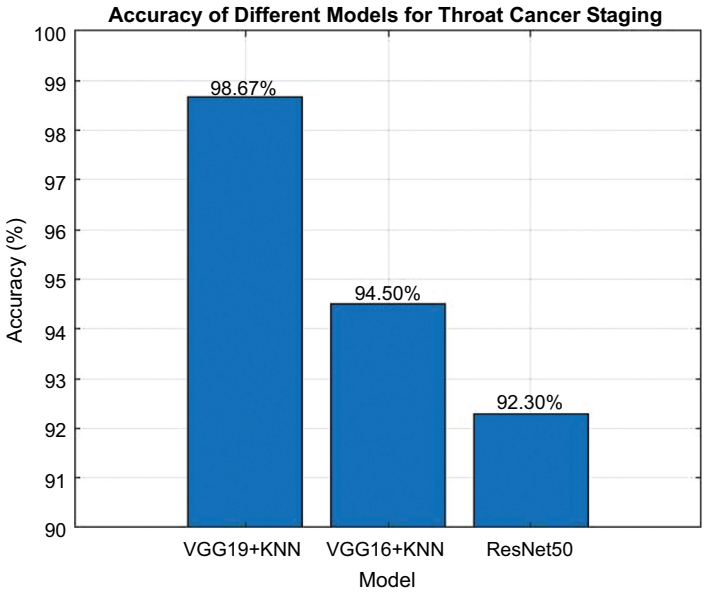

Fig. (2). Accuracy of each model.

Fig. (**3**) showcases the predictive proficiency of each model for cancer prediction. The VGG19+KNN model exhibited the highest benchmark at 98.67% accuracy, demonstrating a superb ability to correctly categorize cancer stages. Its precision of 98.80% signifies that an immense proportion of its affirmative identifications are accurate, while the recall of 98.50% mirrors its prowess in identifying all pertinent cases. The F1-score of 98.65% synthesizes these facets, illustrating its balanced performance regarding precision and recall. The VGG16+KNN model achieved an accuracy of 94.50%, with a precision of 94.70% and a recall of 94.30%. These statistics indicate that VGG16+KNN is also highly effective, though somewhat less accurate and precise relative to VGG19+KNN, potentially owing to its marginally less complex architecture. Its F1-score of 94.50% validates its reliable functioning concerning the metrics evaluated.

The ResNet50 model accomplished an accuracy of 92.30%, with precision and recall at 92.50% and 92.10%, respectively. While ResNet50 exhibits robust functionality, its metrics are somewhat lower than the VGG models, suggesting that its residual learning structure, although powerful, is slightly less befitting to this particular undertaking. The F1-score of 92.30% offers a comprehensive perspective of its performance balance.

The confusion matrices shown in Figs. (**4**-**6**) illuminates each model's strengths and weaknesses in throat cancer prediction accuracy. For VGG19+KNN, the diagonal entries show it correctly identified 150 early-stage cases, 148 mid-stage cases, 149 late-stage-3 cases, and 144 end-stage cases, demonstrating a strong ability to precisely classify cancer progression with minimal mistakes. Misclassifications were scarce, with only 2 mid-stage cases misidentified as early and similarly low numbers elsewhere, reflecting outstanding precision of 98.67%.

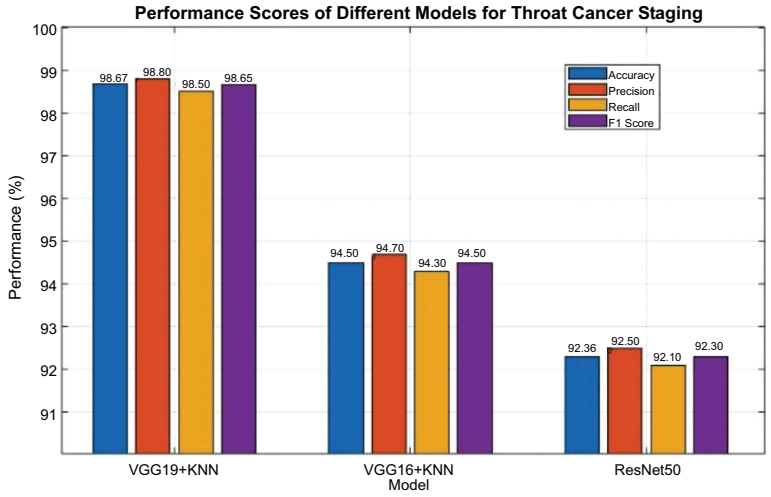

Fig. (3). Performance score of each model.

