IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511153

Deep Learning Based Detection and Invasion-Oriented Stage Classification of Colorectal Cancer

Sharath Kumar B[1], Shrthika A[2], Sinchana[3], Sudeep P[4], Dr. Pramod Kumar P M[5]

Students, Department of Computer Science and Engineering, Vivekananda College of Engineering and Technology, Puttur, Karnataka, India[1-4]

Associate Professor (Guide), Department of Computer Science and Engineering, Vivekananda College of Engineering and Technology, Puttur, Karnataka, India[5]

Abstract: Colorectal cancer prognosis depends on how deeply the tumour invades surrounding tissue, yet invasion-oriented staging has received little attention in deep learning studies. This paper presents a two-stage framework for histopathological patches and whole slide images. A binary classifier first separates normal from tumour tissue, and tumour patches are then assigned one of four invasion-oriented stages, Stage 0 to Stage 3. Both models use a hybrid EfficientNet-B0 and Transformer network with cross attention fusion, trained on STARC-9 patches with Macenko stain normalisation. On held-out patches, the binary and stage models reached accuracies of 99.64 and 99.70 percent. Patch predictions from TCGA-COAD and TCGA-READ slides are aggregated into slide reports; on 160 labelled slides, slide-level stage accuracy was 83.75 percent, showing a clear gap from patch-level results. A Flask web interface exposes the pipeline.

Keywords: colorectal cancer; histopathology; whole slide image; EfficientNet; Transformer; cross attention fusion; invasion staging; deep learning

I. INTRODUCTION

Colorectal cancer (CRC) is one of the major cancers of the human digestive system. Treatment planning and prognosis depend strongly on how deeply the tumour has invaded the surrounding tissue, so accurate detection and staging are essential. Histopathological examination remains the reference method. A pathologist studies cellular morphology, glandular organisation, nuclear abnormality and the relationship between tumour and surrounding layers. This process is slow, requires expertise, and can vary between observers.

Deep learning offers a way to support this workflow. Convolutional neural networks (CNNs) learn discriminative tissue features directly from images, and attention based models can relate distant regions of a slide. A whole slide image (WSI) may contain thousands of tissue regions, so a practical system must tile the slide, classify the tiles and aggregate the results.

This work proposes such a system. A tissue patch is first classified as normal or tumour. Only tumour patches reach the stage classifier, which assigns Stage 0 to Stage 3. The patch predictions are then combined into a slide level summary. The main contributions are: (1) a hybrid EfficientNet-B0 and Transformer architecture with cross attention fusion used for both tasks; (2) a two stage pipeline that avoids staging normal tissue; (3) a slide level aggregation step that reports tumour fraction, stage distribution and the deepest identified stage; and (4) a slide level evaluation that reports the gap between patch level and whole slide performance.

II. LITERATURE REVIEW

A. CNN Based Colorectal Cancer Detection

Deep learning has been applied to CRC histopathology in several studies. Davri et al. [1] reviewed CNN based CRC diagnosis and pointed out that invasion based staging has received little attention. Transfer learning with VGG and ResNet models has been shown to help when labelled histology data is scarce [2]. Lightweight CNNs [3], a hybrid CNN and ranking approach [4] and ResNet variants [5] have also been used for colorectal cancer detection, and Tsai and Tao [6] studied deep learning for classifying colorectal cancer tissue.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511153

B. Datasets and Tissue Taxonomies

Kather et al. [7] set up a multi class CRC tissue taxonomy based on texture features, and Subramanian et al. [8] released the STARC-9 dataset with nine tissue classes. Neither offers staging labels directly.

C. Whole Slide, Transformer and Stain Normalisation Approaches

For whole slide analysis, deep learning has been used to classify lung cancer histopathology images and predict mutations [9], and weakly supervised multiple instance learning makes it possible to use slide level labels without pixel annotations [10]. Transformer based aggregation such as TransMIL captures long range relationships between patches [11], kernel attention transformers steer the model toward diagnostically relevant regions [12], and 2DMamba applies a state space model to giga-pixel whole slide image classification [13]. Stain normalisation is a standard preprocessing step, since stain variation between laboratories changes how tissue looks without changing its morphology [14]. Residual networks [15] remain a common CNN baseline in this area, while EfficientNet

[16] and Transformer self attention [17] provide the two feature branches of our model.

