ISSN: 3107-6513

A Semantic Quality-Gated AI Framework for Reliable Histopathological Diagnosis

Swathi J[1], Monika sree J[2], Harrisha M[3], Bhavadharani[4] E, Karthika D[5]
[1234] *Final-Year Student, Department of Artificial Intelligence and Data Science, Arunai Engineering College, Tamil Nadu, India.*
[5] *Assistant Professor, Department of Artificial Intelligence and Data Science, Arunai Engineering College, Tamil Nadu, India.*
[1]swathij865@gmail.com, [5]karthikadharuman@gmail.com

Abstract
Digital histopathology is gaining importance in computer-aided disease diagnostics. The accuracy of artificial intelligence systems may be lowered due to a variety of problems related to tissue processing, staining, blurriness, folds, dirt, and other imaging artifacts. Poor images may spoil the results of diagnostic modeling. Therefore, this study is focused on the development of a quality-checking system for histopathological research that defines the quality of the slides before the disease is analyzed. To achieve this, a digital image of the prepared tissue slide is taken and analyzed for quality. The approach serves to locate the area of interest using U-Net and to identify artifacts based on ResNet-50 processing. Next, the regions of tissue and artefact are compared to check if the identified artefact has changed the tissue relevant for diagnosis. The outcome of this comparison is the decision whether the slide is accepted or not. The accepted slides then undergo high power analysis or imaging and their staining is normalized by either Macenko or Reinhard approach. The normalized images are processed for preliminary disease detection using ResNet-50 and Vision Transformer. U-Net++ then isolates the area of disease followed by morphological characteristics extraction like tumor size, mitotic index and gland irregularity. The combined features are then processed along with deep features available for severity determination and disease classification using machine learning techniques like Random Forest and XGBoost. The tool is also able to generate report in form of templates and store the analyzed case in cloud storage.

Keywords— *Digital histopathology, Artificial Intelligence, Slide Quality Control, Tissue segmentation, Artefact detection, U-Net, ResNet-50, Vision Transformer, U-Net++, Stain normalization, Disease classification.*

I. INTRODUCTION

Histopathology is an important technique used for diagnosis and categorization of illnesses through microscopic analysis of tissue samples. In the case of conventional diagnosis, pathologists observe stained tissue sections for abnormalities in terms of cellular structures and tissue patterns. The growing volume of pathology data has led to the emergence of the need for computerized methods that can help in the analysis of histopathological images. AI and deep learning are being frequently used in the area of medical imaging. Neural networks have the ability to extract patterns from images and more modern transformerbased models are capable of capturing the relationships between different parts of an image. The application of these technologies is feasible for detection of illnesses and segmentation of tissues. Nonetheless, the accuracy of these technologies is dependent on the quality of the input image. Histopathological images may have tissue folds, blurred images, bubbles of air, residual materials, differences in staining, pen marks, etc. Such artifacts may have an impact on features learned by an AI model and determine its final outcomes. Differences in staining between samples may also lead to the difference in the appearance of the images. Instead of discarding an image solely based on the presence of an artifact, the system first examines the impact of the artifact on the associated tissue region. This enables the quality assessment to take into account the diagnostic significance of the affected region. The proposed framework integrates the processes of tissue segmentation, artifact detection, semantic analysis of overlaps, qualitybased acceptance/rejection, stain normalization, disease detection, disease-region segmentation, feature extraction, severity determination, and classification into one cohesive workflow [11][17].

II. CORE INTENT

A major purpose of the suggested structure is to enhance the dependability of AI-enhanced histopathology assessment by ensuring quality of the input image prior to thorough disease diagnosis. The system begins with isolating the tissue part using U-Net and finding any artifacts by making use of ResNet-50. The identified artifact part is compared to the tissue part to establish whether the artifact interferes with the diagnostically significant tissue [1], [2], [6].

If the image is deemed unacceptable, it is rejected from further inspection. If the image passes the quality gate, it continues to the diagnostic stages. The accepted image is captured at high resolution and processed using stain normalization techniques such as Macenko and Reinhard. The normalized image is then analysed using ResNet-50 and Vision Transformer models for preliminary disease detection. For suspected disease cases, U-Net++ is used to segment the affected region. Morphological measurements such as tumour area percentage, mitotic count, and gland irregularity are extracted from the segmented regions. These features are combined with deep learning features and used for severity assessment and disease classification.The overall intention is to prevent unsuitable

ISSN: 3107-6513

image inputs from directly entering the diagnostic model and to provide a structured pipeline from slide quality assessment to final report generation.

III. SEMANTIC SLIDE QUALITY CONTROL

The quality-control stage is the main decision-making stage of the proposed framework. It includes tissue segmentation, artifact identification, semantic overlap analysis, and quality-based decision-making.

