

Article

Comparative Evaluation of Fusion Strategies Using Multi-Pretrained Deep Learning Fusion-Based (MPDLF) Model for Histopathology Image Classification

Fatma Alshohoumi *  and Abdullah Al-Hamdani

Department of Computer Science, College of Science, Sultan Qaboos University, Muscat 123, Oman; abd@squ.edu.om

* Correspondence: f.alshuhoumi@squ.edu.om

Abstract

Histopathological image analysis remains the cornerstone of cancer diagnosis; however, manual assessment is challenged by stain variability, differences in imaging magnification, and complex morphological patterns. The proposed multi-pretrained deep learning fusion (MPDLF) approach combines two widely used CNN architectures: ResNet50, which captures deeper semantic representations, and VGG16, which extracts fine-grained details. This work differs from previous fusion studies by providing a controlled evaluation of early, intermediate, and late fusion for integrating two pretrained CNN backbones (ResNet50 and VGG16) under single-modality histopathology constraints. To isolate the fusion effect, identical training settings are used across three public H&E datasets. Early fusion achieved the best test performance for the two primary tasks reported here: breast cancer binary classification (accuracy = 0.9070, 95% CI: 0.8742–0.9404; AUC = 0.9707, 95% CI: 0.9541–0.9844) and renal clear cell carcinoma (RCCC) five-class grading (accuracy = 0.8792, 95% CI: 0.8529–0.9041; AUC (OvR, macro) = 0.9895, 95% CI: 0.9859–0.9927). Future work will extend these experiments to additional magnification levels (100×, 200×, and 400×) for breast cancer histopathology images and explore advanced hybrid fusion strategies across different histopathology datasets.

Keywords: histopathology; early fusion; intermediate fusion; late fusion

1. Introduction

Cancer continues to be one of the leading causes of death worldwide [1–3], despite significant achievements in medical imaging technology and advanced therapies. The recent global statistics indicate that cancer caused approximately 10 million deaths in 2020 [2] and nearly 20 million new cases were reported in 2022 [4]. Medical imaging plays an important role in all stages of cancer treatment, involving diagnosing, staging, recommending treatments, and assessing how patients respond to therapy [5,6]. In modern medicine, medical imaging such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), positron emission tomography (PET), and histopathology images are widely used by oncologists to diagnose cancer [7,8]. Although imaging techniques such as CT, MRI, PET scans, ultrasound, and mammograms aid in cancer detection, histopathology images remain the gold standard for confirming malignancy [9–11]. Histopathology images, especially those that are stained with Hematoxylin and Eosin, play an essential role in



Academic Editors: Rafał J. Doniec and Katarzyna Mocny-Pachońska

Received: 24 December 2025

Revised: 5 February 2026

Accepted: 9 February 2026

Published: 16 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

cancer diagnosis, grading, and prognosis [12–14]. Tissue processing for histopathology images preserves its architecture, which assists in identifying disease characteristics and providing essential insights into tumor morphology, structural abnormalities, lymphocytic infiltration, and biological behavior [12,15]. Recent technological advances have improved the quality of histopathology images, which has led to a reduction in disease complications and an improvement in patient outcomes. However, the manual interpretation of histopathology images is prone to inter-observer variability, requires specialized expertise, and is time-consuming [14,16].

As the technology has evolved, histopathological analysis has been digitized via whole-slide images (WSI) [9]. The advancement in digital pathology has facilitated large-scale histopathology images and the development of computational tools, and has improved the efficiency in clinical workflows [17,18]. Due to the advancement of artificial intelligence, deep learning has contributed to digital pathology by offering the potential to improve the accuracy of cancer classification and detection. Recently, many studies have demonstrated the effectiveness and power of deep learning, particularly convolutional neural networks (CNNs), in detecting various diseases by analyzing histopathology images, including lung cancer classification [19,20], breast cancer classification [21,22], and brain tumor classification [23,24], etc. CNNs overcome the limitations of manual feature engineering through their capability to automatically extract hierarchical feature representations [15,25]. Both qualitative data (e.g., architectural patterns such as gland formation, stromal vs. epithelial layout, etc.) and quantitative data (e.g., cell counts, nucleus counts, mitotic counts, etc.) can be extracted from the histopathology images and are essential for accurate diagnoses and monitoring the progression of disease [13,14].

Despite notable advances in deep learning for histopathology image analysis, achieving high accuracy remains one of the key challenges for deep learning models for computational pathology. This is attributed to the heterogeneous and complex nature of cancerous tissues, the lack of large labeled datasets, and variations in image quality caused by differences in staining protocols and different magnification levels [26]. Multi-pretrained deep learning models with fusion strategies have emerged as a potential solution to improve the accuracy of histopathology image classification with limited datasets. This study aims to evaluate the effectiveness of multi-pretrained deep learning models using different fusion strategies for the classification of breast cancer histopathology images and renal clear cell carcinoma (RCCC) grading.

This study employs the proposed model to systematically evaluate practical fusion design choices when integrating multiple pretrained deep learning models under single-modality constraints. The findings offer actionable guidance for selecting optimal fusion strategies in histopathology image analysis, especially in settings where multi-source data are unavailable. By systematically analyzing fusion approaches, this work aims to advance more accurate deep learning frameworks for cancer histopathology analysis. The main contributions of this study are: (1) to present a controlled comparison of early, intermediate, and late fusion while keeping the pretrained backbones and the training protocol fixed and (2) to evaluate the fusion strategies consistently across three public histopathology classification tasks, spanning binary and multiclass settings.

2. Related Work

Recently, technological advancements have transformed the workflow of histopathology image analysis from the traditional practice (manual review of glass slide) to digital workflows supported by whole-slide imaging (WSI) and deep learning [27,28]. This transformation has improved cancer diagnosis, enabling gigapixel-scale image computation, accurate classification, and automated tumor grading [29]. Deep learning offers great

benefits to pathologists and patients by enhancing workflow efficiency, accelerating interpretation speed, and making the diagnosis more reliable [28,30,31]. Many recent studies have highlighted the applications of deep learning, including cancer diagnosis, tumor grading, molecular prediction, and automation of pathological workflows. For instance, a recent study conducted by Y. Ma et al. demonstrated that most studies related to deep learning in histopathology focused on diagnosis (30.9%) and detection (24.2%) tasks and have achieved the highest performance ($AUC \approx 96\%$), affirming that deep learning has become central to cancer diagnostics [32]. Moreover, deep learning can extract features from histopathology images that can help in inferring molecular subtypes, mutations, and treatment response [33]. Furthermore, deep learning models have achieved high accuracy in the automatic grading of meningiomas [34], predicting breast cancer [35], and predicting genetic mutations and protein biomarkers from histopathology images for cancer types such as colorectal carcinoma and melanoma [36], etc. Although deep learning has achieved remarkable advancements in computational pathology, several challenges still limit its implementation into clinical practice, such as the complexity of whole-slide imaging, lack of diverse and well-annotated datasets, model interpretation issues, technical variations across clinics, and ethical and regulatory factors [37–40]. These limitations hinder the effectiveness of the unimodal or single model, which struggles to capture the rich and diverse features that are necessary for achieving high classification accuracy. Recent studies [41–47] have shown that multi-modal approaches using fusion strategies offer promising opportunities to enhance the classification accuracy. By leveraging the strength of multiple pretrained deep learning models and extracting richer features, these approaches have been effective in handling the complexity of histopathology images [48,49]. However, the multi-modal approach requires an effective fusion strategy to combine the extracted features from different pre-trained models to gain better performance.

In biomedical research, multimodal fusion strategies are widely applied to extract the complementary information across modalities such as genomics, clinical data, and medical imaging [50,51]. By applying data fusion strategies, multimodal aims to combine the extracted features into a unified space, which can help to enhance the predictive accuracy, and thus advance precision oncology, through improved diagnosis and treatment selection [52,53]. The most predominant fusion strategies used in multimodal deep learning are generally classified into four broad strategies: early (input/feature-level), intermediate (feature/joint level), late fusion (decision-level), and hybrid fusion (combination of multiple fusion strategies) [52,54]. Early fusion strategy relies on combining low-level features from multiple modalities at the initial stage before feeding them into the deep learning model for further training [52,55]. It is straightforward and efficient, as it involves the concatenation of low-level features [44]. However, it is susceptible to noisy features that can affect the outcomes. [55]. Intermediate fusion is known as joint fusion, and it relies on combining feature vectors that are extracted by separate modality-specific branches at mid-network layers [53,55,56]. This joint representation allows for the model to capture complex and non-linear intermodal relationships [56]. It helps preserve fine-grained morphological details, which are valuable in histopathology tasks [53]. Late fusion strategy relies on averaging the predictions (probabilities) from separate models [52,53,55]. Unlike previously mentioned fusion strategies, late fusion is limited in capturing fine-grained features, as it only aggregates the final outputs of each modality [53]. Hybrid fusion is less common, and it integrates different fusion strategies: for example, integrating features extracted by intermediate fusion and late fusion [53]. In histopathology tasks, hybrid fusion can be used to combine high-dimensional image features with low-dimensional features from other clinical sources [57]. Figure 1 presents an overview of the common fusion strategies:

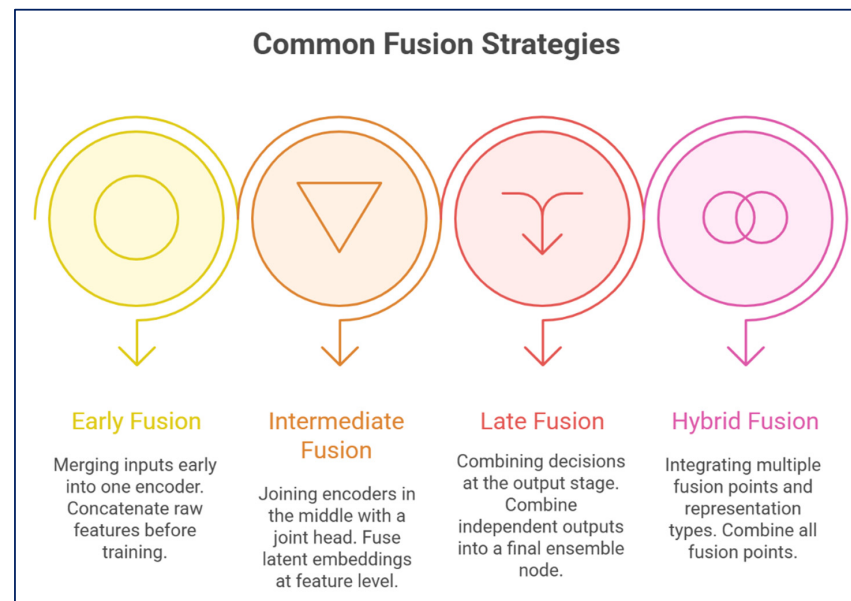


Figure 1. An overview of the common fusion strategies.

Recently, in 2020, S. C. Huang et al. applied different fusion strategies (early, intermediate, and late fusion) to evaluate multimodal deep learning models for automated pulmonary embolism classification by combining different sources of information (computed tomography pulmonary angiography (CTPA) imaging with electronic medical records (EMR) of patients) [54]. The findings showed that late fusion demonstrated the best performance (AUROC = 0.947), outperforming the single modality. In 2022, Steyaert et al. developed and compared multimodal deep learning models that combine histopathology whole slide images and gene expression data for the purpose of improving survival prediction in adult and pediatric brain tumors. ResNet50 was used for extracting features from histopathology images, and a multilayer perception was used for extracting features from gene expression data. The study evaluated early fusion, late fusion, and intermediate fusion (joint fusion). The results showed that multimodal fusion models outperformed single modalities in survival prediction. The study demonstrated that early fusion performed the best and yielded the highest test CS (0.836) among other fusion strategies [58]. In 2023, Cahan et al. developed and compared multimodal deep learning fusion models by fusing two sources of data, CT pulmonary angiography (cTPA) imaging with electronic health record (EHR) tabular data for pulmonary embolism 30-day mortality prediction. The study evaluated different fusion strategies, including early fusion, intermediate fusion, and late fusion. The findings demonstrated that the intermediate fusion strategy achieved the best performance with AUC = 0.96, sensitivity = 90%, and specificity = 94%, outperforming single models that were imaging-only and HER-only [59]. In 2023, Kumar et al. developed a deep learning multimodal fusion model for lung disease classification by integrating chest X-ray images with clinical laboratory data. Different pre-trained CNN backbones, including DenseNet121, DenseNet169, and ResNet50, were used for extracting features from X-ray images, while LSTM and self-attention networks were used for training clinical data. The results showed that intermediate fusion outperformed late fusion and single modalities, achieving a higher F1-Score = 94.75% [48]. In 2023, Zheng et al. developed an application of transfer learning and ensemble learning to improve the classification of histopathology images. A late fusion strategy was employed to combine predictions from multiple pre-trained CNN models, including VGG16, ResNet50, InceptionV3, and DenseNet12. The findings revealed that the ensemble model outperformed single-pretrained models, achieving an accuracy $\approx$ 98% and an F1-Score $>$ 0.97. These results indicated the effectiveness of using transfer learning