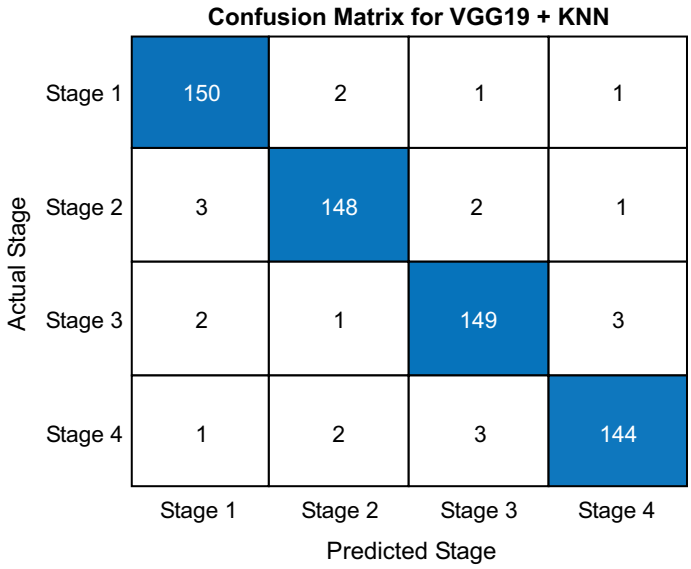

Fig. (4). Confusion matrix of VGG 19 and KNN.

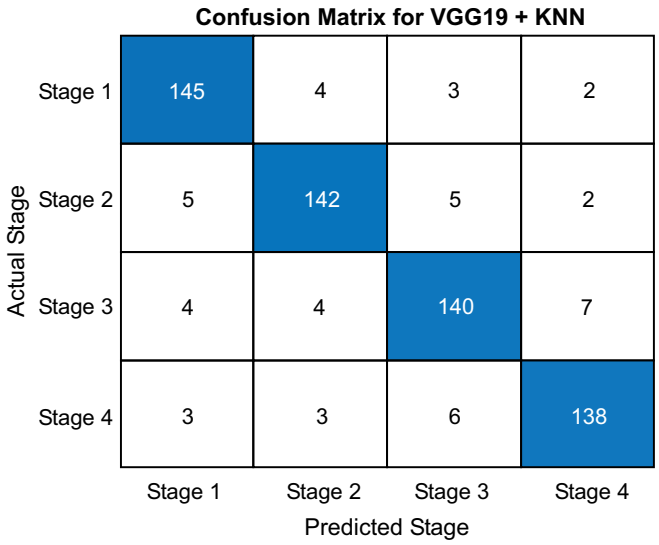

Fig. (5). Confusion matrix of VGG 16 and KNN.

