Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multi-Class Gastric Histopathology Classification Using a Hybrid Swin Transformer–ConvNeXt Model

Usman Nazakat
Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur (22621), Pakistan
Muhammad Saad
Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur (22621), Pakistan
Uzair Khan
Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur (22621), Pakistan
Muhammad Junaid Asif
mjunaid94ee@outlook.com
Artificial Intelligence Technology Centre (AITeC) https://orcid.org/0000-0002-3875-0777
Muhammad Ahsan Aziz
Quaid-i-Azam University
Rana Fayyaz Ahmad
Artificial Intelligence Technology Centre (AITeC), Islamabad 44000, Pakistan

Research Article

Keywords: gastric cancer detection, histopathology, Swin Transformer, ConvNeXt, deep learning, ensemble learning, medical image classification

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11703260/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Multi-Class Gastric Histopathology Classification Using a Hybrid Swin Transformer–ConvNeXt Model

Usman Nazakat[1,2],Muhammad Saad [1,2], Uzair Khan[1,2], Muhammad Junaid Asif [2], Muhammad Ahsan Aziz [3], and Rana Fayyaz Ahmad

[1] School of Computing Sciences, Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur (22621), Pakistan
[2] Artificial Intelligence Technology Centre (AITeC), National Centre for Physics (NCP), Islamabad (44000), Pakistan
[2]Department of Computer Sciences, Quaid-i-Azam University (QAU), Islamabad, Pakistan

***Abstract*—Diagnosing gastric cancer still depends, for the most part, on a pathologist looking at stained tissue slides under a microscope. The disease kills roughly 660,000 people every year, and although early detection makes a real difference to survival, manual slide reading is slow and two experts examining the same slide do not always agree. In this work we built a system that tries to ease that burden. We trained a hybrid deep learning model on the Gastric Cancer Histopathology Tissue Image Dataset (GCHTID), pairing a Swin Transformer, which is good at picking up the broad structural layout of a tissue patch, with a ConvNeXt network, which captures the fine local textures where early malignant changes tend to hide. Every image first passes through a simple preprocessing chain of median filtering and CLAHE contrast enhancement, a combination we settled on after comparing several denoising and enhancement options. The ensemble sorts each patch into one of eight classes: tumor tissue (TUM) plus seven benign types—adipose, debris, lymphocytes, mucus, smooth muscle, normal mucosa, and stroma. On held-out data it reached 75.35% test accuracy (88.76% training, 76.99% best validation), beating the single-model baselines we tried, including ViT and EfficientNet-B3. To make the model usable outside a notebook, we wrapped it in a Flask REST API and a small web interface where a user can upload a slide image, see the prediction and its confidence, download a PDF report, and look back through past results. The system is a research prototype rather than a clinical tool, but the results suggest that combining transformer and convolutional features is a workable direction for computer-assisted pathology.**

***Index Terms*—gastric cancer detection, histopathology, Swin Transformer, ConvNeXt, deep learning, ensemble learning, medical image classification**

I. Introduction

Anyone who has watched a pathologist work through a stack of histopathology slides understands why gastric cancer diagnosis is a bottleneck. Each slide has to be examined carefully, gastric tissue is structurally complicated, and the judgment calls involved mean that two experienced pathologists will sometimes read the same slide differently. Meanwhile the caseload keeps growing: hospitals and laboratories face more samples than their pathologists can comfortably handle, and in many places there simply are not enough trained specialists. With around 660,000 deaths attributed to gastric cancer each year—it remains the fifth deadliest cancer worldwide according to GLOBOCAN estimates [1]—and survival tied closely to how early the disease is caught, delays and inconsistencies in diagnosis have a real human cost.

This is exactly the kind of problem where machine assistance can help. Deep learning has reshaped medical image analysis over the past decade [2], and computational histopathology in particular has grown from a niche topic into an active field with its own surveys and benchmarks [3], [4]. Yet most of the tools we came across during this project stop at a binary question—cancer or not—and say little about what the tissue actually is when it is not cancerous. For a pathologist, that distinction matters. Knowing whether a benign patch is stroma, lymphocytes, or ordinary mucosa provides context for the case as a whole.