with a late fusion strategy [60]. In 2024, Chakravarthy et al. proposed a hybrid fusion strategy (intermediate + late fusion) to combine deep features from multiple CNN models, VGG16, VGG19, ResNet50, and DenseNet121. The study aimed to enhance multiclass classification of breast cancer, using mammography datasets. The extracted features were concatenated and passed through a fully connected classifier. The results revealed that the hybrid model achieved the highest accuracy, reaching 98.83%, outperforming late fusion and single models [61]. In 2025, J. Li developed a hybrid fusion framework to improve the early diagnosis of oral squamous cell carcinoma (OSCC) from histopathology images. The hybrid model combined deep features from a cross-attention vision transformer (CrossViT) with handcrafted texture and color features (LBP, GLCM, FCH) using hybrid feature-level fusion and was classified using an artificial neural network (ANN). The hybrid fusion model was compared with pre-trained models, including ResNet50/101, VGG16/19, EfficientNetB0/B7, ViT, and CrossViT. The results showed that the proposed hybrid model achieved the highest performance with accuracy = 99.36%/99.59%, AUC $\approx$ 0.999, and sensitivity and specificity $\geq$ 99% over both pre-trained models and ViT alone [62]. In 2025, Das et al. implemented an ensemble learning framework using multiple pre-trained deep learning models to improve the accuracy of diagnosing invasive ductal carcinoma (IDC) from histopathology images. Breast histopathology images at three magnification levels (100 $\times$, 200 $\times$, and 400 $\times$) were used. Pre-trained models, including ResNet50, Xception, MobileNetV2, VGG16, and VGG19, were used to extract features from histopathology images, and a late fusion strategy was implemented through the average and weighted ensemble learning of multiple pre-trained models. The obtained results indicated that the weighted average ensemble model (combining ResNet50, VGG16, and VGG19) achieved the best performance across magnification levels, with accuracy = 97.27%, AUC = 0.975, and F1-Score $\approx$ 0.97, while ResNet50 was the best standalone model, achieving approximately 98% accuracy on 100 $\times$ images. The overall results showed that the ensemble model achieved better generalization [41]. In 2025, Asif et al. applied an intermediate fusion strategy through hierarchical concatenation to combine multi-scale deep features extracted from MobileNet (pre-trained on ImageNet) and a custom multi-scale depthwise dilated residual network (MSDDR-Net). Three public brain MRIs (Brain Tumor MRI Dataset, Br35H, and Figshare) were used to assess the proposed model (HMDFF-Net). The results showed that HMDFF-Net achieved 99.31% accuracy, 99.28% precision, 99.25% recall, and an F1-score of 99.26 on the primary MRI multiclass dataset, outperforming DenseNet, ResNet, and transformer-based models [63]. In 2025, Punarselvam et al. conducted a comprehensive evaluation of deep learning models enhanced with feature fusion strategies to improve the classification accuracy of cervical cancer using cytology and histopathology images. Different fusion strategies were evaluated, including early fusion, intermediate fusion, late fusion, and hybrid (CNN + Transformer) fusion. Different pre-trained models, including ResNet50, DenseNet121, EfficientNet-B4, and a hybrid CNN-Transformer, were used. The findings showed the superiority of hybrid fusion, achieving accuracy = 96.2%, precision = 96.0%, recall = 96.5%, F1 = 96.2%, and AUC = 0.981 [64]. In 2025, Sahu proposed a deep learning framework to improve brain tumor classification accuracy from MRI scans. The proposed framework was based on hybrid fusion (intermediate + late fusion) to combine features from different pretrained models, including VGG16, InceptionV3, DenseNet121, Xception, and InceptionResNetV2. The majority voting strategy was used in the hybrid fusion. The proposed ensemble model achieved the highest performance, reaching 99.46% accuracy, 0.995 precision, 0.995 recall, and a 0.992 F1-score, alongside an AUC of 0.995 [65].

The following comparative Table 1 presents a comprehensive overview of thirteen state-of-the-art studies focusing on fusion strategies in multimodal deep learning for medical image analysis across different diseases. The studies were conducted between 2020 and 2025.

Table 1. A comprehensive overview of state-of-the-art studies focusing on fusion strategies in multimodal deep learning.

Study	Year	Disease/Task	Modality	Implemented Model	Fusion Level/Point	Fusion Strategy Description
[54]	2020	Pulmonary embolism classification	Multimodal (CTPA + EMR)	3D CNN (CTPA) + MLP/FFNN (EMR)	Decision-level (late)	Late fusion of modality-specific outputs (probability-level)
[58]	2022	Brain tumor survival prediction	Multimodal (WSI + gene expression)	ResNet50 (WSI) + MLP (gene expression)	Feature-level; decision-level; joint	Early: concatenate image + omics features; joint: end-to-end; late: score fusion
[59]	2023	Pulmonary embolism 30-day mortality prediction	Multimodal (CTPA + EHR)	3D CNN (CTPA) + MLP (EHR tabular)	Feature-level (intermediate)	Intermediate fusion of imaging and tabular embeddings (feature concatenation/joint MLP)
[48]	2023	Lung disease classification	Multimodal (CXR + clinical/lab data)	DenseNet121/169/ResNet50 (CXR) + LSTM/Self-attention (clinical)	Feature-level; decision-level	Intermediate: concatenate image + clinical embeddings; late: score fusion
[62]	2023	Histopathology image classification	Single-modality (histopathology images)	VGG16, ResNet50, InceptionV3, DenseNet121	Decision-level	Late (combine predictions from multiple pretrained CNNs (voting/averaging))
[61]	2024	Breast cancer multiclass classification	Single-modality (mammography)	VGG16, VGG19, ResNet50, DenseNet121	Feature-level + decision-level	Concatenate deep features; final classification with FC layers (hybrid design)
[62]	2025	OSCC histopathology diagnosis	Single-modality (histopathology images)	CrossViT + handcrafted (LBP/GLCM/FCH) + ANN	Feature-level	Hybrid(Fuse CrossViT features with handcrafted features (LBP/GLCM/FCH) → ANN)
[41]	2025	IDC diagnosis across magnifications	Single-modality (histopathology images)	ResNet50, Xception, MobileNetV2, VGG16, VGG19	Decision-level	Late (average/weighted ensemble of pretrained CNN predictions)
[63]	2025	Brain MRI tumor classification	Single-modality (brain MRI; multi-scale features)	MobileNet + custom MSDDR-Net	Feature-level	Intermediate (hierarchical concatenation of multi-scale deep features)
[64]	2025	Cervical cancer classification	Multimodal (SIPaKMeD, Herlev, CRIC (+HPV/metadata))	ResNet50/DenseNet121 / EfficientNet-B4 + CNN-Transformer	Feature-level	Hybrid (feature fusion using CNN and transformer representations)
[65]	2025	Brain tumor MRI classification	Single-modality (brain MRI)	VGG16, InceptionV3, DenseNet121, Xception, InceptionResNetV2	Feature-level + decision-level	Hybrid fusion with majority voting/ensemble decision
[57]	2025	Glioma survival/risk prediction (GBM vs. LGG)	Multimodal (WSI + clinical/omics)	ResNet50 (WSI) + MLP (omics); Cox model	Feature-level; joint; decision-level	Early: concatenate image + clinical/omics; late: combine risk scores
[66]	2025	Multiple pathology downstream tasks (foundation models)	Single-modality (H&E histopathology images)	CHIEF, GigaPath, UNI, TITAN, PRISM + Meta-Encoder	Feature-level	Hybrid (meta-encoder over multiple pathology foundation models)

Abbreviations: 3D CNN = three-dimensional convolutional neural network; ANN = artificial neural network; Cox = Cox proportional hazards model; CXR = chest X-ray; CTPA = computed tomography pulmonary angiography; EHR = electronic health record; EMR = electronic medical record; FC = fully connected; FCH = handcrafted features (color/texture); FFNN = feed-forward neural network; GBM = glioblastoma; GLCM = gray-level co-occurrence matrix; H&E = hematoxylin and eosin; IDC = invasive ductal carcinoma; LBP = local binary pattern; LGG = lower-grade glioma; LSTM = long short-term memory; MLP = multilayer perceptron; MRI = magnetic resonance imaging; MSDDR-Net = multi-scale depthwise dilated residual network; OSCC = oral squamous cell carcinoma; ViT = vision transformer; CrossViT = cross-attention vision transformer; and WSI = whole slide image. CHIEF, GigaPath, UNI, TITAN, and PRISM are pathology foundation model names, as reported in the cited study.

As illustrated in Table 1, the studies highlight the evolution of fusion strategies in deep learning for medical image analysis and demonstrate how the stage of fusion strategy and integration method affect model performance and computational efficiency. Earlier studies of multimodal strategies relied on simple fusion approaches (late fusion), such as Huang et al. (2020) [54], which combined the final predictions from the fused models of different data sources (imaging and EMR). Late fusion achieved strong discriminative power in the classification task (AUROC = 0.947) with minimal complex architecture. In 2022, Steyaert et al. expanded the concept of fusion to early fusion to combine WSI and RNA-seq data for brain tumor prognosis at the early level, showing that the integration of multimodalities at an early stage achieved higher C-index values (0.836–0.919) than single-modal models,

indicating the power of integrating morphological and molecular features in disease prediction. By 2023, studies like Kumar et al. (2023) [48] and Cahan et al. (2023) [59] implemented intermediate fusion to emphasize representation learning between different modalities, such as chest X-ray or 3D-CT imaging with EMR data through latent concatenation and attention mechanisms. Intermediate fusion outperformed image-only or tabular-only models, achieving AUC values of around 93–96%. In the same year, Zheng et al. (2023) [60] focused on utilizing multi-pretrained CNNs, such as VGG16, VGG19, InceptionV3, Xception, ResNet50, and DenseNet201, to extract features from a single modality of breast cancer histopathology images, employing an ensemble approach to combine the final decisions of each pre-trained model. The study opened the door toward applying fusion strategies in different multimodal approaches, relying on multiple pre-trained models and a single data modality, instead of combining different data modalities, such as medical imaging and tabular clinical data. In the years 2024–2025, the research focus shifted to hybrid fusion, which integrates the early fusion and late fusion, or intermediate fusion with late fusion. Chakravarthy et al. (2024) [61] implemented hybrid fusion to fuse features of multi-pretrained deep learning models, achieving higher classification accuracy, reaching 98.83% in the mammography dataset. Similarly, recent studies, such as Sahu (2025) [65] and Punarselvam (2025) [64], demonstrated the effectiveness of hybrid CNN–CNN–Transformer and ensemble approaches in enhancing the classification accuracy across MRI and cytology datasets.

Overall, the comparison among studies between 2020 and 2025, focusing on multimodal learning and fusion strategies, confirms that fusing features from different modalities or multiple deep learning models enhances diagnostic accuracy and outperforms single modality in many tasks. Generally, recent studies demonstrated the effectiveness of integrating features through intermediate fusion and hybrid fusion with attention mechanisms, yielding remarkable improvements in diagnostic and prognostic accuracy. The selection of the fusion point at the early, intermediate, or late stage affects the model's performance. Most studies focused on combining different modalities, such as imaging with tabular data. Although the fusion of different modalities consistently improves the accuracy, the challenge of limited datasets, especially multimodal datasets, remains [63,65]. To help mitigate this limitation, this work shifts toward exploring and evaluating multi-pretrained deep learning models and fusion strategies on a single modality instead of multi-modality (imaging with tabular data). To the best of our knowledge, there is a lack of studies that implement comprehensive experiments on all fusion strategies with a single modality. This work investigates multiple pretrained deep learning models with fusion strategies based on single modality instead of multi-modality and evaluates the implemented fusion strategies among different histopathology datasets.

3. Materials and Methods

This work contributes to offering a comprehensive evaluation of a multi-pretrained deep learning fusion-based approach against single baseline models. This section describes the methodology used in this work and presents details about the dataset descriptions, the proposed pretrained fusion-based approach, and the training setup.

3.1. Datasets Description

This work utilized three publicly available image histopathology datasets. Each dataset was prepared for a specific cancer type and diagnostic task: BreakHis-40× (binary classification), BreakHis-40× (multiclass classification (8 classes) [67], and RCCC (5 classes) [68]. All datasets are stained with Hematoxylin and Eosin. For binary classification, the BreakHis dataset at magnification level (40×) was used to classify breast cancer into benign and malignant tumors. For multiclass classification, two datasets were used: breast cancer and RCCC. In this work, both datasets were used as provided by their sources,

with no additional manual filtering(e.g., removing images due to artifacts, poor quality, or duplication). Table 2 presents details about the datasets used for multiclass classification:

Table 2. Multiclass classification dataset details.

Dataset Type	BreakHis Dataset	Renal Clear Cell Carcinoma Dataset
Description	Public dataset for breast tumors	Kidney histopathology dataset
Total Classes	8 classes (4 benign + 4 malignant)	5 grades (Grade 0 to Grade 4)
Classification Type	Multiclass subtype classification	Multiclass grading
Class Labels	Adenosis (A) Fibroadenoma (F) Tubular Adenoma (TA) Phyllodes Tumor (PT) Ductal Carcinoma (DC) Lobular Carcinoma (LC) Mucinous Carcinoma (MC) Papillary Carcinoma (PC)	Grade 0: Non-cancerous tissue (benign/normal). Grade 1: Very low-grade cancer Grade 2: Low to moderate differentiation. Grade 3: Moderately to poorly differentiated. Grade 4: High-grade, aggressive cancer.
Histopathology Stain	Hematoxylin and Eosin	Hematoxylin and Eosin
Image Magnification	40×	40×
Published date	21 June 2023	December 2022

3.2. Proposed Multi-Pretrained Deep Learning Fusion (MPDLF) Approach

The main objective of this work is to investigate the effectiveness of multi-pretrained deep learning models, specifically ResNet50 and VGG16, using different fusion strategies (early fusion, intermediate fusion, and late fusion) for histopathology image classification. The proposed approach enhances feature richness by combining the strengths of the multiple pretrained models (ResNet50 and VGG16) for histopathology images, which can improve the classification performance with even a single modality. The combination can leverage VGG16's fine-grained details and ResNet50's deeper semantic representations. This work compares the performance of fusion strategies in both binary and multiclass classification tasks. It seeks to provide comprehensive comparisons of different fusion strategies between different histopathology datasets, including breast cancer and RCCC, to identify the optimal fusion strategy that can enhance classification performance and aid in advancing digital pathology. The architecture of the proposed MPDLF model is composed of four components, including the input and processing stage, the feature extraction stage, the fusion strategy, and the classification stage. Figure 2 illustrates the main components of the proposed fusion approach:

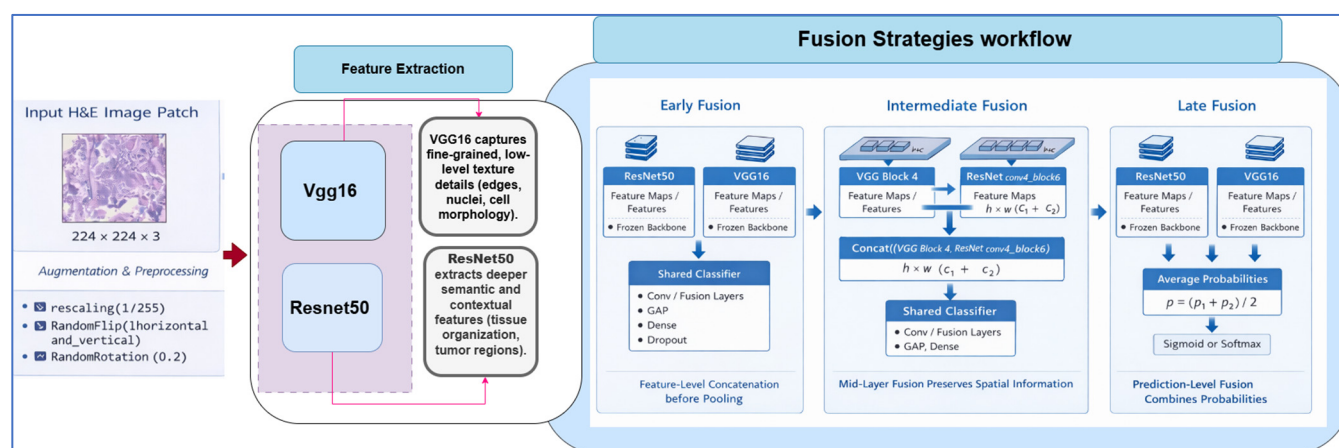


Figure 2. Main components of the proposed model.