The reviewed work leaves three gaps that motivate this paper: few studies combine tumour detection with invasion oriented staging, many models lose spatial context at the patch level, and few report how patch level results carry over to whole slide predictions.

III. METHODS AND MATERIALS

A. STARC-9 and TCGA Data

The binary and stage models are trained and evaluated only on image patches from the STARC-9 dataset [8]. For the binary model the target is Normal (NOR, class 0) versus Tumour (TUM, class 1). For the stage model the four tissue folders, mucus, stroma, muscle and adipose tissue, are labelled Stage 0, Stage 1, Stage 2 and Stage 3 respectively. These tissue types act as a proxy for invasion depth, moving from the mucosa through the submucosa and muscularis to the serosal fat, rather than as formal tumour, node and metastasis staging. Whole slide images from The Cancer Genome Atlas colon adenocarcinoma (TCGA-COAD) and rectum adenocarcinoma (TCGA-READ) cohorts are used only for the final end-to-end prediction and are never used for training or validation.

B. Data Split

To limit leakage between partitions, the split code groups patches by a slide identifier recovered from the file name, either a TCGA barcode or a slide, x and y coordinate pattern, and assigns whole groups to a single partition. Grouping is applied separately within each class. STARC-9 file names carry no recoverable slide or patient identifier, so for the patch data the split is per image. The partitions hold about 60 percent of the data for training, 15 percent for validation and 25 percent for testing. The validation set drives early stopping and checkpoint selection, and the test set is not used until the final evaluation. Class weights inversely proportional to class frequency are computed from the training partition to counter class imbalance.

C. WSI Tiling and Tissue Filtering

Each slide is opened with OpenSlide, and the pyramid level whose magnification is closest to 20 times is selected using the objective power stored in the slide metadata, with 40 times assumed when it is missing. That level is cut into non overlapping tiles of 256 by 256 pixels. A tissue mask is obtained by Otsu thresholding of the saturation channel in HSV colour space, and a tile is kept only if at least 15 percent of its pixels are tissue. Blank and background tiles are discarded, and the kept tiles are saved as PNG files named by slide identifier and tile coordinates.

D. Macenko Stain Normalisation

Colour differences between STARC-9 and TCGA material are reduced with the Macenko method [14], implemented from first principles. Each tile is converted to optical density, and pixels with density below 0.15 are treated as background and excluded. The two principal directions of the remaining values are found by eigen decomposition of their covariance matrix, and the 1st and 99th percentile angles in that plane give the haematoxylin and eosin stain vectors. Stain concentrations are solved by least squares and rescaled so that their 99th percentile matches that of a fixed reference tile, and the tile is rebuilt using the reference stain vectors. The reference is fitted once from a single well stained tile, and if normalisation fails for a tile the original tile is used.

E. Resizing, Normalisation and Augmentation

Every patch is resized to 224 by 224 pixels, so the 256 pixel tiles cut from slides are slightly reduced, and its channels are normalised with the ImageNet mean and standard deviation because the CNN branch starts from ImageNet weights.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.1511153

Training patches are augmented with horizontal and vertical flips, 90 degree rotations, small affine changes (scale 0.9 to 1.1, translation up to 5 percent and rotation within 15 degrees), elastic deformation, colour jitter (brightness and contrast 0.15, saturation 0.1, hue 0.02), Gaussian blur and Gaussian noise. Validation, test and slide inference use resizing and normalisation only. Table I lists the main dataset and preprocessing parameters.

TABLE I KEY DATASET AND PREPROCESSING PARAMETERS

Parameter	Value
Training data	STARC-9 patches [8], used for training, validation and testing
Binary classes	NOR (0) and TUM (1)
Stage classes	Stage 0 to Stage 3 (mucus, stroma, muscle, adipose tissue)
Slide source	TCGA-COAD and TCGA-READ (.svs), used only for final prediction
Tile size and stride	256 by 256 pixels, non overlapping
Magnification	20 times
Model input	224 by 224 by 3, ImageNet mean and standard deviation
Data split	60, 15 and 25 percent for training, validation and testing, per image for STARC-9

F. Hybrid CNN Transformer Architecture

Each patch is processed by two branches. The EfficientNet-B0 branch [16], initialised with ImageNet weights, extracts fine grained local features such as cellular morphology and texture. The Transformer branch splits the patch into tokens and applies a four layer encoder with self attention [17] to learn relationships between tissue regions. The two feature sets are combined with a cross attention fusion module, so that CNN features and Transformer features can attend to each other before classification. The fused representation passes through fully connected layers to produce logits, and the class with the highest probability is selected.