So we set out to build something a bit more complete: a system that not only flags tumor tissue (TUM) but also assigns non-cancerous patches to one of seven histological categories—adipose (ADI), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal mucosa (NOR), and stroma (STR). The technical idea at the center of the work is straightforward. Histopathology images carry information at two scales: the overall arrangement of tissue, and the fine cellular texture within it. Transformers are strong at the former, convolutional networks at the latter, so rather than choosing between them we combined a Swin Transformer [17] with a ConvNeXt [18] and let a weighted ensemble merge their predictions.

Along the way we made three other contributions worth noting. We compared several preprocessing options and found that median filtering followed by CLAHE gave the cleanest images and the best downstream accuracy. We benchmarked our ensemble against single models (ViT, EfficientNet-B3, Swin alone) and against other hybrid pairings, which lets us say with some confidence that the ConvNeXt–Swin combination was the best of the options we tested. And we carried the model all the way to a working prototype—a Flask API, a web front end, PDF reports, and a database that keeps a history of predictions—because a model that only lives in a training notebook is of little use to the people it is meant to help.

The rest of the paper follows the usual structure: Section II reviews related work, Section III covers the dataset, preprocessing, and model, Section IV describes the implementation, Section V presents and discusses the results, and Section VI closes with limitations and plans.

II. Related Work

Deep learning has moved quickly from proof-of-concept to near-clinical performance in medical imaging. The survey by

Litjens *et al.* [2] traced this shift across modalities, and later reviews focused specifically on computational histopathology [3], [4] document how the field matured—including a candid discussion of the obstacles between research prototypes and clinical adoption, obstacles we ran into ourselves and return to in Section V. Two results in particular convinced the wider community that histopathology was ready for deep learning. Coudray *et al.* [5] showed that a CNN could not only classify lung cancer subtypes from slide images but even predict certain gene mutations, and Kather *et al.* [6] classified colorectal histology into nine tissue types—a task closely analogous to our eight-class gastric problem—and linked the predictions to patient survival.

Within gastric cancer specifically, the work splits by imaging modality. On the endoscopy side, Hirasawa *et al.* [7] trained a CNN on more than 13,000 endoscopic images and reached 92.2% sensitivity for lesion detection. On the histopathology side—where our work sits—Iizuka *et al.* [8] classified gastric and colonic whole-slide images into adenocarcinoma, adenoma, and non-neoplastic tissue with AUCs up to 0.97, and Song *et al.* [9] built a gastric diagnosis system that held up under clinical evaluation across multiple hospitals, an important demonstration that these methods can operate outside curated benchmarks. Shi *et al.* [10] pushed in a similar direction for gastric dysplasia in endoscopic specimens. Progress here has also depended on data: the public GasHisSDB database [11], with some 245,000 labeled gastric patches, gave the community a common yardstick, and its authors' follow-up comparison [12] benchmarked everything from linear regression to visual transformers on it.

The architectural landscape has shifted underneath this applied work. Residual networks [13] made very deep CNNs trainable and remain the ancestor of most modern backbones, including EfficientNet [14], whose compound scaling gives a strong accuracy-per-parameter baseline we adopt here. The transformer, originally built for language [15], crossed into vision when Dosovitskiy *et al.* [16] showed with ViT that pure attention can match or beat convolutions given enough data. The Swin Transformer [17] made the idea practical for dense, high-resolution inputs by computing attention within shifted local windows, and ConvNeXt [18] answered from the other side, demonstrating that a carefully modernized CNN remains competitive with transformers.

Closest to our own approach, Yong *et al.* [19] showed on GasHisSDB that ensembles of complementary deep models beat their individual members for gastric cancer detection. Their finding matches the conclusion we reached independently: no single architecture sees everything in a tissue image, and combining models with different inductive biases is a cheap way to recover what each one misses. Two further threads from the literature inform our design choices rather than our architecture. Stain normalization [20] addresses the color variability between laboratories that limits generalization—something we identify as future work—and gradient-based explanation methods such as Grad-CAM [21] offer the visual evidence trails that our test users asked for.

