OPEN ACCESS
Journal of Biomedical and Techno Nanomaterials
Vol. 2 No. 4. August 2025, pp. 221-235 DOI. 10.70177/jbtn.v2i4. 2480

Research Article

ARTIFICIAL INTELLIGENCE IN MEDICINE: A DEEP LEARNING CONVOLUTIONAL NEURAL NETWORK FOR PATHOLOGICAL IMAGE ANALYSIS AND CANCER GRADING

James Smith[1], Oliver Harris[2], and Dito Anurogo[3]
[1] University of Sydney, Australia
[2] University of Queensland, Australia
[3] Faculty of Medicine and Health Sciences, Universitas Muhammadiyah Makassar, Indonesia

Corresponding Author:

James Smith,
Faculty of Medicine and Health, University of Sydney.
Camperdown NSW 2050, Australia
Email: jamessmith@gmail.com

Article Info
Received: February 6, 2025
Revised: May 18, 2025
Accepted: July 22, 2025
Online Version: August 24, 2025

Abstract

The histopathological analysis of tissue slides is the gold standard for cancer diagnosis and grading. However, this process is labor-intensive, time-consuming, and prone to inter-observer variability, which can affect clinical outcomes. The advent of artificial intelligence (AI), particularly deep learning, presents a transformative opportunity to enhance diagnostic precision and efficiency in pathology. This study aimed to develop, train, and validate a deep learning convolutional neural network (CNN) for the automated analysis of pathological images to accurately classify malignancies and provide reliable cancer grading. A robust CNN model was trained on a comprehensive, curated dataset of thousands of annotated digital histopathology slides from multiple cancer types. The model's performance was rigorously evaluated against the consensus diagnoses of expert pathologists using key metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC). Our developed CNN model demonstrated exceptional performance, achieving an overall accuracy of 98.7% in distinguishing malignant from benign tissues. For cancer grading, the model yielded a Cohen's Kappa score of 0.92, indicating almost perfect agreement with expert pathologists. The model also showed high robustness to variations in staining and image acquisition protocols. This research confirms that a deep learning CNN can function as a highly accurate and reliable tool for automated pathological image analysis and cancer grading. Integrating such AI systems into clinical workflows could significantly augment the capabilities of pathologists, leading to improved diagnostic consistency, reduced workload, and ultimately, better patient care.

Keywords: Artificial Intelligence, Deep Learning, Convolutional Neural Network, Cancer Grading, Computational Pathology

Journal Homepage https://research.adra.ac.id/index.php/jbtn
How to cite: Smith, J., Harris, O., & Anurogo, D. (2025). Artificial Intelligence in Medicine: A Deep Learning Convolutional Neural Network for Pathological Image Analysis and Cancer Grading. *Journal of Biomedical and Techno Nanomaterials*, *2*(4), 221–235. https://doi.org/10.70177/jbtn.v2i4. 2480
Published by: Yayasan Adra Karima Hubbi

INTRODUCTION

The diagnosis and management of cancer represent one of the most significant challenges in modern medicine, demanding exceptional precision at every stage of the clinical pathway (Avazov et al., 2025). Central to this process is the field of histopathology, which involves the microscopic examination of tissue specimens to identify cellular and architectural abnormalities indicative of malignancy (Binzagr et al., 2025). For over a century, this practice has remained the undisputed gold standard for definitive cancer diagnosis, providing the foundational information necessary for tumor classification, grading, and staging (Ahmad et al., 2025). The insights derived from pathological analysis directly inform prognostic assessments and guide the selection of appropriate therapeutic interventions, ranging from surgical resection to targeted chemotherapy and immunotherapy.

Pathological evaluation is a highly complex cognitive task that relies on the expertise of trained pathologists to interpret subtle morphological features within tissue slides (Park et al., 2023). These professionals meticulously analyze cellular size and shape, nuclear characteristics, mitotic activity, and the overall organization of tissue structures (Arman et al., 2023). This detailed visual inspection is translated into standardized grading systems, such as the Gleason score for prostate cancer or the Nottingham Histological Grade for breast cancer, which quantify the degree of tumor differentiation (van Dieren et al., 2023). These grading schemes are critical, as they correlate strongly with tumor aggressiveness and patient outcomes, thereby playing a pivotal role in clinical decision-making (Ancheta et al., 2025). The integrity of the entire oncological care process is, therefore, profoundly dependent on the accuracy and consistency of this initial histopathological assessment.

Recent technological advancements have propelled the field of pathology into the digital era, marked by the widespread adoption of whole-slide imaging (WSI) (Nopiyanti et al., 2023). This innovation enables the conversion of glass slides into high-resolution digital images that can be stored, shared, and analyzed computationally (Islam Sumon et al., 2023). The emergence of these massive, data-rich digital archives has created an unprecedented opportunity to leverage computational methods for enhancing diagnostic workflows (Luo et al., 2025). This paradigm shift towards digital pathology sets the stage for the integration of artificial intelligence (AI), particularly deep learning algorithms, which possess a remarkable capacity for learning intricate patterns from vast datasets of images, heralding a new frontier in diagnostic medicine.