The same architecture is used for both tasks. Only the final classification head differs: two outputs for Normal versus Tumour, and four outputs for Stage 0 to Stage 3. In the implementation, the CNN branch projects the 1280 channel EfficientNet-B0 feature map to 256 dimensions, the Transformer divides each 224 pixel patch into 16 by 16 tokens (196 tokens) and uses four encoder layers with four attention heads, and the fused 256 dimensional vector feeds a classifier with a 128 unit hidden layer.

G. Cross Attention Fusion

The two branches see the same normalised patch but produce different token sets. At a 224 pixel input the CNN branch yields a 7 by 7 grid, which is 49 tokens of 256 dimensions, each describing a local neighbourhood. The Transformer branch adds learned position embeddings to its patch tokens, and its encoder layers use GELU activation, a feed forward width of 512 and a dropout of 0.1. The fusion module contains two multi head attention blocks of four heads each. In the first, CNN tokens act as queries over the Transformer tokens, which injects global tissue context into the local features. In the second, Transformer tokens query the CNN tokens, which passes fine local detail back to the global tokens. Each block has a residual connection and layer normalisation.

H. Two Stage Pipeline and Slide Aggregation

The binary model is applied to every tissue patch of a slide. Patches labelled Normal are recorded as normal. Patches labelled Tumour are forwarded to the stage model. A stage prediction is accepted only if its softmax probability is at least 0.55; otherwise the patch is marked unidentified. The slide report contains the number of normal and tumour patches, the tumour fraction and the stage distribution. The deepest invasion stage is the deepest stage (Stage 3, then Stage 2, Stage 1 and Stage 0) whose patch count exceeds 2 percent of the tumour patches, with a minimum of 20 patches. If more than 5 percent of the tumour patches are unidentified, the result is flagged as the stage and beyond, and the report recommends review by a pathologist. Figure 1 illustrates the workflow.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511153

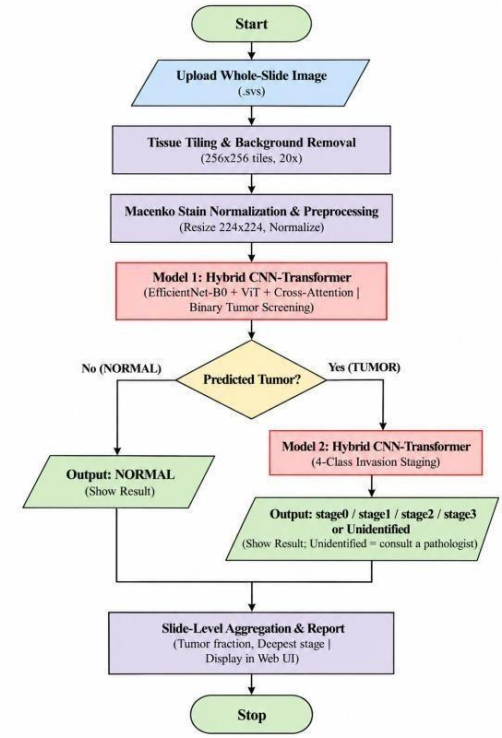


Fig. 1 Flowchart of the proposed two stage system.

I. Confidence Based Rejection

Unidentified is not a trained fifth class, because no labelled set of tumour patches that belong to none of the four tissue types is available. It is a rejection rule applied to the softmax output of the stage model. A prediction is rejected when the highest class probability is below 0.55. The normalised entropy of the distribution, scaled to the range 0 to 1, is also computed and stored with each patch record, with a limit of 1.0 configured, so in the present configuration the confidence threshold is the operative test and the entropy limit can be tightened later. Rejected patches are counted separately in the stage distribution, so uncertain regions are reported and not forced into one of the four stages.