3.3. Data Splitting Strategy and Preprocessing

The BreakHis dataset for eight-class classification was collected through a clinical study conducted at the Pathological Anatomy and Cytopathology Laboratory in Brazil. It contains histopathological images from four benign and four malignant breast tumor subtypes. The dataset includes 7909 microscopic images acquired at four magnification levels ($40\times$, $100\times$, $200\times$, and $400\times$) from 82 patients. In this study, we only focused on $40\times$ images to ensure a consistent comparison with the RCCC dataset, which was evaluated at the same magnification. All samples were obtained from breast tissue biopsies collected during open surgery.

While patient-level partitioning is preferred for evaluating clinical generalization, the primary aim of this study is to provide a controlled and fair comparison of fusion strategies under identical experimental conditions. Therefore, we adopted an image-level stratified hold-out split, following common practice in publicly available histopathology benchmarks. Specifically, within each class, images were randomly shuffled using a fixed seed (seed = 42) and divided into 70% training, 15% validation, and 15% testing subsets. This design ensures reproducibility and preserves class distribution across splits. Nevertheless, because patient identifiers were not used during partitioning, correlations between samples originating from the same patient may exist across subsets, which could lead to optimistic performance estimates. Accordingly, the reported results should be interpreted as comparative evidence for fusion design choices, rather than definitive estimates of patient-level generalization performance. To mitigate this concern, we report F1-based metrics in addition to accuracy and emphasize that all models were evaluated using the same data split, ensuring a fair and consistent comparison. In future work, we will conduct patient-level evaluation to further assess generalization to unseen patients, including experiments on a colon cancer dataset with nine classes.

The KMC kidney histopathology dataset (RCCC dataset), first presented in 2023 by [1], includes images of non-cancerous tissue (Grade 0) and malignant tumors (Grades 1–4). Kidney samples were obtained through surgical biopsy at the Department of Pathology, KMC, Mangalore, India. The data were provided by the authors of [1] upon request. For consistency in evaluating fusion strategies across datasets, we followed the same image-level splitting protocol used for the BreakHis experiments. Before model training, preprocessing steps were applied to the dataset, including image resizing to 224×224 pixels, normalization to the range $[0, 1]$, and augmentation using random rotation and flipping. No stain normalization method (e.g., Macenko/Reinhard) or color augmentation was applied in the current pipeline.

3.4. Feature Extraction

The proposed MPDLF approach used two widely known pretrained CNNs (ResNet50 and VGG16) for feature extraction from histopathology image datasets. The purpose of selecting VGG16 as a feature extractor in this proposed approach is that VGG16's smaller (3×3) convolutional kernels can extract fine-grained features from histopathology images, such as local spatial features, including edges and cell morphology, which are essential for histopathology analysis [69,70]. Meanwhile, the deep layers in ResNet50 help to capture abstract and high-level features [71]. Both models are pretrained on ImageNet. The feature representations extracted from both models provide complementary insight into tissue structures. Combining low-level (fine-grained) features and high-level (semantic) features enhances classification performance [72,73]. All pretrained backbone networks (ResNet50 and VGG16) were used as fixed feature extractors; their weights were fully frozen during training. Only the fusion layers and the final classifier were optimized.

3.5. Fusion Strategies

In this work, three common fusion strategies (early, intermediate, and late) were implemented to fuse the extracted features from a multi-pretrained deep learning model framework combining ResNet50 and VGG16 at different fusion levels. The early fusion strategy (feature-level) integrates the feature maps from both VGG16 and ResNet50 at an early stage through a concatenation technique, and the fused features are passed through a series of convolutional layers for joint feature learning. Early fusion helps to retain or preserve rich spatial and morphological features from both models. Intermediate fusion concatenates the feature maps that are extracted from selected mid-layer blocks in both models (e.g., Block 4 of VGG16 and conv4_block6 of ResNet50). Fusing mid-layer blocks allows for combining semantically meaningful or abstract representations [74]. These intermediate layers were selected because they provide mid-level representations that preserve spatial information that is relevant to histopathological patterns, while avoiding overly low-level features and overly task-specific deep features. Both layers produce feature maps at the same spatial resolution for 224×224 inputs, enabling direct concatenation.

Unlike early and intermediate fusion, late fusion aggregates the final predictions of each model at the decision level, which makes it computationally efficient [75]. However, the ability of late fusion to learn joint morphological patterns is limited because it does not combine the feature maps and therefore fails to capture cross-model feature interactions or spatial patterns [76,77]. The following lines present the mathematical formulation for fusion strategies (early, intermediate, and late):

Early Fusion Mathematical Formulation:

- x : Input (histopathology image).
- $f_R(x)$: Feature vector extracted from ResNet50 branch.
- $f_V(x)$: Feature vector extracted from VGG16 branch.
- z_R : ResNet50's vector after global average pooling (GAP).
- z_V : VGG16's after global average pooling (GAP).
- After global average pooling (GAP) $z_R = \text{GAP}(f_R(x))$, $z_V = \text{GAP}(f_V(x))$
- Fusion by concatenation:

$$z_{\text{Early Fusion}} = z_R \oplus z_V$$

where $z_{\text{Early Fusion}} \in \mathbb{R}^{d_R+d_V}$

- The fused vector at the early stage is fed into the dense layer and then into the final classifier.

Intermediate Fusion Mathematical Formulation:

- x : Input (histopathology image).
- $f_R^{(l_R)}(x)$: Feature map from ResNet50 (layer conv4_block6_out).
- $f_V^{(l_V)}(x)$: Feature map from VGG16 (layer block4_pool).
- Mid-layer output vectors:

$$M_R = f_R^{(l_R)}(x), M_V = f_V^{(l_V)}(x)$$

where

$l_R = \text{conv4_block6_out}$ (ResNet50 mid-layer),

$l_V = \text{block4_pool}$ (VGG16 mid-layer).

- Fusion by concatenation at mid-layer:

$$M_{\text{Intermediate Fusion}} = M_R \oplus M_V$$

$$M_{\text{Intermediate Fusion}} \in \mathbb{R}^{H \times W \times (C_R + C_V)}$$

where

H, W are the spatial height and width of the mid-level feature maps.

C_R and C_V are the number of channels from ResNet50 and VGG16.

- The fused vector at the mid-layer stage is fed into an additional convolutional layer for further processing, followed by a dense layer, and then to the final classifier.

Late Fusion Mathematical Formulation:

- x : Input (histopathology image).
- $p_R(x)$: Class-probability vector predicted by the ResNet50 branch.
- $p_V(x)$: Class-probability vector predicted by the VGG16 branch.
- Late fusion: Averaging the class-probability of both branches.

$$p_{\text{Late Fusion}} = \frac{1}{2}(p_R(x) + p_V(x))$$

Component-wise, for each class, $k \in \{1, \dots, C\}$:

$$p_{\text{Late Fusion}}(k) = \frac{p_R(k) + p_V(k)}{2}$$

- Then, argmax is applied over the fused probabilities to get the final predicted class.

$$\hat{y}_{LF} = \arg \max_{k \in \{1, \dots, C\}} p_{\text{Late Fusion}}(k)$$

3.6. Experimental Setup of the Proposed Multi-Model Fusion Approach

The proposed multi-model fusion approaches were implemented using TensorFlow (version 2.17.0) and Keras (version 3.4.1). Experiments were conducted on a local workstation running Windows 11 Pro (version 25H2, OS build 26200.7623) (Microsoft, Redmond, WA, USA) on a Dell system (Round Rock, TX, USA) equipped with a 13th-generation Intel Core i9-13950HX CPU (2.20 GHz; Santa Clara, CA, USA) and 64 GB RAM. The TensorFlow CPU build was used (CUDA/GPU acceleration was not enabled). Hyperparameter optimization was performed using Keras Tuner (version 1.4.7) with the Random Search algorithm to identify an optimal combination of dense units, dropout rate, and learning rate. Early stopping and model checkpointing were employed to mitigate overfitting and to retain the best-performing weights based on the lowest validation loss. Model performance was evaluated using accuracy and area under the curve (AUC). Receiver operating characteristic (ROC) curves were also plotted to visualize discriminative performance by illustrating the trade-off between true positive rate (sensitivity) and false positive rate (1 – specificity).

The following Table 3 presents the experimental setup details:

Table 3. Experimental setup.

Component	Setting
Data structure	Directory-based split: train/, val/, test/
Image size	224 × 224 pixels
Batch size	16, 32 (Dataset's size dependent).
Maximum epochs	50 (with EarlyStopping)
Label mode	Categorical (one-hot encoding)

Table 3. *Cont.*

Component	Setting
Preprocessing	Rescaling by 1/255
Data augmentation	RandomFlip(“horizontal_and_vertical”), RandomRotation(0.2)±
Backbone architectures	ResNet50 and VGG16, include_top = False, ImageNet weights, frozen
Optimizer	Adam (learning rate selected via Keras Tuner)
Loss function	Categorical cross-entropy (multiclass classification); binary cross-entropy (binary classification)
Evaluation metrics	Accuracy; AUC
Hyperparameter search	Keras Tuner RandomSearch, objective = validation accuracy, max_trials = 10
Tuned hyperparameters (per fusion strategy)	Early Fusion:
	- dense_units: 64–256 (step 64). - dropout_rate: 0.2–0.5 (step 0.1). - learning_rate: $\{1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}\}$.
	Intermediate Fusion:
	- conv1_filters: 128–256 (step 64). - conv2_filters: 64–128 (step 32). - dense_units: 128–256 (step 64). - dropout_rate: 0.2–0.5 (step 0.1). - learning_rate: $\{1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}\}$.
	Late Fusion:
	- dense_resnet: 64–256 (step 64). - dense_vgg: 64–256 (step 64). - learning_rate: $\{1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}\}$.
Callbacks	ModelCheckpoint (best val_loss); EarlyStopping (monitor val_loss, patience = 5, restore best weights)

4. Results and Discussion

This section presents the findings obtained from experiments that were used to evaluate the effectiveness of fusion strategies with multi-pretrained models (ResNet50 and VGG16) on three histopathology datasets: BreakHis-40× (binary classification), BreakHis (multiclass classification, eight classes), and the RCCC grading dataset (five grades). Model accuracies and the area under the curve (AUC) for the training, validation, and test sets were reported to assess the performance of the unimodal models and the proposed multi-pretrained fusion-based model. In addition, the test performance and class-wise analysis for each dataset were reported. The following subsections present the experimental results and evaluation for each dataset:

4.1. Performance of Fusion-Based and Unimodal Models for Binary Classification (BreakHis-40×)

Table 4 shows the dataset splitting into three subfolders, including training, validation, and testing.

Table 4. BreakHis_40× specification (binary classification).

Dataset	Total Size	Training	Validation	Testing
BreakHis_40× (Binary Classification)	1995	1395	298	302

The Table 5 results demonstrate that multi-pretrained models (ResNet50 and VGG16) with the early fusion strategy achieved the best overall performance, with a test accuracy of

0.9073 and an AUC of 0.9709, outperforming the other fusion-based multi-pretrained models and the unimodal baselines (ResNet50 alone or VGG16 alone). The results indicate that combining features from ResNet50 and VGG16 at an early stage encourages complementary learning between the fine-grained features extracted by VGG16 and the deeper contextual representations of ResNet50 [61,64]. Compared to the early fusion strategy, intermediate fusion performed moderately (accuracy = 0.8775; AUC = 0.9684), supporting the usefulness of combining features for richer representations. Unlike early and intermediate fusion, late fusion yielded the lowest results (accuracy = 0.8675; AUC = 0.9114) due to missing the complementary features from both pretrained models and only fusing the final predictions of each model. Overall, these findings show the superiority of feature-level fusion for binary breast cancer classification.

Table 5. Performance of fusion-based and unimodal models for BreakHis-40× (binary classification).

Modality Type	Pre-Trained Model	Fusion Strategy	Accuracy (%)			AUC%		
			Training	Validation	Testing	Training	Validation	Testing
Unimodal	ResNet50	-	0.6867	0.6879	0.6854	0.5805	0.5436	0.5867
Unimodal	VGG16	-	0.916	0.885	0.89	0.968	0.939	0.953
Multi-pretrained	(ResNet50 & VGG16)	Early Fusion	0.9355	0.8993	0.9073	0.9791	0.9587	0.9709
Multi-pretrained	(ResNet50 & VGG16)	Intermediate Fusion	0.8910	0.8859	0.8775	0.9687	0.9633	0.9684
Multi-pretrained	(ResNet50 & VGG16)	Late Fusion	0.8573	0.8523	0.8675	0.9260	0.8993	0.9114

The following Table 6 presents the test-set performance for additional analysis:

Table 6. Test-set performance of fusion-based and unimodal models with additional metrics of BreakHis (40×) for binary classification.