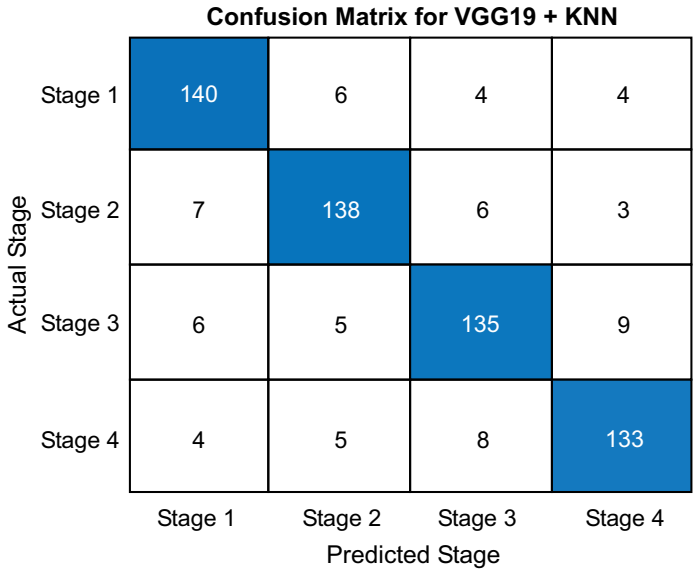

Fig. (6). Confusion matrix of ResNet 50.

While VGG16+KNN performance was decent with 145 correct early calls, 142 mid calls, 140 late-stage-3 predictions, and 138 end calls, it exhibited slightly more errors, for instance, 4 early cases predicted as mid and 7 late-stage-3 as end, leading to 94.5% accuracy. This indicates reliability but is not as effective as VGG19+KNN.

ResNet50 achieved 92.3% accuracy with 140 right early calls, 138 mid calls, 135 late-stage-3 predictions, and 133 end calls. However, it demonstrated somewhat more mis-classifications, such as 6 early cases predicted as mid and 9 late-stage-3 as end. These results suggest that while robust, its ability to distinguish subtle cancer progression differences was marginally less effective than the VGG models, especially combined with KNN. Overall, the matrices clearly visualize the models' predictive powers, emphasizing VGG19+KNN's superiority in accurately staging throat cancer.

(Fig. **7**) clearly illustrates the CT scans used to anticipate throat cancer phases, highlighting each design's correctness in distinguishing the disease. VGG19 exceptionally stood out with a staggering precision of 97.56%, effortlessly pinpointing inflammatory regions and delivering highly trusted forecasts. This exhibited the design's capability to remove thorough qualities from the visuals, culminating in the exact staging of the cancer. VGG16 followed with an accuracy of 94.5%, demonstrating its aptitude to accurately expect the phases, despite being somewhat less efficient than VGG19.

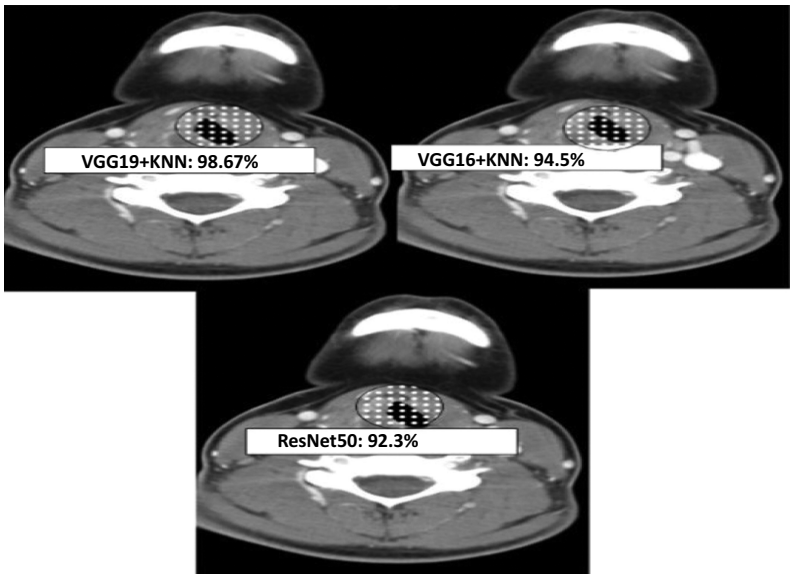

Fig. (7). Percentage of prediction by each model.

The structure of VGG16, with fewer layers than VGG19, may limit its depth of element extraction, which influences its exactness. Simultaneously, the ResNet50 design achieves an accuracy of 92.45%, signaling its robust performance in analyzing the CT visuals. The utilization of residual connections in ResNet50 aids in capturing intricate designs; however, it somewhat lags behind the VGG models concerning accuracy.