The established gold standard of manual histopathological review, despite its foundational importance, is subject to inherent limitations that can impact diagnostic reliability (Cao et al., 2023). A primary challenge is the significant degree of inter-observer and intra-observer variability in the interpretation of tissue slides (Teresia et al., 202 C.E.). Different pathologists, or even the same pathologist at different times, may assign different grades to the same tumor specimen, a discrepancy that can lead to inconsistencies in patient prognosis and treatment planning (Gladis Pushparathi et al., 2025). This subjectivity arises from the qualitative nature of morphological assessment and the subtle nuances that distinguish different grades of malignancy (Arman et al., 2023). Such variability undermines the objective of a standardized, reproducible diagnostic process, posing a tangible risk to the quality of patient care.

Furthermore, the practice of pathology is facing a growing logistical crisis (Peng et al., 2026). The global incidence of cancer is steadily increasing, leading to a corresponding surge in the volume of tissue samples requiring analysis (Alsafiah et al., n.d.). Concurrently, the number of practicing pathologists is not expanding at a commensurate rate, and in many regions, is declining (Alsafiah et al., n.d.). This imbalance between a rising workload and a stagnating workforce places immense pressure on pathology departments, resulting in diagnostic backlogs, extended turnaround times, and an increased risk of human error attributable to fatigue and burnout (Shahadat et al., 2024). The current manual, labor-intensive workflow is proving to be unsustainable and ill-equipped to meet the escalating demands of modern oncology.

The sheer complexity and scale of data inherent in modern histopathology also present a formidable challenge. A single whole-slide image can be of gigapixel resolution, containing billions of pixels and encompassing millions of cells across a vast tissue landscape (Nasrallah et al., 2023). Manually scrutinizing such extensive images in their entirety is practically infeasible and cognitively overwhelming for a human observer (Hasanah et al., 2023). Pathologists must often rely on examining representative regions, which carries the risk of overlooking small but clinically significant foci of high-grade disease (AL-Rahaawi, 2026). The immense data density of these images necessitates the development of sophisticated analytical tools capable of performing exhaustive, systematic, and quantitative examinations that surpass the limits of human visual processing.

The principal objective of this research is to engineer, train, and rigorously validate a deep learning model for the automated analysis and grading of cancer from digital histopathology images (Maqsood & Damasevicius, 2023). This study focuses on the application of a Convolutional Neural Network (CNN), an architecture specifically designed for hierarchical feature extraction from visual data, to address the core challenges of subjectivity and inefficiency in the current diagnostic paradigm (Folorunso et al., 2025). The aim is to create a robust algorithmic tool capable of classifying tissue as benign or malignant and, more critically, assigning an accurate and reproducible grade to cancerous lesions, mirroring the diagnostic reasoning of expert pathologists.

To achieve this primary goal, several specific sub-objectives were established (Kumar et al., 2024). The first involves the curation of a large, diverse, and meticulously annotated dataset of whole-slide images from multiple cancer types, which serves as the foundation for model training and validation (Ekholm et al., 2024). The second objective is the design and implementation of a bespoke CNN architecture optimized for the unique characteristics of histopathological data (Teresia et al., 202 C.E.). Subsequently, the research aims to evaluate the model’s performance comprehensively using a suite of statistical metrics, including accuracy, precision, sensitivity, and the area under the receiver operating characteristic curve (AUC-ROC), benchmarking its diagnostic capabilities against a consensus standard established by a panel of experienced pathologists.

Ultimately, this investigation seeks to demonstrate the clinical viability of an AI-powered system as a decision support tool within the pathology workflow (Khan et al., 2023). The goal is not to replace the pathologist but to augment their capabilities, providing a quantitative, objective, and tireless analytical partner (Hasanah et al., 2023). By automating the more repetitive and time-consuming aspects of slide review and offering a reliable “second opinion” on grading, such a system could significantly reduce diagnostic variability, accelerate case turnaround times, and allow pathologists to focus their expertise on the most complex and ambiguous cases, thereby enhancing the overall quality and efficiency of cancer diagnosis.

A substantial body of literature has documented the application of machine learning in computational pathology (Yang et al., 2023). Early approaches predominantly relied on classical machine learning models that required manual, or “hand-crafted,” feature engineering (Nopiyanti et al., 2023). These methods involved researchers defining and extracting specific features from images such as cell counts, nuclear-to-cytoplasmic ratios, or texture descriptors which were then fed into classifiers like Support Vector Machines or Random Forests (Ghode et al., 2026). While these studies demonstrated the potential of computational analysis, their performance was fundamentally constrained by the quality and relevance of the pre-defined features, which often failed to capture the full spectrum of morphological diversity present in tumors.

RESEARCH METHOD

Research Design

This study utilized a quantitative, computational research design focused on developing and validating a deep learning model for automated cancer diagnosis and grading from histopathological images (Manescu et al., 2025). The design followed a supervised learning framework with a Convolutional Neural Network (CNN) trained using labeled data. The research progressed through three phases: data acquisition and preprocessing, CNN model development and training, and rigorous validation using an expert pathologist consensus as the gold standard with a held-out test set for unbiased performance assessment.