Model	Test Accuracy	Test AUC	Macro Precision	Macro Recall	Macro F1	Weighted Precision	Weighted Recall	Weighted F1
ResNet50 (single)	0.6854	0.5867	0.34	0.50	0.41	0.47	0.69	0.56
VGG16 (single)	0.8907	0.9538	0.85	0.85	0.85	0.87	0.87	0.87
Early Fusion	0.9073	0.9709	0.89	0.90	0.89	0.91	0.91	0.91
Intermediate Fusion	0.8775	0.9684	0.85	0.88	0.86	0.89	0.88	0.88
Late Fusion	0.8675	0.9114	0.87	0.81	0.83	0.87	0.87	0.86

Table 6 summarizes the test performance on BreakHis (40×) for binary classification with additional metrics, and Table A1 (Appendix A) presents the corresponding class-wise metrics. Overall, the fusion-based approaches achieved the best test performance, with early fusion obtaining the highest scores (Macro-F1 = 0.89, Weighted-F1 = 0.91), followed by intermediate fusion (Macro-F1 = 0.86, Weighted-F1 = 0.88). VGG16 (single model) also performed well (accuracy = 0.8907, AUC = 0.9538, Macro-F1 = 0.85, Weighted-F1 = 0.87). However, ResNet50 showed weaker class discrimination (AUC = 0.5867, Macro-F1 = 0.41, Weighted-F1 = 0.56).

The class-wise analysis shows that early fusion detects both classes well with balanced sensitivity (Class 0 recall = 0.86; Class 1 recall = 0.93). In contrast, late fusion exhibits reduced sensitivity for Class 0 (recall = 0.66), despite maintaining high recall for Class 1 (recall = 0.96). Overall, these findings indicate that the fusion approach improves class discrimination and provides more stable class-wise performances compared with a single-model baseline in BreakHis binary classification. Figure 3 presents the ROC curves for each model (Figure 3a–f).

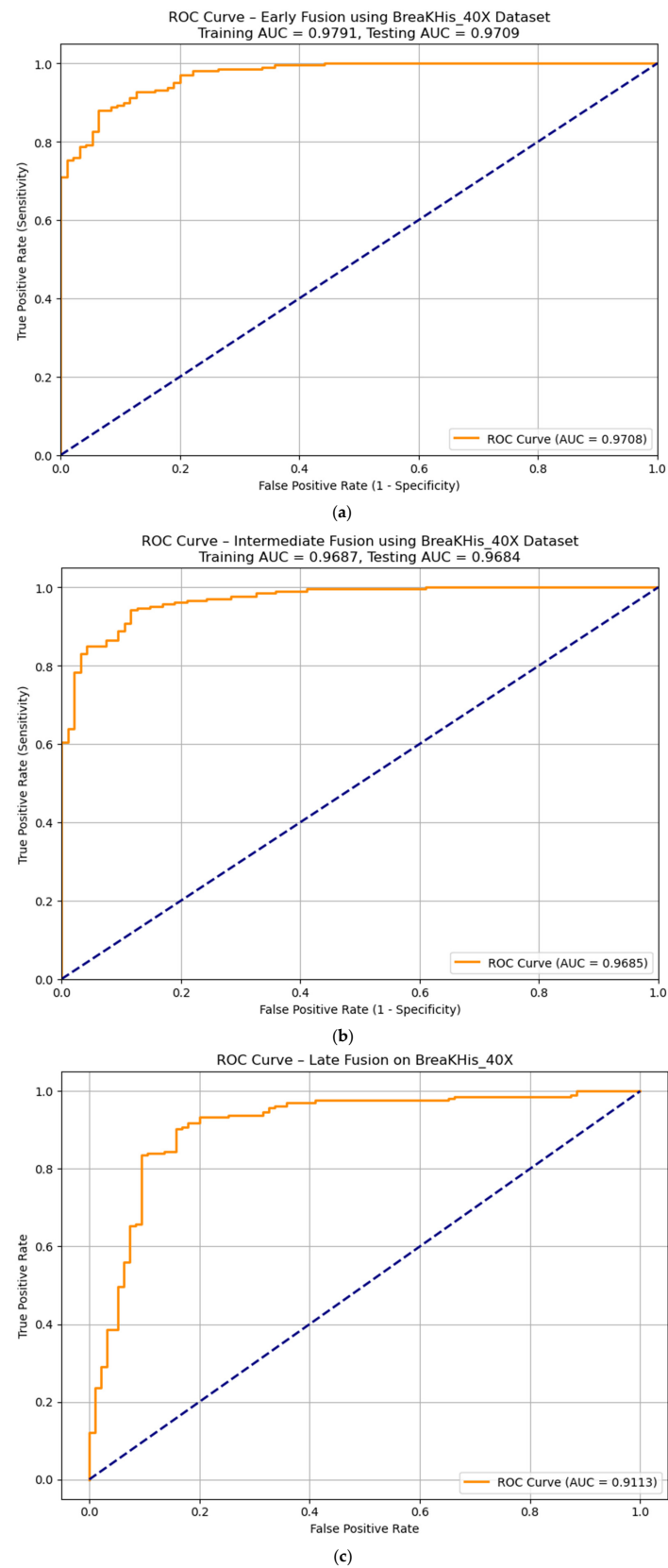


Figure 3. Cont.

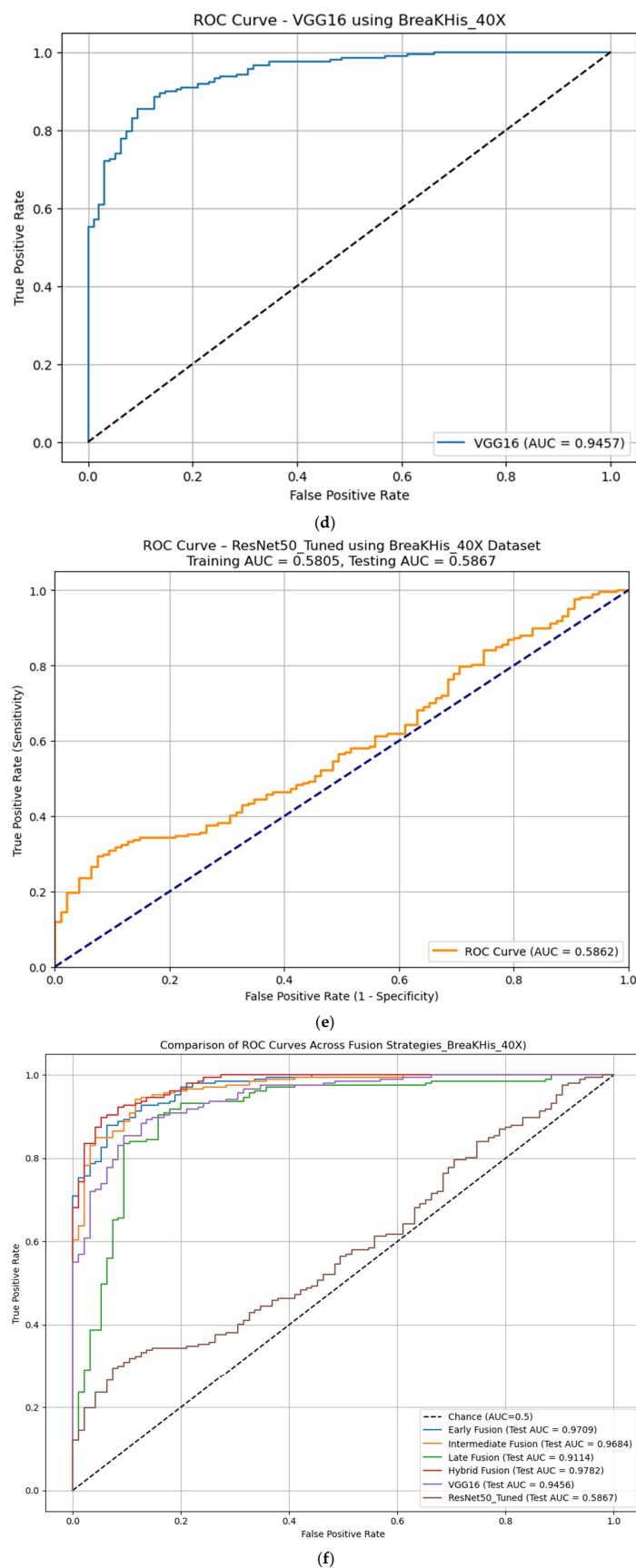


Figure 3. ROC curves of fusion-based and unimodal models for binary classification (BreakHis-40 $\times$): In panels (a–e), the dashed diagonal line indicates chance-level performance (random classifier; AUC = 0.5). (a) shows ROC for early fusion; (b) shows ROC for intermediate fusion; (c) shows ROC for late fusion; (d) shows ROC for VGG16; (e) shows ROC for ResNet50; and (f) shows all comparison ROCs for all models.

4.2. Performance of Fusion-Based and Unimodal Models for BreakHis-40× Multiclass (Eight-Class)

Unlike the binary classification task, multiclass classification is challenging due to the overlapping morphological patterns among the eight histological classes. Table 7 presents the dataset specification of BreakHis multiclassification.

Table 7. BreakHis_40× dataset specification (multiclassification).

Dataset BreakHis_40× (Multiclassification Eight Classes)	Total Size	Training	Validation	Testing
	1995	1393	295	307

Table 8 presents the accuracies obtained by the fusion-based and unimodal models.

Table 8. Performance of fusion-based and unimodal models for BreakHis_40× multiclass (eight-class).

Modality Type	Pre-Trained Model	Fusion Strategy	Accuracy (%)			AUC%		
			Training	Validation	Testing	Training	Validation	Testing
Unimodal	ResNet50	-	0.4336	0.4373	0.4267	0.6386	0.6272	0.6643
Unimodal	VGG16	-	0.7753	0.7119	0.7296	0.9662	0.9418	0.9581
Multi-pretrained	(ResNet50 & VGG16)	Early Fusion	0.7832	0.6915	0.7199	0.9667	0.9379	0.9600
Multi-pretrained	(ResNet50 & VGG16)	Intermediate Fusion	0.6289	0.6203	0.6547	0.9382	0.9193	0.9346
Multi-pretrained	(ResNet50 & VGG16)	Late Fusion	0.5549	0.5424	0.5700	0.9459	0.9196	0.9461

As seen in Table 8, the results demonstrate that combining features at an early stage performs better than other fusion strategies and unimodal models. The early fusion strategy achieved the highest AUC among all models, reaching 0.9600, supporting the effectiveness of fusing features from multi-pretrained models at an early stage. By contrast, ResNet50 alone achieved the lowest performance, with a testing accuracy of 0.4267 and an AUC of 0.6643. Intermediate fusion achieved moderate performance (testing accuracy = 0.6547; AUC = 0.9346). Similarly to the results obtained in the binary classification task, late fusion achieved the lowest performance compared to early and intermediate fusion. This result confirms that late fusion does not guarantee the integration of complementary features and struggles to learn correlations among the combined features [48,59]. Table 9 presents test performance on BreakHis (40×) with an eight-class dataset with additional metrics. Additionally, Tables A2–A6 (Appendix B) show a class-wise performance analysis to give more insight into class imbalance and morphological overlap.

Table 9. Test-set performance with additional metrics for BreakHis_40× multiclass (eight-class).

Model	Test Accuracy	Test AUC	Macro Precision	Macro Recall	Macro F1	Weighted Precision	Weighted Recall	Weighted F1
ResNet50 (single)	0.4267	0.6643	0.05	0.12	0.07	0.18	0.43	0.26
VGG16 (single)	0.7296	0.9581	0.74	0.64	0.67	0.73	0.73	0.71
Early Fusion	0.7199	0.9600	0.71	0.65	0.66	0.72	0.72	0.71
Intermediate Fusion	0.6547	0.9346	0.70	0.57	0.56	0.75	0.65	0.66
Late Fusion	0.5700	0.9461	0.86	0.35	0.42	0.74	0.57	0.50

Overall, VGG16 as a single model and the multi-pretrained model with an early fusion strategy show relatively balanced results across several classes. However, some overlapping classes (e.g., classes 3–6) demonstrated lower recall, ranging between 0.34 and 0.53, which is consistent with the expected similarity in histopathological patterns. Intermediate fusion

showed an improvement in sensitivity for certain classes (e.g., class 6 shows high recall (0.88), but it introduced more false positives, as indicated by reduced precision (0.22) for that class). In contrast, late fusion behaves conservatively, yielding high precision for some classes (e.g., classes 0, 4, 5, and 7) but lower recall for other classes. ResNet50 shows weak class-wise discrimination under the current configuration. Figures 4 and 5 illustrate the confusion matrix of the best-performing fusion strategies (early and intermediate fusion), followed by the ROC plots for each fusion model:

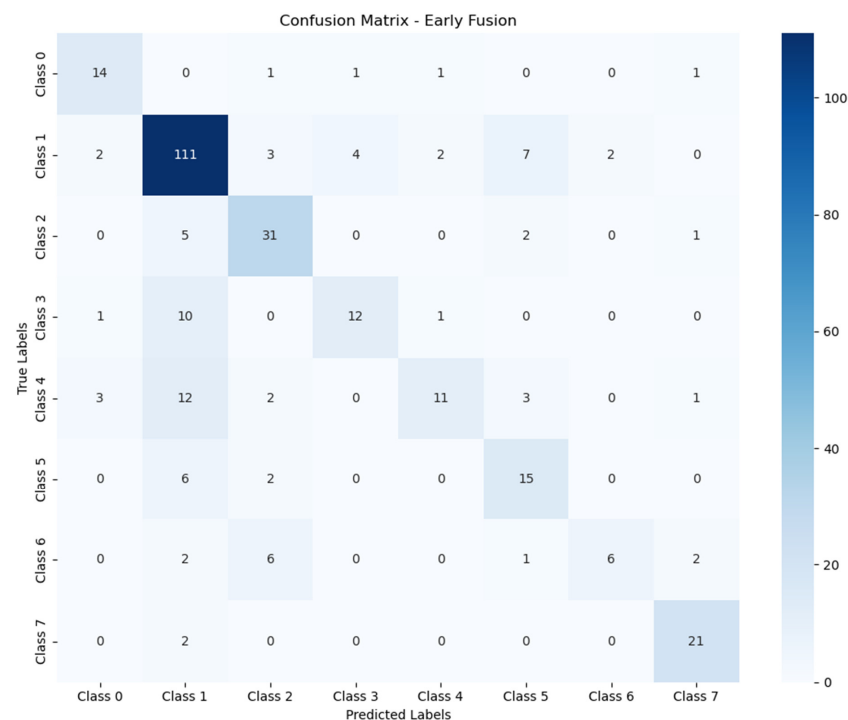


Figure 4. Confusion matrix of early fusion for BreakHis_40 $\times$ multiclass classification.