4.1. Multi-Model Analysis

This study demonstrates how diagnostic precision can be improved by staging throat cancer using a multi-model analysis approach, wherein CNNs are combined with KNN for decision-level fusion. The multi-model approach integrates the strengths of imaging and textual data analysis, hence leveraging the complementary nature of these data modalities.

4.1.1. Overview of the Multi-Model Approach

The CNN models used in this work are VGG16, VGG19, and ResNet50, which have been employed to extract meaningful features from the CT image data. These models have been chosen because of their proven effectiveness in image classification tasks and their capability to capture spatial and hierarchical features from medical images. In parallel, the associated clinical textual data was analyzed using the K-Nearest Neighbors algorithm, which excels at categorizing text-based data into relevant diagnostic categories by identifying similarities in feature space.

4.1.2. Rationale for Multi-Model Integration

The multi-model analysis was thus designed to incorporate both the visual insights from the CT scans and contextual information from the clinical records. Although CNNs provide a high degree of accuracy with regard to the identification and staging of cancerous growth from the images, integrating textual data serves to contextualize these imaging findings by bringing into play patient-specific medical history, symptoms, and diagnostic notes. It is achieved by the fusion of multimodal, which allows the system to take the strengths of both data types into consideration for better robustness and accuracy of predictions.

4.1.3. Decision-Level Fusion

Decision-level fusion was used to integrate CNN-based text analysis with KNN-based imaging analysis. Based on the outputs from different models, the final binary classification decision is taken that combines the information produced by the two models. An individual CNN produces a possibility score for the cancer stage to be in each category via the CT images; their output is combined to provide the ground truth class. The support vector machine provided a probabilistic type of classification based on knowledge from clinical text data KNN. The fusion technique also used a weighted averaging methodology. This means that during all diagnostic decisions, more confident predictions about the output were assigned far greater importance.

4.1.4. Performance Evaluation

To assess the effectiveness of the multi-model approach, a comparative analysis of individual and combined models was conducted:

- VGG16: Achieved an accuracy of 86.2%.
- VGG19: Achieved an accuracy of 87.5%.
- ResNet50: Achieved an accuracy of 92.3%.
- VGG16+KNN: Achieved an accuracy of 94.5%.
- VGG19+KNN: Outperformed all other combinations with an accuracy of 98.67%.

The superior performance of the VGG19+KNN model demonstrates the value of integrating imaging and textual data in diagnostic processes. The decision-level fusion approach ensures that the final model incorporates complementary insights from both modalities, providing a more holistic understanding of the patient's condition.

4.1.5. Significance of Multi-Model Analysis

It forms the basis of how a new multimodal approach can open or further develop advanced deep learning-based methodologies in imaging and machine learning methodologies with text data. The study shows how combining multimodal data overcomes two important weaknesses with unimodal analysis and, in that way, it gives even more emphasis on the necessity of tapping into multiple sources of data to inform decisions clinically. This will provide a really expandable framework for diagnostic enhancement in various medical applications.

4.2. Ablation Studies

We performed an ablation study to evaluate the contribution of each model component alone and combined it to validate each component's necessity and contribution to the proposed framework. These studies aim to isolate and evaluate each element, namely CNNs, KNN, and the decision-level fusion mechanism, to confirm their importance in the framework.

4.2.1. Experimental Setup:

(1) Baseline Models:

• Each CNN model (VGG16, VGG19, and ResNet50) was trained and tested individually using only the CT image data without the integration of textual data.

• The KNN model was trained and tested separately using only the clinical text data.

• Full Framework:

• The complete proposed framework (VGG16 + KNN with decision-level fusion) and (VGG19 + KNN with decision-level fusion) were evaluated to compare its performance against the baseline and intermediate configurations.

4.2.2. Results of Ablation Studies:

(1) CNN Models Alone:

• VGG16: 86.2% accuracy

• VGG19: 87.5% accuracy

• ResNet50: 92.3% accuracy.□□ These results demonstrate that individual CNNs, while effective, lack the contextual understanding provided by clinical text data.