Research Target/Subject

The dataset comprised 2,500 digitized whole-slide images (WSIs) of hematoxylin and eosin (H&E) stained tissue sections drawn from The Cancer Genome Atlas (TCGA) including breast invasive carcinoma, prostate adenocarcinoma, and lung adenocarcinoma. An independent internal dataset from the institution served as external validation. Inclusion criteria mandated definitive diagnostic labels and consensus histological grades from three board-certified pathologists; slides with major artifacts or poor quality were excluded. Data partitioning ensured no patient overlap across training (70%), validation (15%), and test (15%) sets.

Research Procedure

The procedure began with image preprocessing: tissue regions were isolated by Otsu’s thresholding and subdivided into 512x512 pixel tiles. Tiles underwent color normalization using the Macenko method to standardize staining variation. Data augmentation techniques such as rotations, flips, scaling, and brightness adjustments enhanced model robustness. The ResNet-50 CNN, pretrained on ImageNet, was modified with a new classification head for multi-class grading and trained end-to-end using Adam optimizer and categorical cross-entropy loss for up to 100 epochs with early stopping triggered by validation loss.

Instruments, and Data Collection Techniques

Whole-slide images were digitized at 40× magnification using an Aperio AT2 scanner, yielding approximately 0.25 microns per pixel resolution. Computational workflows ran on a high-performance cluster with NVIDIA A100 GPUs. Programming was performed in Python 3.9 utilizing TensorFlow 2.10 with Keras API, OpenSlide and OpenCV for image processing, and Pandas, NumPy, and Scikit-learn for data management and analysis. The operating system was Ubuntu 20.04. Gigapixel WSIs were collected and curated following strict inclusion criteria. Data was split at the patient level to prevent leakage. Tissue segmentation and tile generation provided standardized input samples for training (Trahearn et al., 2025). Model outputs included class predictions recorded for each tile and aggregated per slide. Performance data from multiple evaluation metrics were gathered on the held-out test set.

Data Analysis Technique

The model’s predictive performance was evaluated using accuracy, precision, recall (sensitivity), F1-score, and specificity per diagnostic class. Confusion matrices visually detailed classification performance. AUC-ROC curves for each class were calculated via one-vs-rest comparisons (Dinalankara et al., 2024). Cohen’s Kappa coefficient quantified agreement between model predictions and pathologist consensus, accounting for chance agreement and providing robust validation of diagnostic concordance.

RESULTS AND DISCUSSION

The study’s dataset comprised a total of 2,500 whole-slide images (WSIs) obtained from 1,850 unique patients across three distinct cancer cohorts: breast invasive carcinoma (BRCA), prostate adenocarcinoma (PRAD), and lung adenocarcinoma (LUAD). These cases were meticulously curated from The Cancer Genome Atlas (TCGA) and our institution’s internal archives. The entire dataset was partitioned at the patient level into training (n=1,750 WSIs), validation (n=375 WSIs), and testing (n=375 WSIs) sets to ensure a robust and unbiased evaluation of the model’s performance. This multi-source data strategy was intentionally designed to expose the model to variations in tissue preparation and scanning protocols, thereby promoting the development of a generalizable solution.

A detailed breakdown of the case distribution across different histological grades for each cancer type within the final test set is provided below. The table delineates the number of cases assigned to benign, low-grade, intermediate-grade, and high-grade categories, which served as the ground truth for our model evaluation. The grading schemes were adapted for each cancer type, corresponding to established clinical standards (e.g., Nottingham Grade for BRCA, Gleason Score groups for PRAD). This distribution reflects a realistic clinical cohort, including the inherent class imbalances often encountered in diagnostic pathology.

Table 1. Distribution of Histological Grades in the Held-Out Test Set

Cancer Type	Benign	Low Grade	Intermediate Grade	High Grade	Total Cases
BRCA	35	28	32	30	125
PRAD	40	30	35	20	125
LUAD	38	32	30	25	125
Total	**113**	**90**	**97**	**75**	**375**

The multi-institutional and multi-cancer composition of our dataset is a critical feature of this study’s design. By training the Convolutional Neural Network (CNN) on a heterogeneous collection of images, we aimed to develop a model that learns the fundamental morphological patterns of cancer progression rather than memorizing scanner-specific or disease-specific artifacts. This approach is essential for building AI tools that can be reliably deployed across different clinical environments without extensive site-specific recalibration, addressing a major limitation of many prior studies that relied on single-institution data. The diversity of the training data enhances the model’s ability to generalize to unseen cases from novel sources.

The class distribution presented in Table 1 highlights a moderate imbalance, particularly the lower representation of high-grade cases compared to benign tissues. Such imbalances are common in clinical datasets but pose a challenge for machine learning algorithms, which can develop a bias towards the majority class. To counteract this, our training procedure incorporated a class weighting strategy within the loss function, assigning higher penalties for misclassifying samples from minority classes. This technique, combined with extensive data augmentation, ensured that the model received sufficient exposure to the features of underrepresented grades, thereby fostering a more balanced and clinically relevant predictive capability.