III. Methodology

A. Dataset

All experiments use the Gastric Cancer Histopathology Tissue Image Dataset (GCHTID), which collects high-resolution patches across the eight tissue classes listed in the introduction. We picked it for practical reasons: the images resemble what appears in actual diagnostic work, the classes are reasonably balanced—which spares us from heavy resampling tricks—and the eight-way labeling matches the multi-class problem we wanted to solve. We organized the data into class-wise folders and split it 80/10/10 into training, validation, and test sets, checking that no image appears in more than one split.

B. Preprocessing

Histopathology patches are noisy, and staining and illumination vary from image to image, so some cleanup before training pays off. Our pipeline is deliberately simple. A median filter removes salt-and-pepper noise without blurring the tissue boundaries, and CLAHE then lifts the local contrast so that cellular structures stand out. After that, images are resized to 224×224 to fit the pretrained backbones and normalized with ImageNet statistics. During training we also apply rotation, flips, color jitter, and scaling as augmentation.

We did not arrive at median + CLAHE by guesswork. We tried Non-Local Means, median, and bilateral filtering for denoising, crossed with CLAHE and Gaussian enhancement, and judged the combinations both by eye and by classifier performance. Table I summarizes the outcome and Fig. 1 shows what the different combinations look like on sample patches. Gaussian enhancement mostly just blurred the images; the median filter with CLAHE was the clear winner on both criteria.

TABLE I
Comparison of Preprocessing Techniques

Technique	Category	Effect	Performance
NLM	Noise reduction	Preserves details	Moderate
Median filter	Noise reduction	Removes noise effectively	**Best**
Bilateral filter	Noise reduction	Smooths edges	Good
Gaussian	Enhancement	Slight blur	Not effective
CLAHE	Enhancement	Improves contrast	**Best**

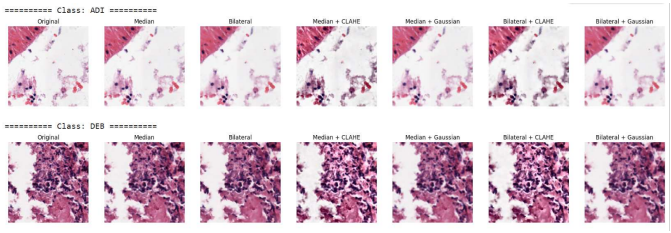


Fig. 1. What the preprocessing options actually do to the images: sample ADI and DEB patches shown original and after each filter/enhancement combination.

C. The Hybrid Model

The model itself is an ensemble of two backbones chosen for their complementary strengths. The Swin Transformer (Swin-Base) [17] handles the global picture: its shifted-window attention builds a hierarchical representation that captures how tissue is organized across the whole patch. ConvNeXt (Base) [18] works the other end, picking up fine textures and local cellular patterns—the sort of subtle detail that distinguishes an early malignant change from its benign neighbors.

Both networks start from ImageNet pretrained weights, with their classification heads replaced to output our eight classes. At inference, we average their class probabilities with weights of 0.7 for Swin and 0.3 for ConvNeXt—values we arrived at empirically—and report the top class along with its confidence. Fig. 2 shows where the model sits inside the larger system, from preprocessing through to the interface and database layers.