A. Tissue segmentation/ROI segmentation

The process begins with the identification of the tissue of interest in the acquired histopathological image. The U-Net model is employed for isolating the tissue from the background and creating the tissue or ROI mask which conveys information about where the tissue is located and is important for analyzing artifacts. The isolation of tissue from the background helps to treat the whole image in a different way when the image quality assessment is performed [1].

B. Artifact detection

Once the tissue has been located, the image is examined for artifacts. ResNet-50 is used for identification of undesired regions like folds, pieces of debris, areas subjected to blurring, air bubbles, and other artifacts in the image [2], [7]. The detected artifact regions are represented as an artifact mask. This mask is used together with the tissue mask in the next stage.

C. Tissue–Artifact Overlap Analysis

The tissue mask and artifact mask are compared to determine the location and extent of their overlap.

An artifact located outside the diagnostically relevant tissue may not require rejection of the complete image. Another way of looking at this is when an artifact overlaps with a critical area of the tissue, it may interfere with the information necessary for diagnosis.

Thus, the decision logic that I have formulated is:
Artifact outside the relevant tissue → Accept
Artifact overlapping with the relevant tissue → Reject

This gives us the advantage of a quality decision that takes tissue into account, rather than simply the presence or absence of an artifact [7], [8].

D. Quality Decision

After semantic analysis, the information is transmitted to the quality gate. If the image meets the required quality standard, it is classified as Accepted and proceeds to the diagnostic pipeline. If there is some important issue that affects the relevant tissue in the image, it is classified as Rejected. Unfortunately, rejected cases have to be sent for re-preparation, re-capturing of the image or manual review [7], [8].

IV. HIGH-RESOLUTION IMAGING AND STAIN NORMALIZATION

Only accepted cases move to high-resolution microscopy. The goal of this stage is to obtain well-detailed tissue data that will be used for subsequent diagnosis models. Histopathological images may differ in terms of colors and stain intensity due to changes in the staining procedure, laboratory conditions, and equipment.These variations can affect the appearance of tissue structures. To reduce such variations, stain normalization is applied to the accepted high-resolution images [5], [6], [9], [10], [12]. The proposed framework considers two commonly used approaches:

- Macenko normalization
- Reinhard normalization

The normalized image provides a more consistent input for the downstream disease-analysis models. The normalization method can be selected based on experimental performance on the target dataset.

V. DISEASE PRESENCE AND PRELIMINARY ANALYSIS

The normalized histopathological image is examined by means of ResNet-50 and Vision Transformer models. Where ResNet-50 is used to extract visual features from tissue, the Vision Transformer is used to evaluate spatial relations between different parts of the image. At this stage, the output is a preliminary disease evaluation. The image is categorized as follows:

- No significant disease pattern → Report
- Suspicious pattern → Further analysis required
- Disease suspected → Segmentation and severity assessment

This stage serves as a preliminary screening step prior to the detailed investigation of the suspected area affected by the disease.

VI. SICKNESS-AREA ANALYSIS AND SEVERITY RESEARCH

In case of suspicion of illness, the system carries out an exhaustive analysis of the afflicted region. The U-Net++ framework is engaged to break down the diseased tissue area. The segmentation result yields a more concrete presentation of the affected region, hence several quantitative indicators can be derived [4], [14], [15]. The suggested approach takes into consideration morphological parameters such as:

- Percentage of tumorous area;
- Mitosis number;
- Correlation of glands;
- Properties of the disease region;
- Other morphological parameters for specific diseases.

These indicators are used to assess severity level. Various machine learning approaches such as Random Forest, XGBoost, or SVM may be investigated within stage in question.

ISSN: 3107-6513

VII. COMBINATION OF DISTINCT ATTRIBUTES AND DISEASE IDENTIFICATION

The current model uses data from all stages of processing to compile combined feature representation. These features include the following:

- Tissue properties
- Deep Learning features
- Disease-region features
- Morphological data
- Severity features

The traditional random forest classifiers or XGBoost classifiers are used for classification after a fusion phase has been completed [13], [16].

The most informative output includes: Classification type → Classification sub-type → Severity classification → Supportive features

Using both image-based deep learning features and quantitative morphological readings enables a comprehensive approach to the combination of various features of interest.