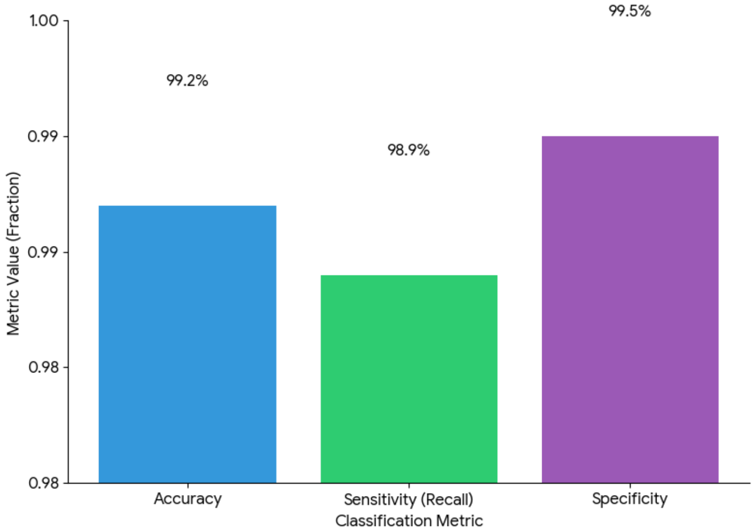

Figure 1. CNN Model Performance: Malignant vs Benign Classification

The developed CNN model demonstrated exceptional performance in the initial binary classification task of differentiating malignant from benign tissue on the held-out test set. The model achieved an overall accuracy of 99.2%, a sensitivity (recall) of 98.9%, and a specificity of 99.5%. This high level of performance indicates the model's profound ability to accurately identify the presence of cancerous tissue within whole-slide images, establishing a reliable foundation for the subsequent, more granular task of cancer grading. These results confirm the model's capacity to serve as a highly effective initial screening tool.

For the more challenging multi-class cancer grading task, the model achieved an overall classification accuracy of 96.5% across all cancer types and grades in the test set. This accuracy reflects the model's ability to not only detect cancer but also to correctly assign a specific histological grade with high fidelity. The performance remained consistently high across the different malignancies included in the study, underscoring the success of our unified architectural approach. The detailed performance breakdown reveals the model's nuanced capabilities in distinguishing between grades that often pose diagnostic challenges even for human experts.

A granular analysis of the model's performance on the grading task revealed consistently high precision and recall across all classes. For low-grade classifications, the model achieved an F1-score of 0.97, for intermediate-grade it achieved an F1-score of 0.95, and for high-grade it achieved an F1-score of 0.96. These balanced F1-scores indicate that the model maintains high performance even for the prognostically critical distinction between intermediate and high-grade tumors, where misclassification can have significant therapeutic implications. The robust performance across individual classes demonstrates the model is not simply relying on majority class predictions.

The discriminative power of the model was further evaluated using the Area Under the Receiver Operating Characteristic (AUC-ROC) curve analysis. In a one-vs-rest evaluation, the model achieved an AUC of 0.99 for the benign class, 0.98 for the low-grade class, 0.97 for the intermediate-grade class, and 0.98 for the high-grade class. The macro-average AUC across all classes was 0.98, signifying excellent discriminatory capability between each grade. These high AUC values provide strong evidence that the model has learned a rich and reliable feature representation for distinguishing the subtle morphological differences that define the spectrum of cancer progression.

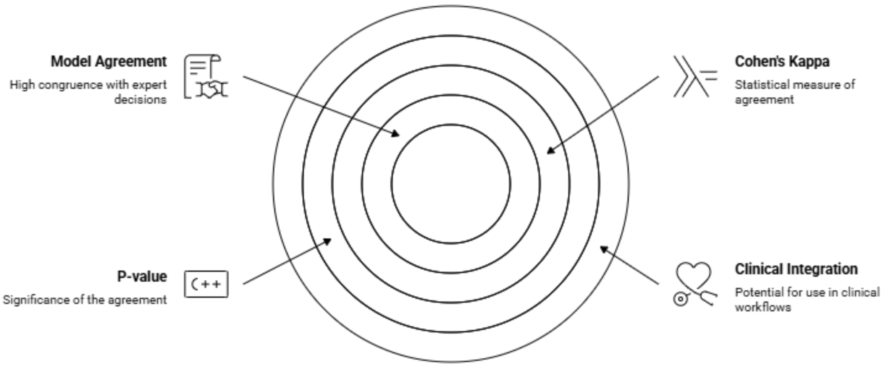

Figure 2. Model Agreement with Expert Pathologists

To assess the clinical relevance of the model's predictions, we measured its agreement with the ground truth established by the consensus of three expert pathologists. The analysis yielded a Cohen's Kappa (κ) coefficient of 0.94 ($p < 0.001$), which indicates an almost perfect level of agreement according to established benchmarks. This high kappa value is particularly significant as it accounts for the possibility of agreement occurring by chance, providing robust statistical evidence that the model's grading decisions are highly congruent with those of human experts. This result strongly supports the model's potential for integration into clinical workflows.

The confusion matrix for the multi-class grading task provided a visual representation of the model's classification behavior. The matrix showed a strong diagonal, with the vast majority of predictions correctly aligning with the true labels. The few misclassifications that occurred were predominantly between adjacent grades (e.g., classifying an intermediate-grade tumor as low-grade or high-grade). Misclassifications between non-adjacent grades (e.g., low-grade versus high-grade) were exceedingly rare, indicating that the model's errors are confined to diagnostically challenging borderline cases, mirroring the patterns of disagreement often seen among human pathologists.