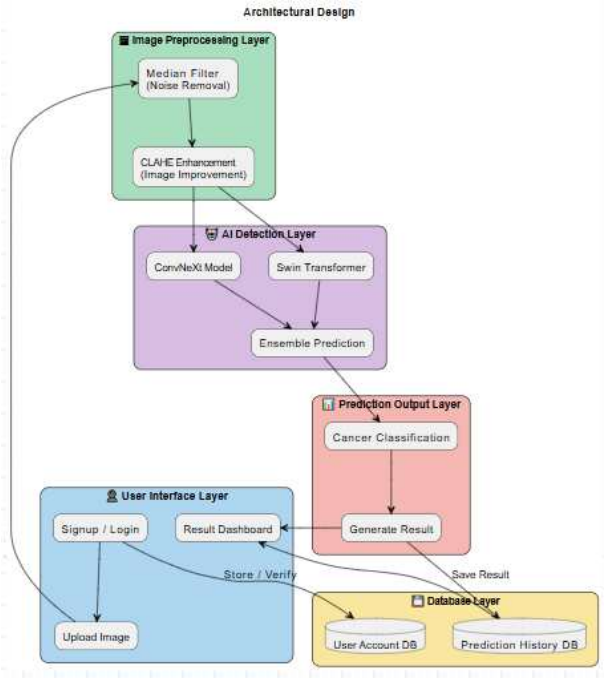


Fig. 2. The system as a whole: preprocessing feeds the two-backbone detection layer, whose ensemble output flows to the prediction layer, the user interface, and the database.

D. Training

Training ran on Kaggle's free T4 GPUs, which shaped several of our choices. Mixed-precision training (AMP) kept iteration times tolerable, and we maintained an exponential moving average (EMA) of the weights, which noticeably steadied validation performance from epoch to epoch. The loss is cross-entropy with label smoothing of 0.1—histopathology classes blur into one another at the boundaries, and we did not want the model to be more confident than the data warrants. We used AdamW with cosine annealing, and gave the ConvNeXt branch a learning rate ten times higher than Swin's (3×10^{-4} versus 3×10^{-5}), since the CNN adapted comfortably at a rate that destabilized the transformer. The full configuration is in Table II.

TABLE II
TRAINING HYPERPARAMETERS

Hyperparameter	Value
Learning rate (Swin)	3×10^{-5}
Learning rate (ConvNeXt)	3×10^{-4}
Batch size	32
Epochs	20
Optimizer	AdamW
Loss function	Cross-entropy (label smoothing $= 0.1$)
Scheduler	Cosine annealing
Ensemble weights	0.7 (Swin) / 0.3 (ConvNeXt)
Precision	Mixed precision (AMP)
Weight averaging	EMA

IV. SYSTEM IMPLEMENTATION

A. From Model to Prototype

A trained checkpoint is not a diagnostic tool, so a good part of the project went into building everything around the model. The stack is unremarkable by design: Python with PyTorch, Torchvision, and the `timm` library on the model side, OpenCV for image handling, Flask for the API, plain HTML/CSS/JavaScript on the front end, and a SQL database (SQLite/MySQL) underneath. We developed in VS Code and Colab, trained on Kaggle, and tested the API with Postman.

The Flask server loads the saved ensemble (`BEST_MODEL.pth`) once at startup and keeps it in memory. A single endpoint, `POST /predict`, accepts an uploaded image, runs it through the same preprocessing used in training, and returns a JSON object with the predicted class, confidence, a plain-language description of what that tissue type means, the cancer status, and a mapped risk level. Round-trip time is about 2–4 seconds per image, depending on the image size and the machine.

On top of the API sits a web interface aimed at people who should never have to see the JSON. A doctor or researcher uploads a slide image, gets the prediction with its confidence and a color-coded status, can download the result as a PDF report, and can browse earlier analyses. Fig. 3 shows the result screen.

B. Data Storage

Persistence is handled by a small relational database, `gastric_db`, with just two tables. `users` holds accounts and supports login and role-based access; `predictions` stores each result—class, confidence, risk level, description, and the image path—linked to its owner through a `user_id` foreign key. One user, many predictions. It is a modest schema, but it gives the prototype something clinical software genuinely needs: a traceable history of what was analyzed and what the system said.

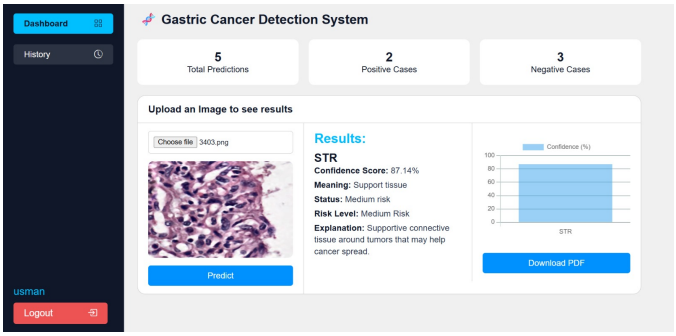


Fig. 3. The result screen of the web interface: uploaded image, predicted class, confidence, and a short diagnostic explanation.