J. Training Configuration

Table II lists the training configuration. Both models were trained with the AdamW optimizer, a learning rate of 0.0003, dropout of 0.3, weight decay of 0.0001 and a batch size of 32. These values are the validated default configuration used for the saved models. The loss is a focal loss with gamma 2 and label smoothing of 0.05, weighted by inverse class frequency to handle class imbalance. The learning rate follows a cosine annealing schedule over up to 40 epochs, training stops early if the validation macro F1 does not improve for seven epochs, and the checkpoint with the highest validation macro F1 is kept. An Optuna based search is included in the training code, but the final models were not trained from a completed search, so these values are reported as the configuration used rather than as optimized values.

The training process aims to achieve good performance while avoiding overfitting. Dropout and weight decay help the models perform well on unseen images, focal loss and class weighting help the models learn from difficult samples and classes with fewer images, and early stopping prevents unnecessary training. The checkpoint with the best validation score is saved for further evaluation.

TABLE II TRAINING CONFIGURATION OF THE HYBRID MODEL

Parameter	Value
CNN branch	EfficientNet-B0, ImageNet weights
Transformer branch	4 encoder layers, 4 heads, 16 by 16 patches (196 tokens)
Fused feature	256, classifier hidden layer of 128 units

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511153

size	
Input size	224 by 224 pixels
Optimizer	AdamW
Learning rate	0.0003, cosine annealing over up to 40 epochs
Weight decay	0.0001
Dropout	0.3
Batch size	32
Loss	Focal loss (gamma 2), label smoothing 0.05, inverse frequency class weights
Early stopping	Patience of 7 epochs on validation macro F1

K. Hyperparameter Search

The training code includes an Optuna study that maximises validation macro F1. Each trial samples a learning rate between 0.00001 and 0.001 and a weight decay between 0.000001 and 0.01, both on a logarithmic scale, a dropout between 0.1 and 0.5, and a batch size of 16, 32 or 64. A trial is trained for at most 15 epochs, a median pruner stops trials that fall behind earlier ones, and 25 trials are configured. The best parameters would then be used for a full training run of up to 40 epochs. As noted above, the reported models use the default configuration and not a completed search.

L. System Implementation and Web Application

The system is implemented in Python with PyTorch. OpenSlide reads WSI formats such as SVS, and the web application accepts .svs, .tif, .tiff and .ndpi uploads. OpenCV and NumPy handle image operations, and scikit-learn supports data splitting and evaluation. A Flask backend loads the saved binary and stage checkpoints and exposes prediction endpoints. The code base is organised into a model module, utility modules for tiling, stain normalisation and data loading, and separate scripts for training, evaluation and slide inference, all reading one configuration file for paths, thresholds and hyperparameters. Because the stage model runs only on patches that the binary model labels as tumour, normal regions never reach the second network, which reduces computation on slides with large normal areas. Each slide produces a JSON report that stores, for every patch, the binary label and confidence and, for tumour patches, the stage label, confidence and entropy, so any slide level call can be traced back to the patches behind it.

The trained models are integrated into one Flask web application with an HTML, CSS and JavaScript front end. A user uploads a slide, which is analysed in a background job while the page shows the current phase, the number of patches classified and a progress grid. The result view then shows the slide level call, the deepest invasion stage, the tumour fraction and the stage distribution. When no tumour is found, the deepest stage is shown as none, and when any tumour patch is unidentified the recommendation advises consulting a pathologist, which makes the decision path auditable.

IV. RESULTS AND DISCUSSION

A. Evaluation Protocol

Each model is evaluated on its own held out test partition, which is not used for training, early stopping or checkpoint selection; the checkpoint is chosen only by validation macro F1. Patch level metrics are accuracy, per class precision, recall and F1 score, confusion matrices, and one versus rest ROC curves with the area under the curve (AUC). The slide level evaluation compares the deepest identified stage of each slide with its known stage and reports a confusion matrix and per stage accuracy.