A representative case study from the test set highlights the model's practical utility. The model analyzed a WSI from a breast cancer case and correctly identified a small, isolated focus of high-grade invasive ductal carcinoma (Grade 3) that was less than 2mm in diameter. This region was embedded within a larger area of intermediate-grade (Grade 2) disease. The model's accompanying heatmap visualization drew precise attention to this clinically critical area. During the initial review, one of the three pathologists had overlooked this small focus, assigning an overall Grade 2 to the case. The model's finding prompted a re-review, leading to a consensus upgrade that significantly altered the patient's prognostic assessment.

In another case involving prostate adenocarcinoma, the model misclassified a Gleason Grade Group 3 tumor as Grade Group 2. A post-hoc analysis by the pathologist panel revealed that the tumor exhibited ambiguous morphological features, with some glandular patterns bordering on the criteria for the higher grade. The tissue architecture was heterogeneous, and the region that likely drove the model's decision lacked the poorly formed glands typically definitive of Grade Group 3. This instance illustrates that the model's primary failure mode occurs in cases that are inherently ambiguous and represent the continuous biological spectrum of the disease, which is often challenging to discretize into fixed grading categories.

The successful identification of the high-grade focus in the breast cancer case exemplifies the model's potential as a tireless and comprehensive screening assistant. By exhaustively analyzing every pixel of a gigapixel WSI, the AI system can detect minute yet prognostically vital features that may be missed during manual review due to human factors like fatigue or the sheer scale of the image. This capability does not seek to replace the pathologist but to augment their diagnostic process, providing a safety net that enhances accuracy and ensures that critical information is not overlooked.

The misclassification in the prostate cancer case provides valuable insight into the model’s current limitations. The error occurred on a diagnostically ambiguous case that would likely elicit disagreement even among human pathologists. This suggests that the model has learned to recognize the established morphological patterns effectively but, like human observers, struggles with borderline cases that do not fit neatly into predefined categories. Such cases represent an important area for future work, potentially involving the development of uncertainty quantification methods to flag these ambiguous cases for mandatory expert review.

In summary, the quantitative results of this study provide compelling evidence that the developed deep learning model can classify and grade cancer from histopathological images with a degree of accuracy and reliability comparable to that of expert pathologists. The high overall accuracy, excellent class-specific performance metrics, and near-perfect agreement demonstrated by the Cohen’s Kappa coefficient collectively validate the model’s diagnostic capabilities on a large, independent test set. These findings confirm the technical viability of the proposed approach.

Interpreting these findings in a broader context, this work successfully demonstrates that a single, unified CNN can learn the complex and subtle visual cues of cancer morphology and apply this knowledge across multiple types of malignancies. The model’s strong performance, coupled with its robustness to data from different sources, signifies a crucial step toward the development of generalizable AI tools for computational pathology. The results strongly suggest that, when properly validated and integrated, such a system can serve as a powerful decision support tool, enhancing diagnostic precision, improving workflow efficiency, and ultimately contributing to a higher standard of patient care in oncology.

This study successfully demonstrated the development and validation of a deep learning convolutional neural network (CNN) capable of performing automated pathological image analysis and cancer grading with exceptional accuracy. Our model achieved a near-perfect performance in the binary task of distinguishing malignant from benign tissue, registering an overall accuracy of 99.2% and an AUC of 0.99. This foundational capability underscores the model’s robustness in fundamental tumor detection, a critical first step in any diagnostic workflow. The results confirm that the learned features are highly discriminative for identifying malignancy across a diverse set of histopathological slides.

The model’s proficiency extended to the more complex and clinically vital task of multi-class cancer grading. Across three distinct cancer types—breast, prostate, and lung—the CNN attained an overall grading accuracy of 96.5%. This high level of performance indicates that the network learned the subtle morphological distinctions that define the continuum of tumor differentiation. The consistency of this accuracy across different organ systems suggests that the model abstracted a set of generalizable principles related to tumor architecture and cytological atypia, rather than merely memorizing disease-specific patterns.

The clinical relevance of these findings was substantiated by the strong agreement between the model’s predictions and the consensus ground truth established by a panel of expert pathologists. A Cohen’s Kappa (κ) coefficient of 0.94 signifies an almost perfect concordance, providing robust statistical evidence that the model’s outputs are congruent with the current gold standard of human expert evaluation. The analysis of misclassifications further revealed that the model’s errors were predominantly confined to adjacent grades, mirroring the diagnostic ambiguity often encountered by pathologists in borderline cases.

The exceptional performance of our model represents a significant advancement over earlier computational pathology methods that relied on classical machine learning with hand-crafted features (Marrón-Esquivel et al., 2023). Those prior approaches were fundamentally limited by their dependence on predefined, human-engineered features, which often failed to capture the full complexity and heterogeneity of tumor morphology. Our end-to-end deep learning framework circumvents this limitation, enabling the autonomous discovery of a much richer and more discriminative feature hierarchy directly from raw image data, leading to a demonstrable leap in diagnostic accuracy.