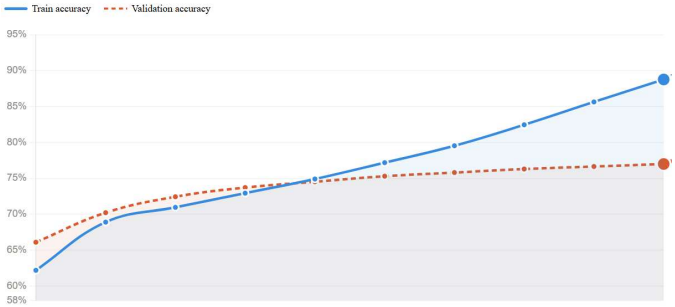


Fig. 4. Training and validation accuracy over the 20 training epochs.

C. How We Tested It

We mixed automated and manual testing. Ten to twelve unit tests cover the pieces that can be checked mechanically: the preprocessing functions, the dataset split (verifying the 80/10/10 partition and the absence of leakage), the model's output format, and the `/predict` endpoint. Integration tests then walked the chain from front end to back end to model to database and back. Finally we ran the whole workflow by hand, including feeding it corrupted files and unsupported formats to confirm it fails gracefully. A small user acceptance round with supervisors and other users found the interface easy to operate; the most common request—one we agree with—was for visual explanations showing *where* in the image the model found evidence for its call.

V. Results and Discussion

A. Overall Performance

On the held-out test set the ensemble reached 75.35% accuracy, with 88.76% on training data and a best validation accuracy of 76.99%. Precision, recall, and F1 all sit close to the accuracy figure (Table III), which tells us the errors are spread across classes rather than concentrated in one or two. The learning curves in Fig. 4 rise smoothly, and the gap that opens between the training and validation curves in later epochs indicates some overfitting—not surprising given how visually intricate histopathology patches are relative to the size of the dataset.

TABLE III
Evaluation Metrics of the Final Hybrid Model

Metric	Value
Training accuracy	88.76%
Validation accuracy (best)	76.99%
Test accuracy	75.35%
Precision	75.85%
Recall	75.35%
F1-score	75.45%
Loss	1.21

B. Was the Hybrid Worth It?

The honest way to justify an ensemble is to show it beats its parts, so we trained the single models and several alternative pairings under the same protocol (Table IV). The single models tell an instructive story. EfficientNet-B3 essentially memorized the training set—99% training accuracy—yet managed only 70% on test data, a textbook case of a local-texture model overfitting. ViT landed at the same 70% test accuracy from the opposite direction, with global features but weak texture sensitivity. Every hybrid we tried beat both, and the ConvNeXt–Swin pairing gave the best validation accuracy of the group, which is why it became the deployed model. Swin + EfficientNet-B3 posted a slightly higher test number (76% versus 75.3%), but its lower training and validation figures made it the less consistent choice overall.

TABLE IV
Model Performance Comparison on GCHTID

Model	Train Acc.	F1 / Val.	Test Acc.
ViT (single)	82.0%	70.0%	70.0%
EfficientNet-B3 (single)	99.0%	71.0%	70.0%
ViT + Swin	85.0%	74.7%	∼73.0%
Swin + EfficientNet-B3	79.4%	75.0%	76.0%
ConvNeXt + Swin (proposed)	**88.7%**	**77.0%**	**75.3%**

C. Where the Model Succeeds and Where It Struggles

The confusion matrix (Fig. 5, 389 test images per class) and the per-class metrics (Fig. 6) show a clear pattern. Adipose, lymphocytes, and smooth muscle are handled well—these tissues simply look distinctive, with fat's characteristic empty vacuoles and lymphocytes' dense dark nuclei. The trouble spots are normal mucosa (NOR) and stroma (STR), which resemble each other and their neighboring classes closely enough that even the ensemble confuses them regularly; they account for the largest share of errors. The tumor class is the encouraging part: TUM recall is the highest of any class, close to 0.80. For a screening-oriented tool that is the right failure mode—occasionally raising a false alarm on benign tissue is far less costly than letting cancer through.