(2) KNN Alone:

• Clinical text data analyzed using KNN achieved an accuracy of 75.3%, highlighting its limitation in capturing the structural complexity of cancer staging solely from textual information.

(3) Full Framework (VGG16 + KNN):

• The complete framework achieved the highest accuracy of 94.5%, demonstrating the necessity of combining image and text data with an effective fusion strategy.

(4) Full Framework (VGG19 + KNN):

• The complete framework achieved the highest accuracy of 98.67%, demonstrating the necessity of combining image and text data with an effective fusion strategy.

4.2.3. The Ablation Studies Confirm That

(1) The combination of *VGG16 and VGG19* with KNN significantly outperforms their standalone counterparts.

(2) The full framework achieves superior performance, validating the necessity of the proposed approach for throat cancer staging.

4.3. Computational Costs and Practical Implementation Challenges

4.3.1. Computational Costs:

(1) Training Phase:

• The CNN models (VGG16, VGG19, ResNet50) used in this study are computationally intensive due to the large number of parameters and the complexity of feature extraction from high-resolution CT images.

• On average, training each CNN model required approximately 12–18 hours on a high-performance GPU (NVIDIA Tesla V100) for the dataset of 650 CT scans. The computational burden was significantly reduced by employing transfer learning, which allowed fine-tuning of pretrained models instead of training full CNN architectures from scratch.

(2) Inference Phase:

• The prediction phase, that is, the actual inference, is far less demanding and requires about 20-50 milliseconds per CT scan on the same hardware. This easily extends real-time diagnosis to even clinically moderate computational infrastructure.

(3) KNN Integration:

• Most of the computational cost of the KNN algorithm depends on the size of the textual dataset. In this study, preprocessing and classification of textual data were completed in less than 10 milliseconds per record on a CPU, hence computationally efficient.

(4) Fusion Mechanism:

• The decision-level fusion adds minimal overhead to the overall system, requiring negligible computational time to combine the outputs of the CNN and KNN models.

4.3.2. Practical Implementation Challenges:

(1) Hardware Requirements:

• Whereas the clinical use of deep learning models such as VGG19 requires either access to high-performance GPUs or cloud-based computational resources, neither is always available in many hospitals, especially in resource-poor settings.

(2) Integration with Clinical Workflow:

• The integration of the proposed framework into existing clinical workflows presents challenges related to compatibility with hospital information systems (HIS) and picture archiving and communication systems (PACS).

• Seamless integration would require custom APIs or middleware to enable data exchange between the AI system and clinical databases.

(3) Data Privacy and Security:

• Handling sensitive medical data for training and inference necessitates stringent data privacy measures to comply with regulations such as HIPAA and GDPR.

• Encryption and secure storage solutions are essential to ensure patient confidentiality during model deployment.

(4) Generalizability Across Institutions:

• Variations in imaging protocols, scanner types, and clinical documentation across institutions may affect the model's performance. To address this challenge, standardizing data preprocessing steps is essential.

(5) Interpretability:

• Most of the deep learning models are black boxes, and clinicians might find interpreting the results difficult. The integration of explainable AI techniques would enhance trust in clinical practice.

CONCLUSION

This investigation illustrates the noteworthy capability of gaining profound knowledge of designs in expanding the exactness of throat disease arranging through the investigation of CT images and joining with restorative records. The relative assessment of VGG19, VGG16, and ResNet50 models, combined with K-Nearest Neighbors (KNN) for true information, uncovers that VGG19+KNN accomplishes the most astounding precision of 98.67%, showing transcending execution in anticipating disease levels. VGG16+KNN takes after with an exactness of 94.5%, mirroring its solid yet somewhat less exact limit at VGG19. ResNet50, while strong with an exactness of 92.3%, demonstrates a somewhat brought-down execution correlation with the VGG models. The perplexity lattices additionally illuminate the models' qualities and shortcomings, featuring VGG19’s exceptional precision and recollect, with some sentences longer than others. All in all, the coordination of multimodal information joining propelled CNN structures with content-based KNN investigation proves to be a ground-breaking way to deal with, fundamentally improving analytic exactness and offering important experiences for restorative applications in throat disease administration.