VIII. BLOCK DIAGRAM

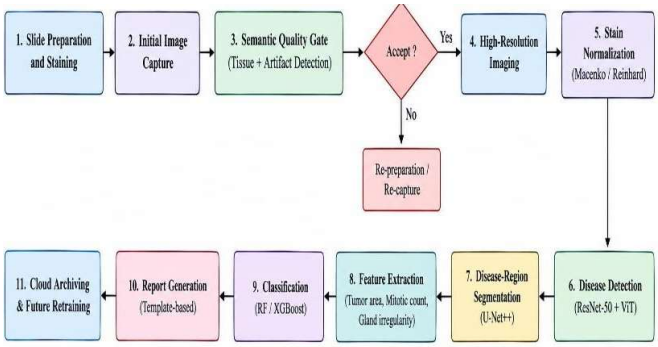


Fig 1. Block diagram of proposed hispathology analysis framework

IX. ALGORITHM METHOD

The suggested system is based on a series of quality control processes, digitally processing the image, applying deep learning, and using machine learning [4], [5], [7], [11].

Step 1: Preparation of Slide and First Image.

- Prepare the histological slide.
- Use a certain coloring method.
- Take the first digital photo. 4. Send the photo to the semantic quality control system.

Step 2: Segmentation of the Tissue.

- Pre-process the taken image.
- Use U-Net for identification of the tissue area.
- Make a tissue mask. 4. Distinguish tissue from the background.

Step 3: Detection of Artifacts.

- Send the taken image to ResNet-50.
- Detect possible artifacts.
- Determine the place of artifacts.
- Prepare an artifact representation.

Step 4: Semantic Overlap Analysis.

- Compare the tissue mask and the artifact mask.
- Find out whether the artifacts cover relevant tissue.
- Calculate if the artifact influenced the diagnostic region.
- Use the quality decision rules determined earlier. 5. Make the Accept or Reject statement.

Step 5: Decision on Basis of Quality.

- If the image does not meet the required quality standards, mark it as Reject.
- Send rejected cases for re-preparation, re-capturing, or manual verification.
- If the image meets the specified quality standards, mark it as Accept.
- Send accepted cases for high-resolution imaging.

Step 6: High-Resolution Imaging and Stain Standardization

- Obtain the high-resolution microscopic image.
- Choose a suitable stain standardization approach.
- Use Macenko or Reinhard standardization approach.
- Produce the standardized image for the diagnostic purpose.

Step 7: Disease Presence Detection

- Feed the standardized image to ResNet-50.
- Process the image through Vision Transformer.
- Check for presence of disease-related patterns. 4. If the area looks suspicious, refer to detailed inspection.

Step 8: Disease Area Segmentation

- Employ U-Net++ on suspected disease area.
- Create the disease area mask.
- Establish the damaged body tissue area. 4. Form the segmented area for feature extraction.

Step 9: Morphological Feature Extraction

- Determine tumor area percentage when necessary.
- Identify mitotic count if data is available.
- Include gland or tissue irregularity measurement.
- Extract further morphological characteristics specific to the disease.

Step 10: Severity Assessment

- Integrate the extracted morphological information.
- Identify severity target based on the training data.
- 3.Train & Validate Random Forest, XGBoost, or SVM.
- Produce the severity prediction.

ISSN: 3107-6513

Step 11: Consolidation of Features and Classification

- Gather the features from deep learning.
- Gather morphology and diseased region features.
- Combine taken features.
- Employ Random Forest or XGBoost.
- Make the disease label and subtype where applicable.

Step 12: Report Creation

- Collect the result of quality control.
- Merge disease, segmentation, severity, and classification results.
- Produce a structured report following template.
- Make measurements and images used for analysis part of the report.

Step 13: Archiving Data Online

- Store original and processed images.
- Keep quality control results and segmentation results.
- Keep features gathered and predictions from model.
- Store created report.
- Ensure availability of archive data for future models' improvements.

X. RELATED WORK

A. Automated Quality Control in Digital Pathology

A number of different computer-aided pathology methods have been developed to help reveal quality issues in digital histopathology imagery. These methods include artifact detection and identifying areas associated with those issues. The significance of quality assurance techniques has thus been highlighted [7], [8], [17].

B. Histopathological Image Segmentation

Various deep learning network architectures have been employed in the past for detection of tissues, cell types and other disease-related regions in histopathological images. In particular, U-Net-based technique is widely used due to the ability it gives for performing the pixelwise analysis of images [1], [4].

C. Stain Normalization

Variations in staining can result in considerable differences between histopathological images. Methods such as Macenko and Reinhard have been used to reduce these colour and stain variations. Stain normalization can provide more consistent inputs for computational pathology models [5], [6], [9], [10], [12].

D. Deep Learning for Histopathological Classification

CNN-based architectures such as ResNet have been applied to histopathological image classification. More recent studies have also investigated transformer-based architectures for medical image analysis. These models can learn morphological patterns from tissue images and support disease classification[11], [13], [16].