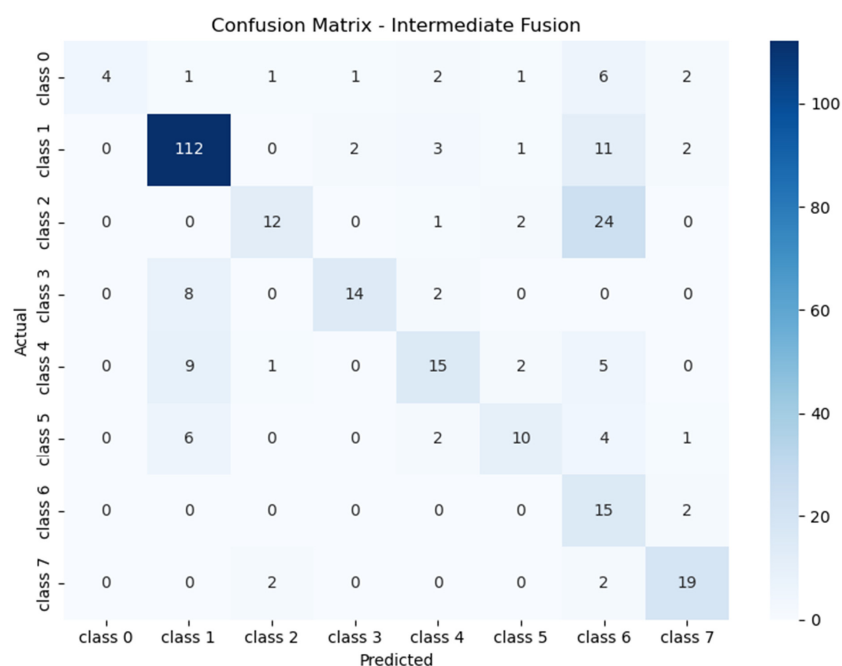
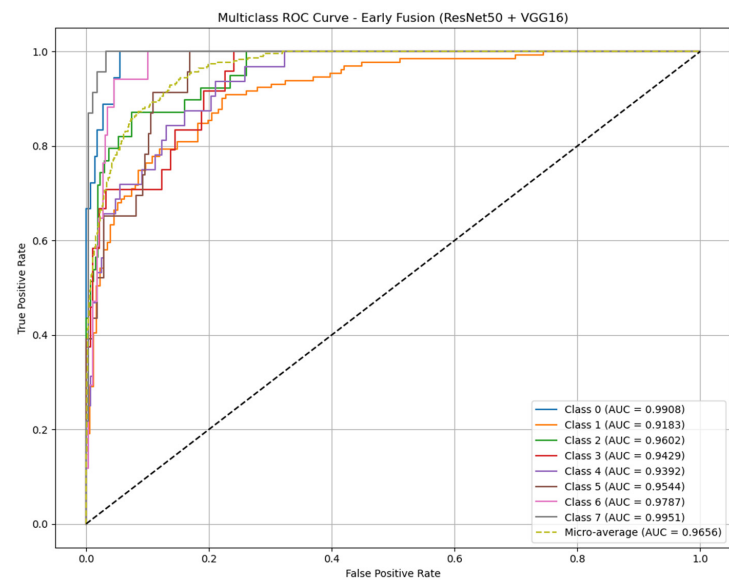
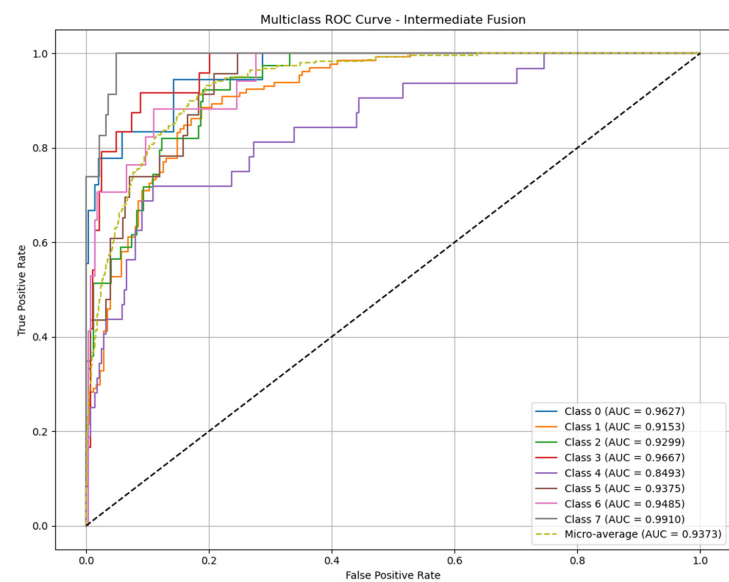


Figure 5. Confusion matrix of intermediate fusion for BreakHis_40 $\times$ multiclass classification.

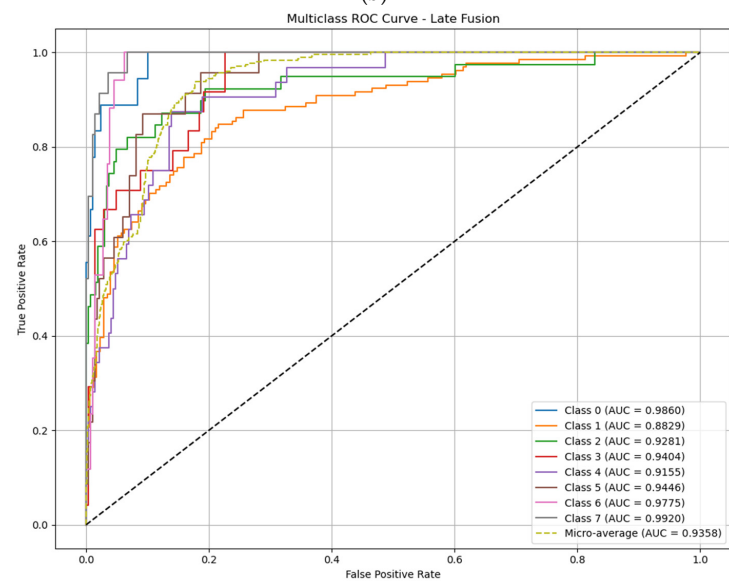
Figure 6 presents the ROC curves (Figure 6a–e) illustrating model performance on the BreakHis 40 $\times$ multiclass (eight-class) task:



(a)



(b)



(c)

Figure 6. Cont.

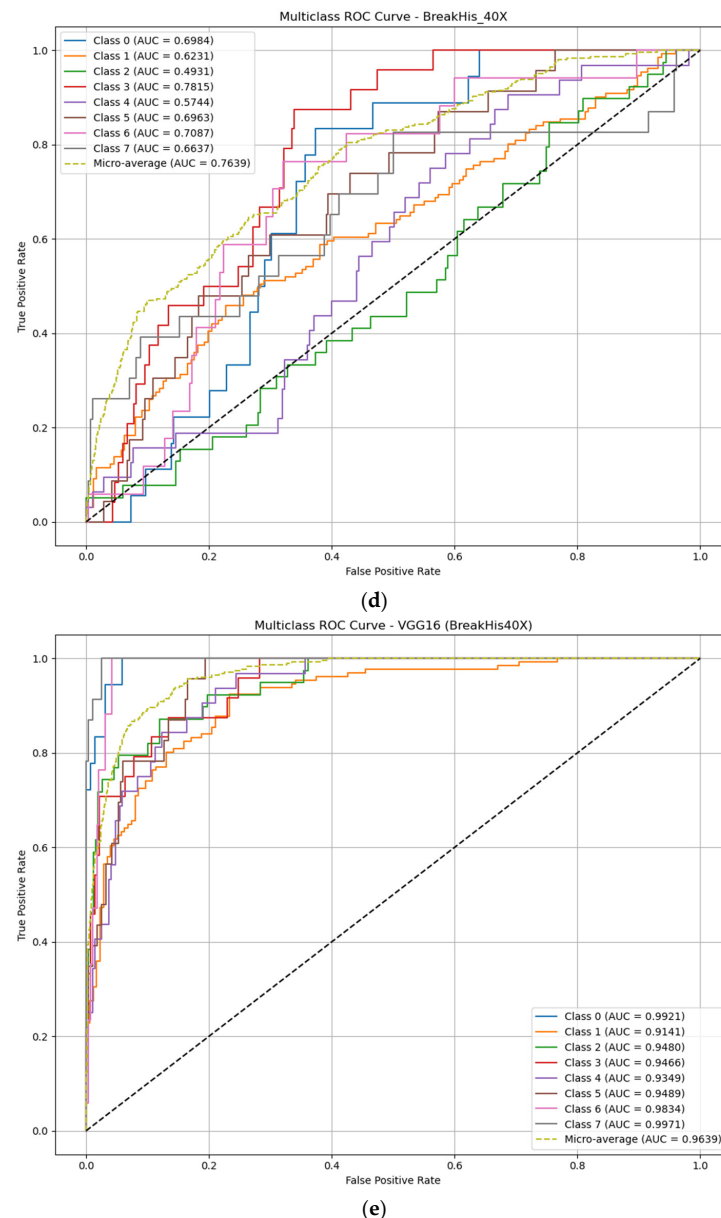


Figure 6. ROC curves of model performance for BreakHis_40 $\times$ multiclass (eight-class): the dashed diagonal line indicates chance-level performance (random classifier; AUC = 0.5) (a) shows ROC for early fusion; (b) shows ROC for intermediate fusion; (c) shows ROC for late fusion; (d) shows ROC for ResNet50; and (e) shows ROC for VGG16.

4.3. Performance of Fusion-Based and Unimodal Models for Multiclass: (RCCC) (Five-Class Grading)

Table 10 presents the dataset specification of RCCC, followed by Table 11, which presents the performance of fusion-based models and a single model for RCCC. As illustrated in Table 11, the performance on the (RCCC) grading task followed a similar trend to the BreakHis datasets. Similarly, early fusion achieved the best testing accuracy (0.8793) and the highest AUC (0.9889) across all RCCC experiments (multi-pretrained and unimodal models). These findings offer clear evidence that combining or fusing low- and high-level representations improves the model's ability to discriminate subtle differences or feature variations between RCCC grades. The cancer grading task requires careful recognition of subtle nuclear and cytoplasmic differences. ResNet50 alone on the RCCC grading dataset achieved the lowest performance, with a testing accuracy of 0.4331 and an AUC of 0.7344 and thus fails to capture these differences.

Table 10. Renal Clear Cell Carcinoma dataset specification.

Dataset (Multiclass Classification (5-Class Grading).	Total Size	Training	Validation	Testing
	4003	2800	598	605

Table 11. Performance of fusion-based and unimodal models for the Renal Clear Cell Carcinoma dataset.

Modality Type	Pre-Trained Model	Fusion Strategy	Accuracy (%)			AUC%		
			Training	Validation	Testing	Training	Validation	Testing
Unimodal	ResNet50	-	0.4521	0.4147	0.4331	0.7456	0.7167	0.7344
Unimodal	VGG16	-	0.8711	0.8679	0.8463	0.9863	0.9834	0.9821
Multi-pretrained	(ResNet50 & VGG16)	Early Fusion	0.9050	0.8946	0.8793	0.9926	0.9902	0.9889
Multi-pretrained	(ResNet50 & VGG16)	Intermediate Fusion	0.8829	0.8562	0.8645	0.9888	0.9853	0.9843
Multi-pretrained	(ResNet50 & VGG16)	Late Fusion	0.8746	0.8445	0.8479	0.9745	0.9680	0.9697

Table 12 presents the RCCC test-set performance with additional metrics:

Table 12. Renal Clear Cell Carcinoma test-set performance with additional metrics (multiclass; 5 grades/classes).

Model	Test Accuracy	Test AUC	Macro Precision	Macro Recall	Macro F1	Weighted Precision	Weighted Recall	Weighted F1
ResNet50 (single)	0.4331	0.7344	0.48	0.44	0.42	0.48	0.43	0.42
VGG16 (single)	0.8463	0.9821	0.85	0.85	0.84	0.85	0.85	0.85
Early Fusion	0.8793	0.9889	0.88	0.88	0.87	0.88	0.88	0.88
Intermediate Fusion	0.8645	0.9843	0.87	0.86	0.86	0.87	0.86	0.87
Late Fusion	0.8479	0.9697	0.85	0.84	0.84	0.85	0.85	0.84

Overall, the fusion-based models (early fusion and intermediate fusion) performed well on the test set. Early fusion achieved the best overall results (Macro-F1 0.87, Weighted-F1 0.88), followed closely by intermediate fusion (Macro-F1 0.86, Weighted-F1 0.88). In contrast, ResNet50 performed much worse (Macro-F1 and 0.42), indicating weak class discrimination under the current configuration.

Class-wise analysis, Appendix C, Tables A7–A11, shows that the fusion models performed consistently well for Grades 0, 1, and 4 (for example, early fusion F1-scores: 0.97, 0.91, and 0.92, respectively). In comparison, Grade 2 was the most challenging class, with lower recall across all fusion strategies (early 0.65, intermediate 0.77, late 0.63), suggesting morphological overlap with adjacent grades. Overall, fusion strategies improved grading robustness and reduced errors for most grades. Misclassifications were mainly concentrated in the intermediate grades, which is expected due to histopathological similarity between neighboring severity levels. Figures 7 and 8 illustrate the confusion matrix of best-performing fusion strategies (early and intermediate fusion).