LIMITATION AND FUTURE WORK

We are aware that the results given in this paper are obtained based on a single dataset of 650 CT scans and may be difficult to generalize. Although this dataset was prepared very carefully for the representation of all stages of throat cancer, we admit that more datasets would be needed in order to validate the strength and applicability of the proposed approach within larger populations and different clinical conditions.

To address this limitation, we plan to incorporate the following in future research:

(1) External Validation:

• Apply the trained models to independent, publicly available datasets to assess their performance in different clinical environments.

• Perform cross-institutional studies to evaluate the generalizability across diverse patient demographics and imaging equipment.

(2) Augmentation of Dataset:

• Increase dataset size through collaboration with multiple hospitals and research institutions.

• Incorporate data from different imaging modalities (e.g., MRI or PET scans) and diverse medical records to test the adaptability of the model.

(3) Domain Adaptation Techniques:

• Implement domain adaptation methods to train the model on datasets with different characteristics (e.g., resolution, noise levels) while preserving diagnostic accuracy.

(4) Generalization Metrics:

• Evaluate the model's performance using statistical measures of generalization, such as leave-one-institution-out validation or bootstrap sampling.

We believe these steps will confirm the wider applicability of the proposed multi-model system and further validate its utility in clinical decision-making on throat cancer staging.

AUTHORS' CONTRIBUTION

It is hereby acknowledged that all authors have accepted responsibility for the manuscript's content and consented to its submission. They have meticulously reviewed all results and unanimously approved the final version of the manuscript.

LIST OF ABBREVIATIONS

HIS	=	Hospital information systems
PACS	=	Picture archiving and communication systems
CNN	=	Convolutional neural networks
KNN	=	K-Nearest Neighbors

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The authors confirm that the data supporting the findings of this study is available within manuscript.

FUNDING

None

CONFLICT OF INTEREST

Dr. Vinayakumar Ravi is the Associate Editorial Board Member of the journal The Open Bioinformatics Journal.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Dikshit P, Dey B, Shukla A, Singh A, Chadha T, Sehgal VK. Prediction of Breast Cancer using Machine Learning Techniques Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing (IC3-2022) Association for Computing Machinery. New York, NY, USA, 24 Oct 2022, pp. 382–387. http://dx.doi.org/10.1145/3549206.3549274

[2] Ghosh S, Dhar S, Kumar A, Jana ND. Sciencedirect melanoma skin skin cancer cancer detection detection using using ensemble ensemble of of machine machine melanoma learning models models considering considering deep deep feature feature embeddings embeddings learning. Procedia Comput Sci 2024; 235: 3007-15. http://dx.doi.org/10.1016/j.procs.2024.04.284

[3] Lei CS, Hou YC, Pai MH, Lin MT, Yeh SL. Effects of quercetin combined with anticancer drugs on metastasis-associated factors of gastric cancer cells: *In vitro* and *in vivo* studies. J Nutr Biochem 2018; 51: 105-13. http://dx.doi.org/10.1016/j.jnutbio.2017.09.011 PMID: 29125991

[4] Zhao Y, Li X, Zhou C, *et al*. A review of cancer data fusion methods based on deep learning. Inf Fusion 2024; 108(24): 102361. http://dx.doi.org/10.1016/j.inffus.2024.102361

[5] Boeckmann L, Berner J, Kordt M, *et al*. Synergistic effect of cold gas plasma and experimental drug exposure exhibits skin cancer toxicity *in vitro* and *in vivo*. J Adv Res 2023; 57: 181-96. http://dx.doi.org/10.1016/j.jare.2023.06.014 PMID: 37391038