B. Binary Model on Patch Data

The binary model was tested on 2500 patches, 1250 per class, and Fig. 2 shows its confusion matrix. Of the normal patches, 1249 were classified correctly and 1 was classified as tumour. Of the tumour patches, 1242 were classified correctly and 8 were classified as normal. The overall accuracy is 99.64 percent (2491 of 2500). Precision is 99.36 percent for normal and 99.92 percent for tumour, and recall is 99.92 percent for normal and 99.36 percent for tumour, giving F1 scores of 99.66 percent and 99.64 percent. The ROC curves for both classes have an AUC of 1.000, as Fig. 3 shows. Although overall accuracy is high, the tumour class has more missed detections (8) than false alarms (1), which matters clinically because a missed tumour is the more serious error.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511153

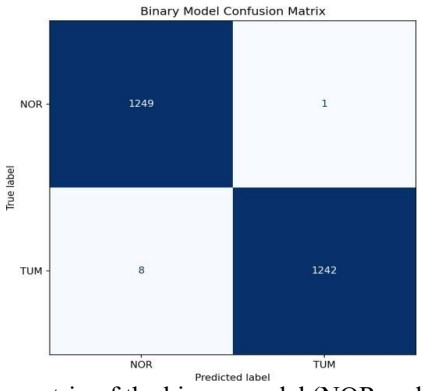


Fig. 2 Confusion matrix of the binary model (NOR and TUM patches).

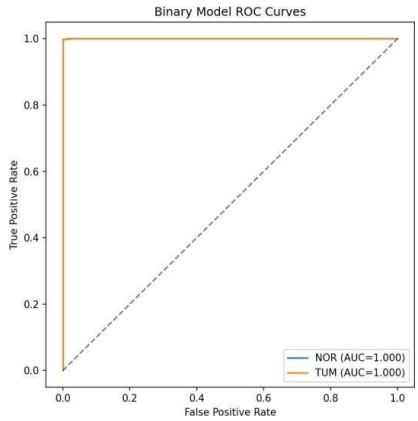


Fig. 3 ROC curves of the binary model.

C. Stage Model on Patch Data

The stage model was tested on 5001 held out patches of the four tissue classes. Fig. 4 shows the confusion matrix and Fig. 5 the ROC curves. Stage 0 (mucus) had 1247 correct out of 1250, Stage 1 (stroma) had 1245 out of 1250, Stage 2 (muscle) had 1243 out of 1250, and Stage 3 (adipose) had 1251 out of 1251. Most errors occurred between neighbouring stages. The overall accuracy is 99.70 percent, and the one versus rest ROC curves for all four stages have an AUC of 1.000. Muscle (Stage 2) was the hardest class, with 7 of 1250 patches misclassified, 4 as mucus and 3 as stroma, while adipose (Stage 3) had no errors at all. Stroma and muscle were confused with each other most often, with 3 patches mistaken in each direction. In total only 15 of 5001 test patches were misclassified, and the recall of every stage stayed above 99.4 percent, so no single stage dominates the error.

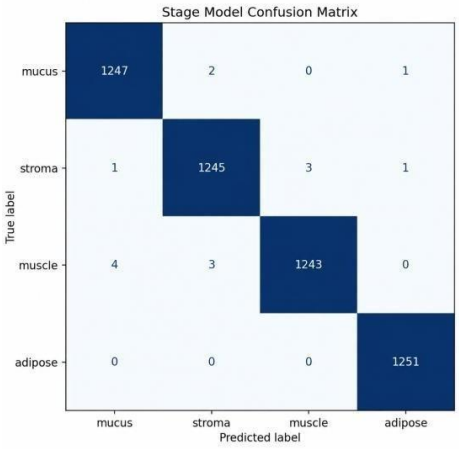


Fig. 4 Confusion matrix of the stage model (Stage 0 mucus, Stage 1 stroma, Stage 2 muscle, Stage 3 adipose).

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511153

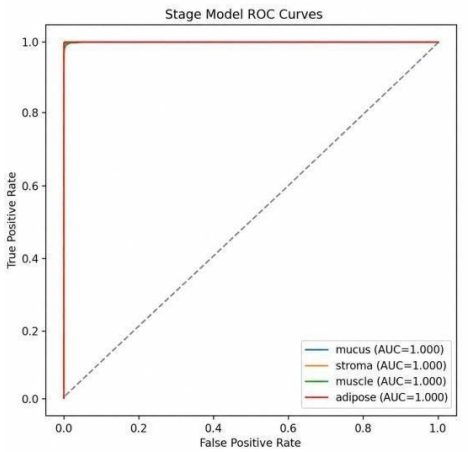


Fig. 5 ROC curves of the stage model for the four classes.