Figure 9 presents the ROC curves (Figure 9a–e) illustrating model performance on RCCC (5-grades) task:

In addition to the dataset-specific results reported above, Table 13 provides a compact cross-task summary of the best test-set performance and the absolute improvements (Δ) versus the strongest unimodal baseline (VGG16). Confidence intervals (95%) were estimated using bootstrap resampling of test predictions, and Δ values were computed using paired bootstrap resampling of the same test indices ($B = 2000$). This summary facilitates direct comparison of fusion effectiveness across tasks.

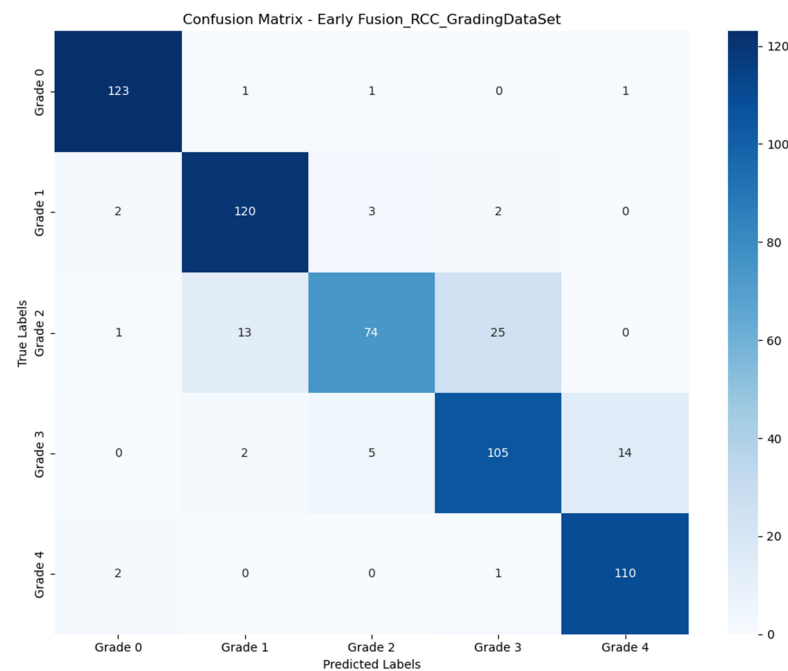


Figure 7. Confusion matrix of early fusion for RCCC dataset.

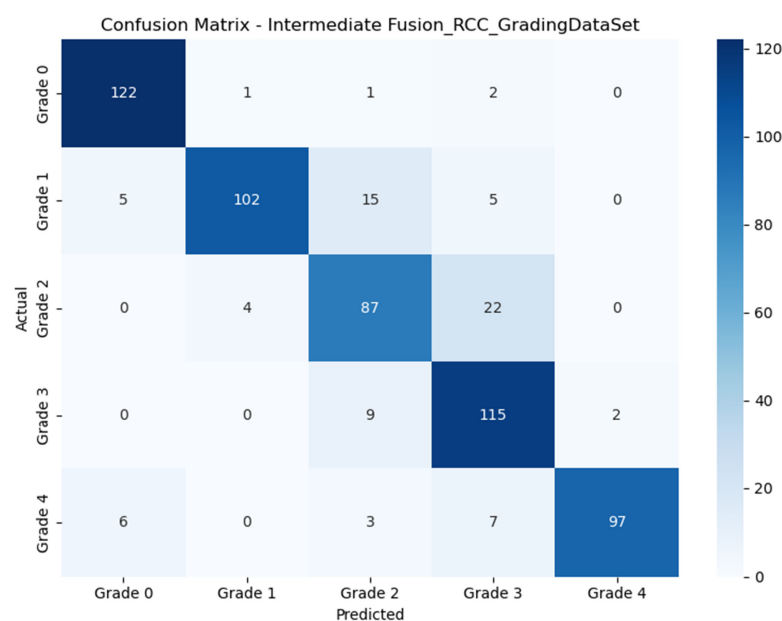
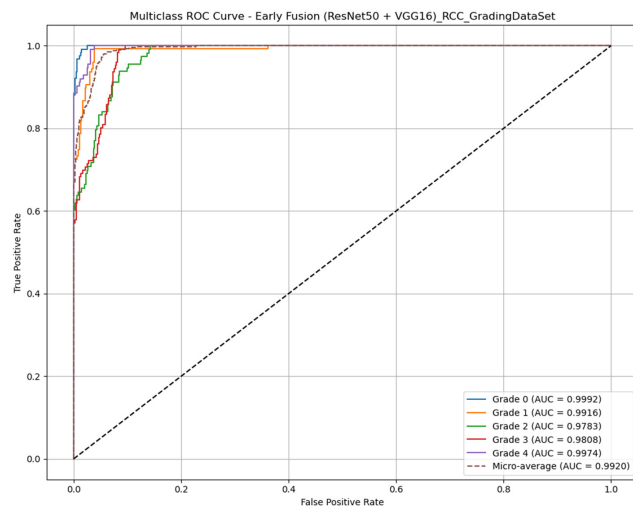


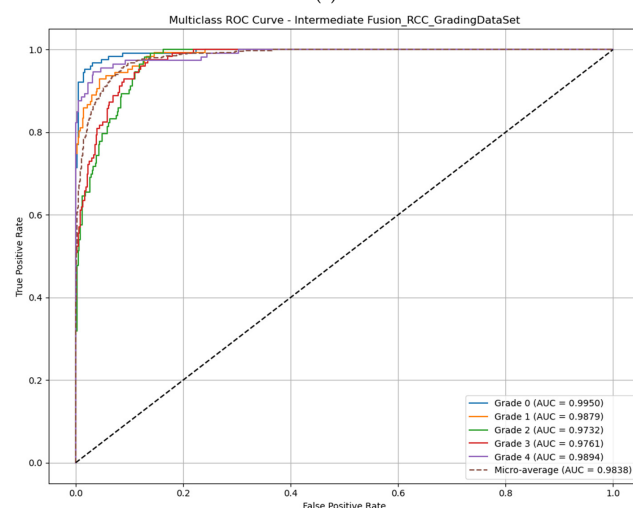
Figure 8. Confusion matrix of intermediate fusion for RCCC dataset.

Overall, the comparative evaluation across all datasets demonstrates that combining representations from multiple pretrained models can enhance performance in histopathology image classification. Across nearly all experiments, early fusion consistently achieved the best results, outperforming single-backbone baselines and the other fusion strategies. In contrast, late fusion (decision-level combination) generally yielded the weakest performance, suggesting that it is less effective at exploiting complementary representations learned by each pretrained model. Nevertheless, late fusion still outperformed the ResNet50 single-model baseline, indicating that incorporating fusion—even at the decision level—can provide measurable benefits. Collectively, these findings highlight the importance of adopting feature-level fusion to integrate multi-pretrained representations within a single-modality histopathology setting.

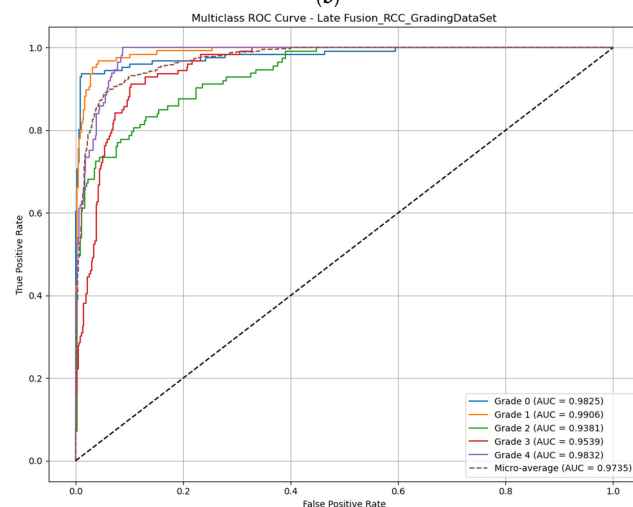
The proposed MPDLF model could be considered as a second-reader/triage tool to flag high-risk or low-confidence cases and to output calibrated probabilities with uncertainty to support reliable decision-making. However, such use requires rigorous validation and safety assessment. Recent evidence also suggests that LLM-based image analysis can support clinical decision-making in medicine (e.g., dermatology), motivating analogous exploration in computational pathology after appropriate validation and safeguarding [78,79].



(a)



(b)



(c)

Figure 9. Cont.

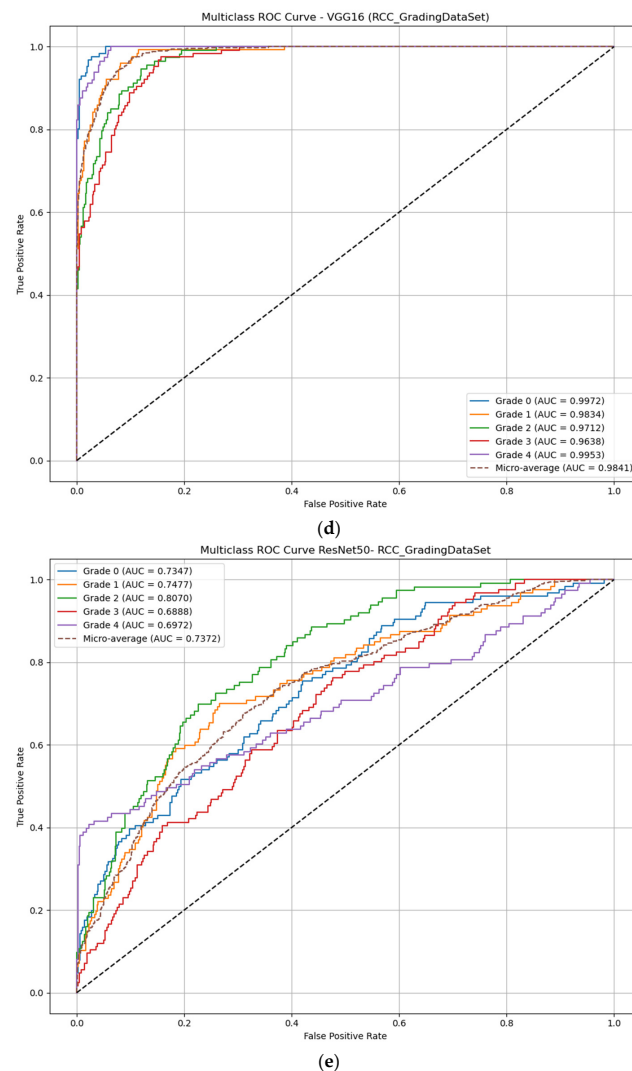


Figure 9. ROC curves of model performance for RCCC (multiclass classification (5 classes)): the dashed diagonal line indicates chance-level performance (random classifier; AUC = 0.5) (a) shows ROC for early fusion; (b) shows ROC for intermediate fusion; (c) shows ROC for late fusion; (d) shows ROC for VGG16; and (e) shows ROC for ResNet50.

Table 13. Compact cross-task summary of best test-set performance (accuracy and AUC) with 95% bootstrap confidence intervals and absolute deltas (Δ) versus the strongest unimodal baseline (VGG16).

Task	Best Model	Accuracy (95% CI)	AUC (95% CI)	Baseline (VGG16) Accuracy (95% CI)	Baseline AUC (95% CI)	Δ Acc (95% CI)	Δ AUC (95% CI)
BreakHis binary (40 $\times$; N = 302)	Early Fusion	0.9070 (0.8742–0.9404)	0.9707 (0.9541–0.9844)	0.8905 (0.8543–0.9238)	0.9538 (0.9308–0.9729)	+0.0165 (0.0000–0.0364)	+0.0170 (0.0080–0.0271)
BreakHis multiclass (40 $\times$; 8 classes; N = 307)	VGG16	0.7306 (0.6808–0.7785)	0.9582 (0.9463–0.9695)	0.7306 (0.6808–0.7785)	0.9582 (0.9463–0.9695)	0.0000	0.0000
RCCC grading (5 classes; N = 605)	Early Fusion	0.8792 (0.8529–0.9041)	0.9895 (0.9859–0.9927)	0.8461 (0.8149–0.8744)	0.9822 (0.9768–0.9870)	+0.0331 (0.0083–0.0595)	+0.0072 (0.0040–0.0109)

Future work will strengthen the clinical relevance of this study by conducting additional validation, including patient/slide-level evaluation (when identifiers are available), external validation across scanners/laboratories and staining protocols, and robustness analysis under strain and magnification variability. Finally, although pretrained backbones

were kept fully frozen to ensure a controlled comparison, partial fine-tuning may further improve performance and will be investigated in future work.

5. Conclusions

Shifting toward multimodality can improve computational pathology by leveraging complementary representations from multiple pretrained models, even when learning from a single data source. This work investigated multi-pretrained, deep-learning-based fusion strategies for histopathology image classification across three datasets and diagnostic tasks: binary breast cancer classification, multiclass (eight-class) breast cancer classification, and five-class (RCCC) grading. The proposed fusion-based models were evaluated against a unimodal (single-model) baseline, demonstrating that the choice of fusion strategy (early, intermediate, or late) substantially influences the classification performance.

Across the evaluated settings, early fusion provided consistent gains over the strongest unimodal baseline (VGG16) for BreakHis binary classification (40×) and RCCC five-class grading, yielding statistically supported improvements in AUC and modest but positive improvements in accuracy under paired bootstrap confidence intervals. Early fusion benefits from combining complementary representations from ResNet50 and VGG16 at an early stage, enabling richer feature learning by integrating fine-grained texture and morphological cues with deeper semantic features. Intermediate fusion generally achieved moderate performance by integrating representations at a mid-network level, whereas late fusion, which operates at the decision level, was comparatively less effective in exploiting feature complementarity.

In contrast, for BreakHis multiclass classification, VGG16 remained the strongest model, indicating that the advantages of feature-level fusion are task-dependent and may vary with class separability, morphological overlap, and the complexity of the decision boundary. Overall, these findings highlight the importance of carefully selecting fusion designs when constructing multi-pretrained models for histopathology image analysis. Future work will extend the evaluation to additional magnification levels and large datasets and will explore advanced hybrid fusion strategies to improve robustness and generalizability across diverse histopathological settings.