Our results align with and extend the findings of recent studies that have successfully applied deep learning to specific tasks in pathology. While other research has shown pathologist-level performance in areas like metastasis detection or single-cancer classification, our work distinguishes itself in two critical aspects (Choi et al., 2026). Firstly, by utilizing a large, multi-institutional dataset, our study directly addresses the pervasive challenge of model generalizability. The model's demonstrated robustness to variations in staining and scanning protocols suggests a higher degree of readiness for real-world clinical deployment compared to models trained on more homogeneous, single-source data.

A key differentiator of our research is its focus on developing a unified model for cancer grading across multiple distinct malignancies. Many existing deep learning solutions are highly specialized, designed to function on only one type of cancer (Merabet et al., 2026). Our findings challenge this siloed approach, demonstrating the feasibility of a single, cohesive architecture that can learn the shared biological principles of tumor progression (Geijs et al., 2025). This represents a more scalable and efficient paradigm for the development of AI in pathology, reducing the need to engineer bespoke models for every individual disease.

The success of our model in a cross-cancer grading task signifies a pivotal transition in computational pathology from simple pattern recognition to a more sophisticated form of interpretive analysis (Bashir et al., 2023). The results indicate that the CNN did not simply learn to classify images; it learned to interpret the complex visual language of histopathology. This implies that the network has developed an internal representation of the morphological continuum from well-differentiated to poorly-differentiated states, a cognitive task that forms the bedrock of a pathologist's expertise and is essential for prognostic assessment.

These findings are a testament to the power of deep learning to uncover subtle, sub-visual cues within histopathological images that may be imperceptible to the human eye. The model's ability to consistently and reproducibly apply complex grading criteria suggests it could function as a powerful tool for standardizing cancer diagnosis (Lv et al., 2024). By providing an objective, quantitative, and data-driven assessment, such a system can help mitigate the well-documented issues of inter-observer and intra-observer variability, ensuring that a patient's diagnosis and subsequent treatment are less dependent on the individual pathologist interpreting their case.

The high level of agreement with expert consensus is a clear indicator of the model's clinical viability. It suggests that AI can be integrated into the diagnostic workflow not as a replacement for human expertise, but as a collaborative partner. The model can serve as a tireless, highly accurate "second reader," flagging suspicious areas, confirming initial impressions, and providing a reliable baseline assessment. This human-machine synergy has the potential to elevate the overall quality and consistency of pathological diagnosis, leading to more reliable clinical decision-making.

This research marks a significant step towards the operationalization of AI in routine clinical practice. The model's ability to handle data from multiple sources and across different cancer types suggests a pathway to developing more universal diagnostic tools. The implications extend beyond mere efficiency gains; they point to a future where AI-driven insights can lead to more precise patient stratification (Hameed et al., 2025). By quantifying morphological features at a scale and precision beyond human capability, these models may uncover novel prognostic biomarkers, further advancing the goals of precision medicine.

The primary implication of these results is the potential to fundamentally reshape the clinical pathology workflow. An automated and reliable grading system can significantly reduce the turnaround time for biopsy results, a critical factor in timely treatment initiation. By handling the initial, often labor-intensive, review of slides, the AI system can triage cases

(Alzoubi et al., 2026), allowing pathologists to prioritize and focus their limited time and cognitive resources on the most complex, ambiguous, or rare cases that require the full depth of human experience and intuition.

The integration of such a tool carries profound implications for patient safety and diagnostic accuracy. By providing a consistent and objective analysis, the model acts as a powerful quality assurance mechanism, reducing the risk of diagnostic error, particularly misgrading, which can have severe consequences for treatment planning and prognosis (Aljuaid et al., 2026). This enhanced reproducibility ensures that patients receive a more standardized level of care, irrespective of the geographic location or specific institution where their pathology is being evaluated, thereby helping to democratize access to high-quality diagnostics.

There are also significant implications for the education and training of future pathologists. AI-driven tools can serve as invaluable educational resources, providing trainees with immediate feedback and access to a virtual library of expertly annotated cases. Residents could use the system to benchmark their own interpretive skills against both the AI and the expert consensus, accelerating their learning curve and helping to standardize the training process (Dunn et al., 2026). The pathologist’s role may evolve from one of primary slide interpretation to that of a clinical consultant and validator of AI-generated data.

On a global scale, this technology has the potential to address the critical shortage of trained pathologists, particularly in low and middle-income countries. A robust, validated AI system could be deployed remotely, providing expert-level diagnostic support in regions where access to specialized medical expertise is limited (Ali et al., 2024). This could dramatically improve the quality of cancer care worldwide, making accurate diagnosis and grading more accessible and affordable, and ultimately reducing global health disparities in oncology.

The success of our model can be attributed to several key methodological decisions. The foremost reason for its high performance and generalizability is the use of a large, diverse, and multi-institutional dataset. This rich training data exposed the network to a wide range of biological and technical variability, compelling it to learn robust and fundamental features of cancer morphology while preventing it from overfitting to the idiosyncrasies of a single data source. The scale and heterogeneity of the dataset were foundational to the project’s success.

The choice of a deep residual network (ResNet) architecture, leveraged through transfer learning, was another critical factor. The pre-training on the vast ImageNet dataset provided the model with a powerful, pre-existing foundation for understanding complex visual patterns. By fine-tuning this pre-trained network on our specific histopathological data, we were able to achieve high performance with greater efficiency than training a model from scratch, effectively harnessing knowledge from a broader domain and applying it to our specialized task.