D. Slide Level Evaluation

To evaluate the complete pipeline, the slide level prediction was compared with the known stage of 160 labelled whole slide images from TCGA-READ and TCGA-COAD, with 40 slides per stage. Fig. 6 shows the resulting confusion matrix, and Table III gives the corresponding counts.

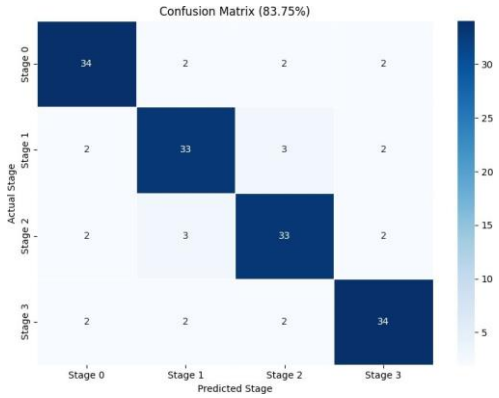


Fig. 6 Final slide level prediction confusion matrix.

TABLE III SLIDE LEVEL CONFUSION MATRIX (ROWS ACTUAL, COLUMNS PREDICTED)

Actual	Stage 0	Stage 1	Stage 2	Stage 3	Total	Correct
Stage 0	34	2	2	2	40	85.0%
Stage 1	2	33	3	2	40	82.5%
Stage 2	2	3	33	2	40	82.5%
Stage 3	2	2	2	34	40	85.0%

Of the 160 slides, 134 were predicted correctly and 26 were misclassified, giving an overall slide level accuracy of 83.75 percent. Because every class contains 40 slides and every predicted column also sums to 40, the per class precision, recall and F1 score all equal the per class accuracy, and the macro averaged values are 83.75 percent. The errors are spread fairly evenly. The largest off diagonal counts, 3 slides, occur between Stage 1 and Stage 2, which are adjacent stages and the most difficult to separate. Averaged over the four stages, the slide level precision, sensitivity and F1 score are each 83.75 percent, and the macro averaged specificity is 94.58 percent.

E. Discussion and Limitations

The patch level models reach near perfect accuracy on held out patches, but the slide level accuracy is 83.75 percent. This gap is expected. A slide contains many patches, including background, stroma, normal tissue and tumour regions that were not seen in the same proportions during training. Errors in individual patches, unidentified regions and stain variation can accumulate when patch predictions are aggregated. The slide level result should therefore be read as a

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.1511153

more realistic estimate of end use performance than the patch level figures. Improving the aggregation rule, increasing the diversity of slide data, and evaluating on more slides from multiple centres are the most direct ways to reduce this gap.

The class balance supports the reported figures. The binary test set holds 1250 patches per class, so a classifier that always predicted one class would score 50 percent, and the slide set holds 40 slides per stage, so chance level is 25 percent. The slide level accuracy is therefore well above chance even though it is lower than the patch level result.

Several limits apply. The models are trained on STARC-9 patches but applied to TCGA slides, so scanner and staining differences may remain even after Macenko normalisation. Patches from one slide are highly correlated, and the STARC-9 split is per image and not per patient, so the patch level figures are likely optimistic. The slide set comes from a single public source, and there is no external or prospective validation. The stage model recognises tissue type, so a stage label shows which tissue a patch flagged as tumour resembles and does not measure invasion directly. Unidentified is a rejection rule and not a learned class, and its threshold is a fixed setting, so the confidence scores should not be read as calibrated probabilities. Clinical utility therefore remains unproven, and the system should be used only as decision support.

V. CONCLUSION AND FUTURE SCOPE

This paper presented a two stage hybrid CNN Transformer framework for colorectal cancer detection and invasion oriented stage classification from whole slide images. The binary model reached 99.64 percent accuracy and the stage model reached 99.70 percent accuracy on held out patches. On 160 labelled slides, the complete pipeline reached 83.75 percent slide level accuracy, with errors concentrated between neighbouring stages. The system provides a web interface for slide upload, patch level visualisation and a slide report.