E. Disease Region Segmentation and Quantitative Analysis

Segmentation networks such as U-Net++ can be used to identify disease-related regions. The resulting regions can support quantitative measurements such as affected tissue area and other morphological characteristics. Combining these measurements with deep learning features provides another approach for disease analysis [4], [14], [15].

F. Proposed Quality-Gated Diagnostic Framework

The proposed work combines these stages into a single workflow. The main focus is the semantic quality gate, where tissue segmentation and artifact detection are combined to determine whether an artifact affects diagnostically relevant tissue. The quality decision is then used to control whether the image proceeds to normalization and downstream disease analysis [7], [8], [11], [17].

XI. CONCLUSION

This article describes a quality-oriented mechanism for AI supported histopathological assessment. The central premise here is that only the right kind of images get to be processed through the diagnostic process. Rather than only applying the notion of whether something is an artifact, a proposed quality gate deals with the relation of any artifact to its corresponding tissue area. The apparatus incorporates U-Net-based tissue segmentation, ResNet-50-based artifact identification, analysis of tissue and artifact overlap, quality-based acceptance/rejection, and disease identification, among others. The images that passed through the gate will require the processing by other deep learning or machine learning methods to get the results related to the disease. The final stage is automatic generation of reports and storage of the processed cases on the cloud. The proposed method is universal in the sense that it unifies slide quality assessment with diagnostic analysis. Future research will involve training individual models on appropriate datasets and developing diseasespecific criteria.

REFERENCES

[1] Ronneberger O, Fischer P, Brox T: U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. 2015, 9351:234–241. DOI: 10.1007/978-3-319-24574-4_28.

[2] He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770–778. DOI: 10.1109/CVPR.2016.90.

[3] Dosovitskiy A, Beyer L, Kolesnikov A, et al.: An image is worth 16 × 16 words: Transformers for image recognition at scale. International Conference on Learning Representations. 2021.

[4] Zhou Z, Siddiquee MM, Tajbakhsh N, Liang J: UNet++: A nested U-Net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. 2018, 11045:3–11. DOI: 10.1007/978-3-030-00889-5_1.
[5] Macenko M, Niethammer M, Marron JS, et al.: A method for normalizing histology slides for quantitative analysis. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2009:1107–1110. DOI: 10.1109/ISBI.2009.5193250.
[6] Reinhard E, Adhikhmin M, Gooch B, Shirley P: Color transfer between images. IEEE Computer Graphics and Applications. 2001, 21(2):34–41. DOI: 10.1109/38.946629.
[7] Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A: HistoQC: An open-source quality control tool for digital pathology slides. JCO Clinical Cancer Informatics. 2019, 3:1–7. DOI: 10.1200/CCI.18.00157.
[8] Weng Z, Seper A, Pryalukhin A, et al.: GrandQC: A comprehensive solution to quality control problem in digital pathology. Nature Communications. 2024, 15:10685. DOI: 10.1038/s41467-024-54769-y.
[9] Niethammer M, Borland D, Marron JS, Woosley J, Thomas NE: Appearance normalization of histology slides. Machine Learning in Medical Imaging. 2010, 6357:58–66. DOI: 10.1007/978-3-642-15948-0_8.
[10] Vahadane A, Peng T, Sethi A, et al.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE Transactions on Medical Imaging. 2016, 35(8):1962–1971. DOI: 10.1109/TMI.2016.2529665.
[11] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al.: A survey on deep learning in medical image analysis. Medical Image Analysis. 2017, 42:60–88. DOI: 10.1016/j.media.2017.07.005.
[12] Tellez D, Litjens G, Bandi P, Bulten W, Bokhorst JM, Ciompi F, et al.: Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical Image Analysis. 2019, 58:101544. DOI: 10.1016/j.media.2019.101544.
[13] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva VWK, Busam KJ, et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine. 2019, 25:1301–1309. DOI: 10.1038/s41591-019-0508-1.
[14] Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al.: Classification of breast cancer histology images using convolutional neural networks. PLOS ONE. 2017, 12(6):e0177544. DOI: 10.1371/journal.pone.0177544.
[15] Bandi P, Geessink O, Manson Q, van Dijk M, Balkenhol M, Hermsen M, et al.: From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge. IEEE Transactions on Medical Imaging. 2019, 38(2):550–560. DOI: 10.1109/TMI.2018.2867350.
[16] Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al.: Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nature Medicine. 2019, 25:1054–1056. DOI: 10.1038/s41591-019-0462-y.
[17] Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B: Histopathological image analysis: A review. IEEE Reviews in Biomedical Engineering. 2009, 2:147–171. DOI: 10.1109/RBME.2009.2034865.