Author Contributions: Conceptualization, F.A.; methodology, F.A.; software, F.A.; validation, F.A. and A.A.-H.; formal analysis, F.A.; investigation, F.A.; resources, F.A.; data curation, F.A.; writing—original draft preparation, F.A.; writing—review and editing, F.A.; visualization, F.A.; supervision, A.A.-H.; project administration, A.A.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data sets used in this study are publicly available. BreakHis-40× (binary classification), BreakHis-40× (multiclass classification (eight classes) [67], and Renal Clear Cell Carcinoma (RCC) (5 classes) [68].

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNNs	Convolutional Neural Networks
MPDLF	Multi-Pretrained Deep Learning Fusion
RCCC	Renal Clear Cell Carcinoma

CT	Computed Tomography
MRI	Magnetic Resonance Imaging
US	Ultrasound
PET	Positron Emission Tomography
WSI	Whole-Slide Images (WSI)
EMR	Electronic Medical Records
CTPA	CT Pulmonary Angiography Imaging
HER	Electronic Health Record
CrossVit	Cross-Attention Vision Transformer
ANN	Artificial Neural Network

Appendix A

Class-wise tables (precision/recall/F1 with support) for BreakHis (40× for binary classification):

Table A1. BreakHis (40×) binary class-wise test performance (precision/recall/F1, with supports).

Model	Class	Support	Precision	Recall	F1-Score
ResNet50	0.0	95	0.00	0.00	0.00
	1.0	207	0.69	1.00	0.81
VGG16	0.0	95	0.80	0.79	0.79
	1.0	207	0.90	0.91	0.91
Early Fusion	0.0	95	0.85	0.86	0.85
	1.0	207	0.94	0.93	0.93
Intermediate Fusion	0.0	95	0.75	0.91	0.82
	1.0	207	0.95	0.86	0.91
Late Fusion	0.0	95	0.89	0.66	0.76
	1.0	207	0.86	0.96	0.91

Appendix B

Class-wise tables (precision/recall/F1 with support) for BreakHis (40× eight classes):

Table A2. Class-wise test performance BreakHis (40×) and ResNet50 (single model).

Class	Support	Precision	Recall	F1-Score
0	18	0.00	0.00	0.00
1	131	0.43	1.00	0.60
2	39	0.00	0.00	0.00
3	24	0.00	0.00	0.00
4	32	0.00	0.00	0.00
5	23	0.00	0.00	0.00
6	17	0.00	0.00	0.00
7	23	0.00	0.00	0.00

Table A3. Class-wise test performance BreakHis (40×) VGG16 (single model).

Class	Support	Precision	Recall	F1-Score
0	18	0.71	0.83	0.77
1	131	0.71	0.91	0.80

Table A3. *Cont.*

Class	Support	Precision	Recall	F1-Score
2	39	0.82	0.69	0.75
3	24	0.83	0.42	0.56
4	32	0.71	0.38	0.49
5	23	0.65	0.48	0.55
6	17	0.75	0.53	0.62
7	23	0.75	0.91	0.82

Table A4. BreakHis (40×) early fusion class-wise test performance.

Class	Support	Precision	Recall	F1-Score
0	18	0.70	0.78	0.74
1	131	0.75	0.85	0.80
2	39	0.69	0.79	0.74
3	24	0.71	0.50	0.59
4	32	0.73	0.34	0.47
5	23	0.54	0.65	0.59
6	17	0.75	0.35	0.48
7	23	0.81	0.91	0.86

Table A5. BreakHis (40×) intermediate fusion class-wise test performance.

Class	Support	Precision	Recall	F1-Score
0	18	1.00	0.22	0.36
1	131	0.82	0.85	0.84
2	39	0.75	0.31	0.44
3	24	0.82	0.58	0.68
4	32	0.60	0.47	0.53
5	23	0.62	0.43	0.51
6	17	0.22	0.88	0.36
7	23	0.73	0.83	0.78

Table A6. BreakHis (40×)—late fusion class-wise test performance.

Class	Support	Precision	Recall	F1-Score
0	18	1.00	0.50	0.67
1	131	0.50	0.98	0.66
2	39	0.94	0.44	0.60
3	24	0.80	0.17	0.28
4	32	1.00	0.09	0.17
5	23	1.00	0.09	0.16
6	17	0.60	0.18	0.27
7	23	1.00	0.35	0.52

Appendix C

Class-wise tables (precision/recall/F1 with support) for RCCC:

Table A7. ResNet50 (single model) RCCC class-wise performance.

Grade	Support	Precision	Recall	F1-Score
0	126	0.62	0.32	0.42
1	127	0.43	0.51	0.47
2	113	0.34	0.75	0.47
3	126	0.36	0.20	0.26
4	113	0.65	0.42	0.51

Table A8. VGG16 (single model) RCCC class-wise performance.

Grade	Support	Precision	Recall	F1-Score
0	126	0.93	0.94	0.93
1	127	0.88	0.83	0.86
2	113	0.83	0.71	0.77
3	126	0.74	0.81	0.78
4	113	0.85	0.94	0.89

Table A9. Early fusion—RCCC class-wise performance.

Grade	Support	Precision	Recall	F1-Score
0	126	0.96	0.98	0.97
1	127	0.88	0.94	0.91
2	113	0.89	0.65	0.76
3	126	0.79	0.83	0.81
4	113	0.88	0.97	0.92

Table A10. Intermediate fusion—RCCC class-wise performance.

Grade	Support	Precision	Recall	F1-Score
0	126	0.92	0.97	0.94
1	127	0.95	0.80	0.87
2	113	0.76	0.77	0.76
3	126	0.76	0.91	0.83
4	113	0.98	0.86	0.92

Table A11. Late fusion—RCCC class-wise performance.

Grade	Support	Precision	Recall	F1-Score
0	126	0.92	0.93	0.92
1	127	0.87	0.94	0.91
2	113	0.88	0.63	0.73
3	126	0.76	0.81	0.78
4	113	0.83	0.91	0.87

References

1. American Cancer Society The Global Cancer Burden | American Cancer Society. Available online: <https://www.cancer.org/about-us/our-global-health-work/global-cancer-burden.html> (accessed on 10 August 2025).
2. WHO Cancer. Available online: <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed on 10 August 2025).
3. Dattani, S.; Samborska, V.; Ritchie, H.; Roser, M. Cancer—Our World in Data. Available online: <https://ourworldindata.org/cancer> (accessed on 10 August 2025).
4. WHO Global Cancer Burden Growing, Amidst Mounting Need for Services. Available online: <https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services> (accessed on 10 August 2025).
5. Wang, J.; Hesketh, R.L.; Gore, J.C.; Brindle, K.M. The Need for Evidence-Based, Outcome-Focused Medical Imaging Research for Cancer Management. *npj Imaging* **2025**, *3*, 1–4. [\[CrossRef\]](#)
6. Fass, L. Imaging and Cancer: A Review. *Mol. Oncol.* **2008**, *2*, 115–152. [\[CrossRef\]](#)
7. Ullah Khan, S.; Ahmad Khan, M.; Azhar, M.; Khan, F.; Lee, Y.; Javed, M. Multimodal Medical Image Fusion towards Future Research: A Review. *J. King Saud Univ. Comput. Inf. Sci.* **2023**, *35*, 7361–7382. [\[CrossRef\]](#)
8. Schillaci, O.; Scimeca, M.; Toschi, N.; Bonfiglio, R.; Urbano, N.; Bonanno, E. Combining Diagnostic Imaging and Pathology for Improving Diagnosis and Prognosis of Cancer. *Contrast Media Mol. Imaging* **2019**, *2019*, 6. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Thatha, V.N.; Karthik, M.G.; Gaddam, V.G.; Krishna, D.P.; Venkataramana, S.; Lella, K.K.; Pamula, U. Histopathological Image Based Breast Cancer Diagnosis Using Deep Learning and Bio Inspired Optimization. *Sci. Rep.* **2025**, *15*, 1–24. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Ma, L.; Guo, H.; Zhao, Y.; Liu, Z.; Wang, C.; Bu, J.; Sun, T.; Wei, J. Liquid Biopsy in Cancer Current: Status, Challenges and Future Prospects. *Signal Transduct. Target. Ther.* **2024**, *9*, 336. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Mallick, I. An Overview of Histopathology. Available online: <https://www.verywellhealth.com/histopathology-2252152> (accessed on 10 August 2025).
12. Gurcan, M.N.; Boucheron, L.E.; Can, A.; Madabhushi, A.; Rajpoot, N.M.; Yener, B. Histopathological Image Analysis: A Review. *IEEE Rev. Biomed. Eng.* **2009**, *2*, 147–171. [\[CrossRef\]](#)
13. Wang, Y.; Coudray, N.; Zhao, Y.; Li, F.; Hu, C.; Zhang, Y.; Imoto, S.; Tsirigos, A.; Webb, G.I.; Daly, R.J.; et al. HEAL: An Automated Deep Learning Framework for Cancer Histopathology Image Analysis. *Bioinformatics* **2021**, *37*, 4291–4295. [\[CrossRef\]](#)
14. Bülent, Y. Histopathology Image Classification: Highlighting the Gap between Manual Analysis and AI Automation. *Front. Oncol.* **2024**, *13*, 1325271. [\[CrossRef\]](#)
15. Laak, J.; Litjens, G.; Ciompi, F. To the Clinic. *Nat. Med.* **2021**, *27*, 775–784. [\[CrossRef\]](#)
16. Mezei, T.; Kolcs, M.; Gurzu, S. Image Analysis in Histopathology and Cytopathology: From Early Days to Current Perspectives. *J. Imaging* **2024**, *10*, 252. [\[CrossRef\]](#)
17. Fu, Y.; Huang, Z.; Deng, X.; Xu, L.; Liu, Y. Artificial Intelligence in Lymphoma Histopathology: Systematic Review. *J. Med. Internet Res.* **2025**, *27*, e62851. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Zadvornyi, T. Digital Pathology As an Innovative Tool for Improving Cancer Diagnosis and Treatment. *Exp. Oncol.* **2024**, *46*, 289–294. [\[CrossRef\]](#)
19. Lakshmanaprabu, S.K.; Mohanty, S.N.; Shankar, K.; Arunkumar, N.; Ramirez, G. Optimal Deep Learning Model for Classification of Lung Cancer on CT Images. *Future Gener. Comput. Syst.* **2019**, *92*, 374–382. [\[CrossRef\]](#)
20. Riquelme, D.; Akhloufi, M.A. Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans. *Ai* **2020**, *1*, 28–67. [\[CrossRef\]](#)
21. Golatkar, A.; Anand, D.; Sethi, A. Classification of Breast Cancer Histology Using Deep Learning. In Proceedings of the Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, 27–29 June 2018; pp. 837–844.
22. Liu, M.; Hu, L.; Tang, Y.; Wang, C.; He, Y.; Zeng, C.; Lin, K.; He, Z.; Huo, W. A Deep Learning Method for Breast Cancer Classification in the Pathology Images. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5025–5032. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Saleh, A.; Sukaik, R.; Abu-Naser, S.S. Brain Tumor Classification Using Deep Learning. In Proceedings of the 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech), Gaza, Palestine, 28–29 August 2020; pp. 131–136.
24. Mohsen, H.; El-Dahshan, E.-S.A.; El-Horbaty, E.-S.M.; Salem, A.-B.M. Classification Using Deep Learning Neural Networks for Brain Tumors. *Future Comput. Inform. J.* **2018**, *3*, 68–71. [\[CrossRef\]](#)
25. Abdelsamea, M.M.; Zidan, U.; Senousy, Z.; Gaber, M.M.; Rakha, E.; Ilyas, M. A Survey on Artificial Intelligence in Histopathology Image Analysis. *WIREs Data Min. Knowl. Discov.* **2022**, *12*, e1474. [\[CrossRef\]](#)
26. Imran, M.T.; Shafi, I.; Ahmad, J.; Fasih, M.; Butt, U.; Villar, S.G. Virtual Histopathology Methods in Medical Imaging—A Systematic Review. *BMC Med. Imaging* **2024**, *24*, 318. [\[CrossRef\]](#)
27. Mubarak, M. Move from Traditional Histopathology to Digital and Computational Pathology: Are We Ready? *Indian J. Nephrol.* **2022**, *32*, 414–415. [\[CrossRef\]](#)