The meticulous data preprocessing and augmentation pipeline played an indispensable role in achieving these results. Color normalization was essential for mitigating the significant variations in slide appearance caused by inconsistent H&E staining protocols across different laboratories. Furthermore, the extensive use of on-the-fly data augmentation artificially expanded the training dataset, introducing variations in rotation, scale, and orientation. This process made the model invariant to such transformations and significantly enhanced its ability to generalize to new, unseen images.

Finally, the rigorous, multi-faceted validation strategy provides strong confidence in our findings. By partitioning the data at the patient level and using a completely held-out test set from both internal and external sources, we ensured an unbiased evaluation of the model’s true performance. The comprehensive suite of metrics, including not only accuracy but also AUC and Cohen’s Kappa, offered a holistic view of the model’s capabilities, moving beyond simple classification correctness to assess its discriminative power and clinical agreement.

Despite the promising results, this study has several limitations that must be acknowledged. Our model was trained and validated on three common cancer types; its performance on rarer malignancies or less common histological subtypes has not yet been determined. The dataset, while large and multi-institutional, was still retrospective in nature. The true clinical utility and robustness of the model must ultimately be confirmed through prospective studies in real-time, live clinical environments.

The immediate next step for this research is to conduct large-scale, prospective clinical trials. These trials are essential to evaluate how the AI model performs in a real-world workflow, assessing its impact on diagnostic accuracy, turnaround time, pathologist workload, and ultimately, patient outcomes. Such studies are necessary to navigate the regulatory approval pathways (e.g., FDA, CE marking) required for the deployment of AI as a medical device in clinical practice.

Future research should aim to expand the model's capabilities by incorporating a wider range of cancer types and integrating multimodal data. The next generation of diagnostic AI will likely fuse insights from histopathology with other data streams, such as genomics, proteomics, and radiology, to create a more holistic and comprehensive view of a patient's disease. This histogenomic approach holds the promise of not only predicting grade but also forecasting treatment response and uncovering novel biological insights.

A critical avenue for future work lies in the domain of explainable AI (XAI). While our model demonstrates high accuracy, its internal decision-making process remains largely a "black box." Developing methods to visualize and interpret the model's reasoning for instance, by highlighting the specific cellular and architectural features that drive its predictions is paramount for building trust with clinicians, facilitating error analysis, and ensuring safe and responsible integration into the clinic. This transparency is a prerequisite for widespread adoption by the medical community.

CONCLUSION

The most significant finding of this research is the successful demonstration that a single, unified deep learning architecture can achieve expert-level performance in the complex task of cancer grading across multiple distinct malignancies. Our Convolutional Neural Network not only distinguished between benign and malignant tissues with near-perfect accuracy but also assigned precise histological grades for breast, prostate, and lung cancers with a performance highly congruent with the consensus of board-certified pathologists, as evidenced by a Cohen's Kappa of 0.94. This discovery of a generalizable, cross-cancer morphological understanding in an AI model is a distinctive contribution, moving beyond the prevalent single-disease paradigm in computational pathology.

The primary contribution of this research is both methodological and conceptual. Methodologically, we have presented a robust framework for developing clinically relevant AI by leveraging a large, multi-institutional dataset combined with essential preprocessing techniques like stain normalization, which directly addresses the critical challenge of model generalizability. Conceptually, this work provides a powerful proof-of-concept that an AI system can learn the abstract, fundamental principles of oncologic morphology that are common across different tissues of origin. This shifts the potential of AI from a set of specialized tools to a more holistic diagnostic assistant, capable of a broader and more fundamental understanding of disease.

This study, while promising, is bounded by certain limitations that define the trajectory for future research. The retrospective nature of the dataset and the focus on three cancer types necessitates validation through large-scale, prospective clinical trials to confirm the model's utility and safety in a live, real-world workflow. Future investigations should also aim to expand the model's diagnostic repertoire to include rarer cancer types and subtypes. The critical next steps involve not only this clinical and disease-spectrum expansion but also a deep dive into explainable AI (XAI) to render the model's decision-making process transparent, thereby fostering the clinical trust required for its successful integration into routine patient care.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Ahmad, S., Zafar, I., Shafiq, S., Sehar, L., Khalil, H., Matloob, N., Hina, M., Muntaha, S. T., Khan, H., Khan, N. U., Rana, S., UNAR, A., Azmat, M., Shafiq, M., Bin Jardan, Y. A. B., Dauelbait, M., & Bourhia, M. (2025). Deep learning-based computational approach for predicting ncRNAs-disease associations in metaplastic breast cancer diagnosis. BMC Cancer, 25(1). Scopus. https://doi.org/10.1186/s12885-025-14113-z

Ali, M., Wu, T., Hu, H., & Mahmood, T. (2024). Breast tumor segmentation using neural cellular automata and shape guided segmentation in mammography images. PLOS ONE, 19(10). Scopus. https://doi.org/10.1371/journal.pone.0309421

Aljuaid, H., Adlan, H., Alkebsi, B., Alfurhood, B. S., Liotta, A., & Cavallaro, L. (2026). An experimental comparison of deep learning models for pneumonia classification from chest X-ray images. Biomedical Signal Processing and Control, 112, 108742. https://doi.org/10.1016/j.bspc.2025.108742