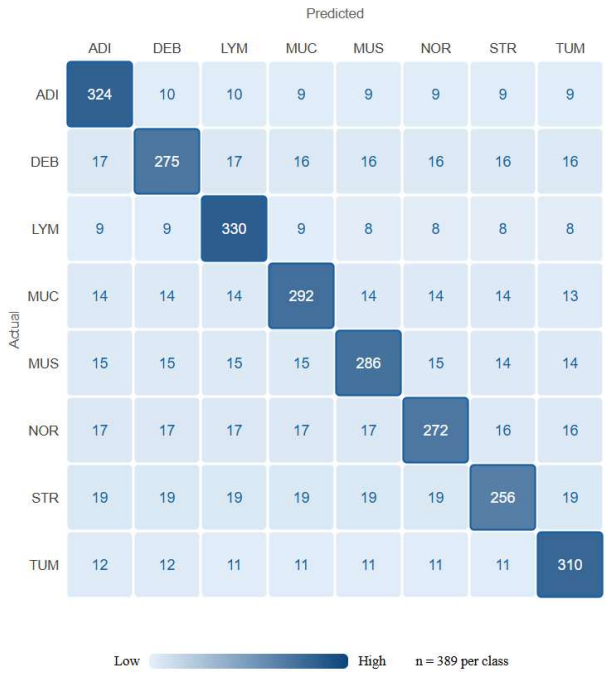


Fig. 5. Confusion matrix on the test set (389 images per class). Most confusion involves the visually similar NOR and STR classes.

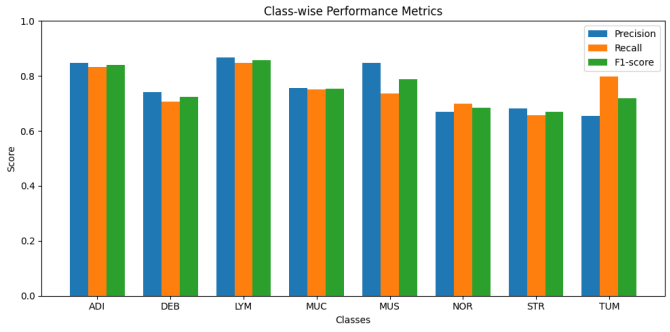


Fig. 6. Per-class precision, recall, and F1-score. Note the high recall on TUM and the dip on NOR and STR.

D. Limitations

We want to be candid about what this system is and is not. The 75% test accuracy is respectable for an eight-way problem on a student budget, but it is well short of clinical grade, and the NOR/STR confusion suggests the ceiling will not rise much without stain normalization or more varied training data. Our compute was limited to free cloud GPUs, so hyperparameter tuning was necessarily shallow—the ensemble weights and learning rates were chosen from a small number of trials, not a systematic search. Everything was trained and tested on one public dataset, so we cannot say how the model behaves on slides from a different scanner, staining protocol, or patient population. Ambiguous cases still need an expert's eyes. And as a prototype, the system carries only basic security—nowhere near what handling real patient data would demand.

VI. Conclusion and Future Work

We built and evaluated a system that classifies gastric histopathology patches into eight tissue types, pairing a Swin Transformer's global view with ConvNeXt's local texture sensitivity in a weighted ensemble. With a median + CLAHE preprocessing front end, the ensemble reached 75.35% test accuracy and outperformed every single-model baseline we tried, and it now runs end to end behind a Flask API and web interface, complete with PDF reports and a stored prediction history. The main lesson of the project matches what others have found [19]: in histopathology, models with different inductive biases genuinely see different things, and combining them is one of the most reliable ways to improve on either alone.