28. Pressman, C. Realizing the Potential of Digital Pathology. Ph.D. Thesis, University of Texas Health Science Center, Houston, TX, USA, 2024.
29. Yao, I.Z.; Dong, M.; Hwang, W.Y.K. Deep Learning Applications in Clinical. *MAYO Clin. Proc. Digit. Health* **2025**, *3*, 100253. [\[CrossRef\]](#)
30. Luan, H.; Yang, K.; Hu, T.; Hu, J.; Liu, S.; Li, R.; He, J. Review of Deep Learning-Based Pathological Image Classification: From Task-Specific Models to Foundation Models. *Future Gener. Comput. Syst.* **2025**, *164*, 107578. [\[CrossRef\]](#)
31. Huang, T.; Huang, X.; Yin, H. Deep Learning Methods for Improving the Accuracy and Efficiency of Pathological Image Analysis. *Sci. Prog.* **2025**, *108*, 1–34. [\[CrossRef\]](#)
32. Ma, Y.; Jamdade, S.; Konduri, L.; Sailem, H. AI in Histopathology Explorer for Comprehensive Analysis of the Evolving AI Landscape in Histopathology. *npj Digit. Med.* **2025**, *8*, 156. [\[CrossRef\]](#)
33. Unger, M.; Kather, J.N. Deep Learning in Cancer Genomics and Histopathology. *Genome Med.* **2024**, *16*, 44. [\[CrossRef\]](#)
34. Noori Mirtaheri, P.; Akhbari, M.; Najafi, F.; Mehrabi, H.; Babapour, A.; Rahimian, Z.; Rigi, A.; Rahbarbaghbani, S.; Mobaraki, H.; Masoumi, S.; et al. Performance of Deep Learning Models for Automatic Histopathological Grading of Meningiomas: A Systematic Review and Meta-Analysis. *Front. Neurol.* **2025**, *16*, 1536751. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Ramamoorthy, P.; Reddy, B.; Reddy, R.; Askar, S.; Abouhawwash, M. Histopathology-Based Breast Cancer Prediction Using Deep Learning Methods for Healthcare Applications. *Front. Oncol.* **2024**, *14*, 1300997. [\[CrossRef\]](#)
36. Couture, H. Deep Learning-Based Prediction of Molecular Tumor Biomarkers from H & E: A Practical Review. *J. Pers. Med.* **2022**, *12*, 2022. [\[CrossRef\]](#)
37. Tiwari, A.; Ghose, A.; Hasanova, M.; Faria, S.S.; Mohapatra, S.; Adeleke, S.; Boussios, S. The Current Landscape of Artificial Intelligence in Computational Histopathology for Cancer Diagnosis. *Discov. Oncol.* **2025**, *16*, 438. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Cooper, M.; Ji, Z.; Krishnan, R.G. Machine Learning in Computational Histopathology: Challenges and Opportunities. *Genes Chromosom. Cancer* **2023**, *62*, 540–556. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Huang, Q.; Wu, S.; Ou, Z.; Gao, Y. Computational Pathology: A Comprehensive Review of Recent Developments in Digital and Intelligent Pathology. *Intell. Oncol.* **2025**, *1*, 139–159. [\[CrossRef\]](#)
40. Komura, D.; Ochi, M.; Ishikawa, S. Machine Learning Methods for Histopathological Image Analysis: Updates in 2024. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 383–400. [\[CrossRef\]](#)
41. Das, H.S.; Borah, K.; Bora, K. Ensemble Learning Approach for Detecting Breast Invasive Ductal Carcinoma from Histopathological Images. *Pathol. Pract.* **2025**, *272*, 156041. [\[CrossRef\]](#)
42. Ravi, P.; Balasundaram, J.K.; Chinnappan, R.; Ramasamy, S. Comparative Analysis of Transfer Learning Models for Breast Cancer Detection: Leveraging Pre-Trained Networks for Enhanced Diagnostic Accuracy. *Biomed. Pharmacol. J.* **2025**, *18*, 1343–1352. [\[CrossRef\]](#)
43. Rahaman, M.M.; Millar, E.K.A.; Meijering, E. Generalized Deep Learning for Histopathology Image Classification Using Supervised Contrastive Learning. *J. Adv. Res.* **2024**, *75*, 389–404. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Li, Y.; El Habib Daho, M.; Conze, P.H.; Zeghlache, R.; Le Boité, H.; Tadayoni, R.; Cochener, B.; Lamard, M.; Quéllec, G. A Review of Deep Learning-Based Information Fusion Techniques for Multimodal Medical Image Classification. *Comput. Biol. Med.* **2024**, *177*, 108635. [\[CrossRef\]](#)
45. Cherkottu, S.B.; Vijayan, S.; Mahil, J.; Chembukkavu, J.; Gnanaprakash, V.; Arshath Raja, R.; Yuvaraj, N. AI-Driven Ensemble Deep Learning Framework for Automated Neurological Disorder Diagnosis from MRI Scans. *J. Neonatal Surg.* **2025**, *14*, 187–197. [\[CrossRef\]](#)
46. Addisu, E.G.; Yirga, T.G.; Yirga, H.G.; Yehuala, A.D. Transfer Learning-Based Hybrid VGG16-Machine Learning Approach for Heart Disease Detection with Explainable Artificial Intelligence. *Front. Artif. Intell.* **2025**, *8*, 1–18. [\[CrossRef\]](#)
47. Kumar, R.; Pan, C.T.; Lin, Y.M.; Yow-Ling, S.; Chung, T.S.; Janesha, U.G.S. Enhanced Multi-Model Deep Learning for Rapid and Precise Diagnosis of Pulmonary Diseases Using Chest X-Ray Imaging. *Diagnostics* **2025**, *15*, 1–32. [\[CrossRef\]](#)
48. Kumar, S.; Ivanova, O.; Melyokhin, A.; Tiwari, P. Informatics in Medicine Unlocked Deep-Learning-Enabled Multimodal Data Fusion for Lung Disease Classification. *Inform. Med. Unlocked* **2023**, *42*, 101367. [\[CrossRef\]](#)
49. Mejri, S.; Oueslati, A.E. Dermoscopic Images Classification Using Pretrained VGG-16 and ResNet-50 Models. In *Proceedings of the 2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP), Sousse, Tunisia, 11–13 July 2024*; IEEE: Piscataway, NJ, USA, 2024; Volume 1, pp. 342–347. [\[CrossRef\]](#)
50. Lipkova, J.; Chen, R.J.; Chen, B.; Lu, M.Y.; Barbieri, M.; Shao, D.; Vaidya, A.J.; Chen, C.; Zhuang, L.; Williamson, D.F.K.; et al. Artificial Intelligence for Multimodal Data Integration in Oncology. *Cancer Cell* **2022**, *40*, 1095–1110. [\[CrossRef\]](#)
51. Steyaert, S.; Pizurica, M.; Nagaraj, D.; Khandelwal, P.; Hernandez-Boussard, T.; Gentles, A.J.; Gevaert, O. Multimodal Data Fusion for Cancer Biomarker Discovery with Deep Learning. *Nat. Mach. Intell.* **2023**, *5*, 351–362. [\[CrossRef\]](#)
52. Yang, H.; Yang, M.; Chen, J.; Yao, G.; Zou, Q.; Jia, L. Multimodal Deep Learning Approaches for Precision Oncology: A Comprehensive Review. *Brief. Bioinform.* **2025**, *26*, bbae699. [\[CrossRef\]](#) [\[PubMed\]](#)

53. Ali, F.N.; Evgin, I. A Comprehensive Investigation of Multimodal Deep Learning Fusion Strategies for Breast Cancer Classification. *Artif. Intell. Rev.* **2024**, *57*, 327. [\[CrossRef\]](#)
54. Huang, S.C.; Pareek, A.; Zamanian, R.; Banerjee, I. Multimodal Fusion with Deep Neural Networks for Leveraging CT Imaging and Electronic Health Record: A Case-Study in Pulmonary Embolism Detection. *Sci. Rep.* **2020**, *10*, 22147. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Guarrasi, V.; Mogensen, K.; Tassinari, S.; Qvarlander, S.; Soda, P. Timing Is Everything: Finding the Optimal Fusion Points in Multimodal Medical Imaging. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN), Rome, Italy, 30 June–5 July 2025*; IEEE: Piscataway, NJ, USA, 2025.
56. Guarrasi, V.; Aksu, F.; Maria, C.; Di, F.; Rofena, A.; Ruffini, F.; Soda, P. A Systematic Review of Intermediate Fusion in Multimodal Deep Learning for Biomedical Applications. *Image Vis. Comput.* **2025**, *158*, 105509. [\[CrossRef\]](#)
57. Shirae, S.; Debsarkar, S.S.; Kawanaka, H.; Aronow, B.; Prasath, V.B.S. Multimodal Ensemble Fusion Deep Learning Using Histopathological Images and Clinical Data for Glioma Subtype Classification. *IEEE Access* **2025**, *13*, 57780–57797. [\[CrossRef\]](#)
58. Steyaert, S.; Qiu, Y.L.; Zheng, Y.; Mukherjee, P.; Vogel, H. Multimodal Data Fusion of Adult and Pediatric Brain Tumors with Deep Learning. *medRxiv* **2022**. [\[CrossRef\]](#)
59. Cahan, N.; Klang, E.; Marom, E.M.; Soffer, S.; Barash, Y.; Burshtein, E.; Konen, E.; Greenspan, H. Multimodal Fusion Models for Pulmonary Embolism Mortality Prediction. *Sci. Rep.* **2023**, *13*, 7544. [\[CrossRef\]](#)
60. Zheng, Y.; Li, C.; Zhou, X.; Chen, H.; Xu, H.; Li, Y.; Zhang, H.; Li, X.; Sun, H.; Huang, X.; et al. Application of Transfer Learning and Ensemble Learning in Image-Level Classification for Breast Histopathology. *Intell. Med.* **2023**, *3*, 115–128. [\[CrossRef\]](#)
61. Chakravarthy, S.; Surbhi, N.B.; Khan, B.; Mahesh, V.V.K.T.R. Multi-Class Breast Cancer Classification Using CNN Features Hybridization. *Int. J. Comput. Intell. Syst.* **2024**, *7*, 191. [\[CrossRef\]](#)
62. Li, J. Fusion Feature-Based Hybrid Methods for Diagnosing Oral Squamous Cell Carcinoma in Histopathological Images. *Front. Oncol.* **2025**, *15*, 1551876. [\[CrossRef\]](#)
63. Asif, S.; Wang, E.; Yang, V. HMDFF-Net: Hierarchical Multi-Scale Dilated Feature Fusion Network for Accurate Multiclass Diagnosis of Brain Tumors. *Expert Syst. Appl.* **2025**, *287*, 128272. [\[CrossRef\]](#)
64. Punarselvam, E.; Karthikeyan, M.; Jeyabharathy, S.; Suthar, M.B.; Chavda, K.; Suthar, T.M.; Babu, A.J.; Anusha, P. Quantitative Analysis of Deep Learning Frameworks with Integrated Feature Fusion for Cervical Cancer Classification and Detection. *J. Inf. Syst. Eng. Manag.* **2025**, *10*, 86–99. [\[CrossRef\]](#)
65. Sahu, P.K. LPIF-Based Image Enhancement and Hybrid Ensemble Models for Brain Tumor Detection. *Connect. Sci.* **2025**, *37*, 2518983. [\[CrossRef\]](#)
66. Gao, R.; Yang, Z.; Yuan, X.; Wang, Y.; Xia, Y. Features Fusion or Not: Harnessing Multiple Pathological Foundation Models Using Meta-Encoder for Downstream Tasks Fine-Tuning. *bioRxiv* **2025**. [\[CrossRef\]](#)
67. Pereira, M. BreakHis—Breast Cancer Histopathological Database. Available online: <https://data.mendeley.com/datasets/jxwvdwhpc2/1> (accessed on 19 June 2024).
68. Mahmood, T.; Wahid, A.; Hong, J.S.; Kim, S.G.; Park, K.R. A Novel Convolution Transformer-Based Network for Histopathology-Image Classification Using Adaptive Convolution and Dynamic Attention. *Eng. Appl. Artif. Intell.* **2024**, *135*, 108824. [\[CrossRef\]](#)
69. Chowdhury, A.A.; Mahmud, S.M.H.; Uddin, P.; Kadry, S.; Kim, J.; Nam, Y. Nuclei Segmentation and Classification from Histopathology Images Using Federated Learning for End-Edge Platform. *PLoS ONE* **2025**, *20*, e0322749. [\[CrossRef\]](#)
70. He, J.; Wang, X.; Wang, Z.; Xie, R.; Zhang, Z.; Liu, T.; Cai, Y.; Chen, L. Interpretable Deep Learning Method to Predict Wound Healing Progress Based on Collagen Fibers in Wound Tissue. *Comput. Biol. Med.* **2025**, *191*, 110110. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Liu, S.; Himel, G.M.S.; Wang, J. Breast Cancer Classification With Enhanced Interpretability: DALAResNet50 and DT Grad-CAM. *IEEE Access* **2024**, *12*, 196647–196659. [\[CrossRef\]](#)
72. Wang, D.; Guo, L.; Zhong, J.; Yu, H.; Tang, Y.; Peng, L.; Cai, Q.; Qi, Y.; Zhang, D. A Novel Deep-Learning Based Weighted Feature Fusion Architecture for Precise Classification of Pressure Injury. *Front. Physiol.* **2024**, *15*, 1304829. [\[CrossRef\]](#)
73. Gupta, R.K.; Manhas, J. Improved Classification of Cancerous Histopathology Images Using Color Channel Separation and Deep Learning. *J. Multimed. Inf. Syst.* **2021**, *8*, 175–182. [\[CrossRef\]](#)
74. Stahlschmidt, S.R.; Ulfenborg, B.; Synnergren, J.; Richard, S. Multimodal Deep Learning for Biomedical Data Fusion: A Review. *Brief. Bioinform.* **2022**, *23*, bbab569. [\[CrossRef\]](#)
75. Jiao, T.; Guo, C.; Feng, X.; Chen, Y.; Song, J. A Comprehensive Survey on Deep Learning Multi-Modal Fusion: Methods, Technologies and Applications. *Comput. Mater. Contin.* **2024**, *80*, 1–35. [\[CrossRef\]](#)
76. Manzoor, F.; Gupta, V.; Pinky, L.; Wang, Z. A Systematic Review of Multimodal Deep Learning and Machine Learning Fusion Techniques for Prostate Cancer Classification. *medRxiv* **2025**. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Nikolaou, N.; Salazar, D.; RaviPrakash, H.; Gonçalves, M.; Mulla, R.; Burlutskiy, N.; Markuzon, N.; Jacob, E. A Machine Learning Approach for Multimodal Data Fusion for Survival Prediction in Cancer Patients. *npj Precis. Oncol.* **2025**, *9*, 128. [\[CrossRef\]](#) [\[PubMed\]](#)

78. Boostani, M.; Bánvölgyi, A.; Zouboulis, C.C.; Goldfarb, N.; Suppa, M.; Goldust, M.; Lorincz, K.; Kiss, T.; Nádudvari, N.; Holló, P.; et al. Large Language Models in Evaluating Hidradenitis Suppurativa from Clinical Images. *J. Eur. Acad. Dermatol. Venereol. JEADV* **2025**, *39*, e1052–e1055. [[CrossRef](#)] [[PubMed](#)]
79. Boostani, M.; Bánvölgyi, A.; Goldust, M.; Cantisani, C.; Pietkiewicz, P. Diagnostic Performance of GPT-4o and Gemini Flash 2.0 in Acne and Rosacea. *Int. J. Dermatol.* **2025**, *64*, 1881–1882. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.