Future work will focus on improving the slide level aggregation rule, evaluating on larger and more diverse multi centre datasets, adding region level heatmaps for explainability, and integrating the system with clinical decision support and electronic health records. Further directions include completing the hyperparameter search, calibrating the confidence threshold on a labelled set of uncertain patches, grouping every split by slide, and compressing the models for deployment on modest hardware.

ACKNOWLEDGMENT

The authors thank their guide, Dr. Pramod Kumar P M, Associate Professor, Department of Computer Science and Engineering, VCET Puttur, for guidance throughout the project.

REFERENCES

[1]. A. Davri, E. Birbas, T. Kanavos, G. Ntritsos, N. Giannakeas, A. T. Tzallas, and A. Batistatou, “Deep learning on histopathological images for colorectal cancer diagnosis: A systematic review,” Diagnostics, vol. 12, no. 4, Art. no. 837, 2022, doi: 10.3390/diagnostics12040837.

[2]. A. Ben Hamida, M. Devanne, J. Weber, C. Truntzer, V. Derangère, F. Ghiringhelli, G. Forestier, and C. Wemmert, “Deep learning for colon cancer histopathological images analysis,” Comput. Biol. Med., vol. 136, Art. no. 104730, 2021, doi: 10.1016/j.compbiomed.2021.104730.

[3]. A. S. Sakr, N. F. Soliman, M. S. Al-Gaashani, P. Pławiak, A. A. Ateya, and M. Hammad, “An efficient deep learning approach for colon cancer detection,” Appl. Sci., vol. 12, no. 17, Art. no. 8450, 2022, doi: 10.3390/app12178450.

[4]. A. Karthikeyan, S. Jothilakshmi, and S. Suthir, “Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm,” Meas. Sens., vol. 31, Art. no. 100976, 2024.

[5]. D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, “Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer,” Procedia Comput. Sci., vol. 179, pp. 423–431, 2021.

[6]. M.-J. Tsai and Y.-H. Tao, “Deep learning techniques for the classification of colorectal cancer tissue,” Electronics, vol. 10, no. 14, Art. no. 1662, 2021, doi: 10.3390/electronics10141662.

[7]. J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, “Multi-class texture analysis in colorectal cancer histology,” Sci. Rep., vol. 6, Art. no. 27988, 2016, doi: 10.1038/srep27988.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.1511153

[8]. B. Subramanian, R. Jeyaraj, M. N. Peterson, T. Guo, N. Shah, C. Langlotz, A. Y. Ng, and J. Shen, "STARC-9: A large-scale dataset for multi-class tissue classification for CRC histopathology," in Advances in Neural Information Processing Systems, vol. 38, 2025.
[9]. N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula,
[10]. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," Nat. Med., vol. 24, no. 10, pp. 1559–1567, 2018, doi: 10.1038/s41591-018-0177-5.
[11]. G. Campanella et al., "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," Nat. Med., vol. 25, no. 8, pp. 1301–1309, 2019, doi: 10.1038/s41591-019-0508-1.
[12]. Z. Shao, H. Bian, Y. Chen, Y. Wang, J. Zhang, and Y. Zhang, "TransMIL: Transformer-based correlated multiple instance learning for whole slide image classification," in Advances in Neural Information Processing Systems, vol. 34, pp. 2136–2147, 2021.
[13]. Y. Zheng, J. Li, J. Shi, F. Xie, J. Huai, M. Cao, and Z. Jiang, "Kernel attention transformer for histopathology whole slide image analysis and assistant cancer diagnosis," IEEE Trans. Med. Imaging, vol. 42, no. 9, pp. 2726–2739, 2023, doi: 10.1109/TMI.2023.3264781.
[14]. J. Zhang, A. T. Nguyen, X. Han, V. Q.-H. Trinh, H. Qin, D. Samaras, and M. S. Hosseini, "2DMamba: Efficient state space model for image representation with applications on giga-pixel whole slide image classification," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2025, pp. 3583–3592, doi:
[15]. 1109/CVPR52734.2025.00339.
[16]. M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in Proc. IEEE Int. Symp. Biomed. Imaging (ISBI), 2009, pp. 1107–1110, doi: 10.1109/ISBI.2009.5193250.
[17]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[18]. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), vol. 97, pp. 6105–6114, 2019.
[19]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.

 This work is licensed under a Creative Commons Attribution 4.0 International License