Several directions follow naturally from the limitations above. The first priority is explainability—Grad-CAM [21] or SHAP overlays showing which regions drove a prediction—both because our test users asked for it and because trust in a diagnostic aid depends on it. Training on larger, multi-institutional data with stain normalization [20], and augmentation methods such as MixUp and CutMix, should chip away at the NOR/STR confusion. Beyond the model, cloud deployment would remove the local-hardware requirement, integration with hospital information systems would let the tool meet pathologists inside their existing workflow, and a mobile-friendly interface would help in resource-limited settings. Extending the approach to other cancer types where deep learning has already shown promise, such as colorectal [6] or lung [5], is a longer-term goal. None of these steps makes the system clinical on its own, but together they trace a plausible path from the prototype described here toward a tool a pathologist might actually keep open on a second screen.

Acknowledgment

The authors thank the School of Computing Sciences at Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology for its guidance throughout this work, and the Artificial Intelligence Technology Centre (AITeC) at the National Centre for Physics (NCP) for providing the computational facilities that made the experiments possible.

References

[1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J. Clin.*, vol. 74, no. 3, pp. 229–263, 2024, doi: 10.3322/caac.21834.

[2] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[3] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Med. Image Anal.*, vol. 67, art. 101813, Jan. 2021, doi: 10.1016/j.media.2020.101813.

[4] A. Shmatko, N. Ghaffari Laleh, M. Gerstung, and J. N. Kather, "Artificial intelligence in histopathology: Enhancing cancer research and clinical oncology," *Nature Cancer*, vol. 3, pp. 1026–1038, 2022, doi: 10.1038/s43018-022-00436-4.

[5] N. Coudray *et al.*, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature Med.*, vol. 24, pp. 1559–1567, 2018, doi: 10.1038/s41591-018-0177-5.

[6] J. N. Kather *et al.*, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLoS Med.*, vol. 16, no. 1, art. e1002730, Jan. 2019, doi: 10.1371/journal.pmed.1002730.

[7] T. Hirasawa *et al.*, "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," *Gastric Cancer*, vol. 21, no. 4, pp. 653–660, 2018, doi: 10.1007/s10120-018-0793-2.

[8] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki, "Deep learning models for histopathological classification of gastric and colonic epithelial tumours," *Sci. Rep.*, vol. 10, art. 1504, 2020, doi: 10.1038/s41598-020-58467-9.

[9] Z. Song *et al.*, "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," *Nature Commun.*, vol. 11, no. 4294, 2020, doi: 10.1038/s41467-020-18147-8.

[10] Z. Shi, C. Zhu, Y. Zhang *et al.*, "Deep learning for automatic diagnosis of gastric dysplasia using whole-slide histopathology images in endoscopic specimens," *Gastric Cancer*, vol. 25, pp. 751–760, 2022, doi: 10.1007/s10120-022-01294-w.

[11] W. Hu *et al.*, "GasHisSDB: A new gastric histopathology image dataset for computer aided diagnosis of gastric cancer," *Comput. Biol. Med.*, vol. 142, art. 105207, Mar. 2022, doi: 10.1016/j.compbiomed.2021.105207.

[12] W. Hu *et al.*, "A comparative study of gastric histopathology subsize image classification: From linear regression to visual transformer," *Frontiers Med.*, vol. 9, 2022, doi: 10.3389/fmed.2022.1072109.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[14] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, USA, Jun. 2019, pp. 6105–6114.

[15] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.

[16] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. 9th Int. Conf. Learn. Represent. (ICLR)*, Virtual Event, May 2021.

[17] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 10012–10022.

[18] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 11976–11986.

[19] M. P. Yong, Y. C. Hum, K. W. Lai, Y. L. Lee, C.-H. Goh, W.-S. Yap, and Y. K. Tee, "Histopathological gastric cancer detection on GasHisSDB dataset using deep ensemble learning," *Diagnostics*, vol. 13, no. 10, p. 1793, May 2023, doi: 10.3390/diagnostics13101793.

[20] M. Macenko *et al.*, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, Boston, MA, USA, Jun. 2009, pp. 1107–1110.

[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 618–626.