[6] Torres MP, Rachagani S, Purohit V, *et al*. Graviola: A novel promising natural-derived drug that inhibits tumorigenicity and metastasis of pancreatic cancer cells *in vitro* and *in vivo* through altering cell metabolism. Cancer Lett 2012; 323(1): 29-40. http://dx.doi.org/10.1016/j.canlet.2012.03.031 PMID: 22475682

[7] Sami H, Sagheer M, Riaz K, Mehmood MQ, Zubair M. Machine Learning-Based Approaches for Breast Cancer Detection in Microwave Imaging 2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium). Singapore, Singapore, 04-10 Dec 2021, pp. 72-73. http://dx.doi.org/10.23919/USNC-URSI51813.2021.9703518

[8] Bao G, Xu R, Wang X, *et al*. Identification of lncRNA signature associated with pan-cancer prognosis. IEEE J Biomed Health Inform 2021; 25(6): 2317-28. http://dx.doi.org/10.1109/JBHI.2020.3027680 PMID: 32991297

[9] Maurya S, Tiwari S, Mothukuri MC, Tangeda CM, Nandigam RNS, Addagiri DC. A review on recent developments in cancer detection using machine learning and deep learning models. Biomed Signal Process Control 2023; 80(P2): 104398. http://dx.doi.org/10.1016/j.bspc.2022.104398

[10] Abdelaziz Ismael SA, Mohammed A, Hefny H. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. Artif Intell Med 2020; 102: 101779. http://dx.doi.org/10.1016/j.artmed.2019.101779 PMID: 31980109

[11] Cheng AS, Guan Q, Su Y, Zhou P, Zeng Y. Integration of machine learning and blockchain technology in the healthcare field: A literature review and implications for cancer care. Asia Pac J Oncol Nurs 2021; 8(6): 720-4. http://dx.doi.org/10.4103/apjon.apjon-2140 PMID: 34790856

[12] Lai KC, Kuo CL, Ho HC, *et al*. Diallyl sulfide, diallyl disulfide and diallyl trisulfide affect drug resistant gene expression in colo 205 human colon cancer cells *in vitro* and *in vivo*. Phytomedicine 2012; 19(7): 625-30. http://dx.doi.org/10.1016/j.phymed.2012.02.004 PMID: 22397993

[13] Jha A, Verma G, Khan Y, Mehmood Q, Rebholz-Schuhmann D, Sahay R. Deep Convolution Neural Network Model to Predict Relapse in Breast Cancer 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). Orlando, FL, USA, 17-20 Dec 2018, pp. 351-358. http://dx.doi.org/10.1109/ICMLA.2018.00059

[14] Wang YR, Yang SY, Chen GX, Wei P. Barbaloin loaded polydopamine-polylactide-TPGS (PLA-TPGS) nanoparticles against gastric cancer as a targeted drug delivery system: Studies *in vitro* and *in vivo*. Biochem Biophys Res Commun 2018; 499(1): 8-16. http://dx.doi.org/10.1016/j.bbrc.2018.03.069 PMID: 29534962

[15] van Tienderen GS, Conboy J, Muntz I, *et al*. Tumor decellularization reveals proteomic and mechanical characteristics of the extracellular matrix of primary liver cancer. Biomater Adv 2023; 146: 213289. http://dx.doi.org/10.1016/j.bioadv.2023.213289 PMID: 36724550

[16] Wu X, Yin C, Ma J, *et al*. Polyoxypregnanes as safe, potent, and specific ABCB1-inhibitory pro-drugs to overcome multidrug resistance in cancer chemotherapy *in vitro* and *in vivo*. Acta Pharm Sin B 2021; 11(7): 1885-902. http://dx.doi.org/10.1016/j.apsb.2020.12.021 PMID: 34386326

[17] Jyoti K, Bhatia RK, Martis EAF, *et al*. Soluble curcumin amalgamated chitosan microspheres augmented drug delivery and cytotoxicity in colon cancer cells: *In vitro* and *in vivo* study. Colloids Surf B Biointerfaces 2016; 148: 674-83. http://dx.doi.org/10.1016/j.colsurfb.2016.09.044 PMID: 27701049