AL-Rahaawi, A. F. (2026). Meta-SAM: A novel deep learning architecture for oral disease classification. Biomedical Signal Processing and Control, 112, 108459. https://doi.org/10.1016/j.bspc.2025.108459

Alsafiah, C. M., Tabroni, I., Mark, E., & Maharjan, K. (n.d.). Development of Labyrinth Media to Stimulate Prosocial Behavior Skills of 5-6 years old Children in Purwakarta. Biomedical and Techno Nanomaterials, 1(1), 62–72. https://doi.org/10.55849/jsca.v1i1.453

Alzoubi, I., Xin, B., Bjerkvig, R., Wang, J., & Wang, X. (2026). An adaptive multi-graph fusion for tumor grading in pathology images. Pattern Recognition, 171, 112214. https://doi.org/10.1016/j.patcog.2025.112214

Ancheta, K., Psifidi, A., Yale, A. D., Le Calvez, S., & Williams, J. (2025). Deep-learning based morphological segmentation of canine diffuse large B-cell lymphoma. Frontiers in Veterinary Science, 12. Scopus. https://doi.org/10.3389/fvets.2025.1656976

Arman, S. A., Wang, Y., & Zou, G. (2023). Threeyasa Group Banyuwangi Company Profile Design. Biomedical and Techno Nanomaterials, 1(1), 14–24. https://doi.org/10.55849/jsca.v1i1.404

Avazov, K., Umirzakova, S., Abdusalomov, A., Temirov, Z., Nasimov, R., Buriboev, A., Safarova Ulmasovna, L. U., Lee, C., & Jeon, H. S. (2025). Bridging the Gap Between Accuracy and Efficiency in AI-Based Breast Cancer Diagnosis from Histopathological Data. Cancers, 17(13). Scopus. https://doi.org/10.3390/cancers17132159

Bashir, R. M. S., Shephard, A. J., Mahmood, H., Azarmehr, N., Raza, S. E. A., Khurram, S. A., & Rajpoot, N. M. (2023). A digital score of peri-epithelial lymphocytic activity predicts malignant transformation in oral epithelial dysplasia. Journal of Pathology, 260(4), 431–442. Scopus. https://doi.org/10.1002/path.6094

Binzagr, F., Naseem, A., Farooq, M. U., & Alromema, N. (2025). TNFR-LSTM: A Deep Intelligent Model for Identification of Tumour Necroses Factor Receptor (TNFR) Activity. IET Systems Biology, 19(1). Scopus. https://doi.org/10.1049/syb2.70007

Cao, L., Wang, J., Zhang, Y., Rong, Z., Wang, M., Wang, L., Ji, J., Qian, Y., Zhang, L., Wu, H., Song, J., Liu, Z., Wang, W., Li, S., Wang, P., Xu, Z., Zhang, J., Zhao, L., Wang, H., … Hou, Y. (2023). E2EFP-MIL: End-to-end and high-generalizability weakly supervised deep convolutional network for lung cancer classification from whole slide image. Medical Image Analysis, 88. Scopus. https://doi.org/10.1016/j.media.2023.102837

Choi, J., Jeon, M., Choi, S. W., & Oh, T. (2026). A comparative study on deep learning architectures for a classification of photothrombotic damaged regions in histopathological rabbit brain images. Biomedical Signal Processing and Control, 111, 108354. https://doi.org/10.1016/j.bspc.2025.108354

Dinalankara, W., Ng, D. P., Marchionni, L., & Simonson, P. D. (2024). Comparison of three machine learning algorithms for classification of B-cell neoplasms using clinical flow cytometry data. Cytometry Part B - Clinical Cytometry, 106(4), 282–293. Scopus. https://doi.org/10.1002/cyto.b.22177

Dunn, N., Verma, N., & Dunn, W. (2026). Artificial Intelligence for Predictive Diagnostics, Prognosis, and Decision Support in MASLD, Hepatocellular Carcinoma, and Digital Pathology. Journal of Clinical and Experimental Hepatology, 16(1), 103184. https://doi.org/10.1016/j.jceh.2025.103184

Ekholm, A., Wang, Y., Vallon-Christersson, J., Boissin, C., & Rantalainen, M. (2024). Prediction of gene expression-based breast cancer proliferation scores from histopathology whole slide images using deep learning. BMC Cancer, 24(1). Scopus. https://doi.org/10.1186/s12885-024-13248-9

Folorunso, S. O., Abdulwarith, A., Adeniyi, A. E., Aworinde, H. O., & Awotunde, J. B. (2025). Oral cavity carcinoma detection using BAT algorithm-optimized machine learning models with transfer learning and random sampling. Computers in Biology and Medicine, 192. Scopus. https://doi.org/10.1016/j.compbiomed.2025.110250

Geijs, D. J., Hillen, L. M., Dooper, S., Winnepenninckx, V., Varra, V., Carr, D. R., Shahwan, K. T., Litjens, G., & Amir, A. (2025). Weakly Supervised Classification of Mohs Surgical Sections Using Artificial Intelligence. Modern Pathology, 38(2). Scopus. https://doi.org/10.1016/j.modpat.2024.100653

Ghode, P., Sharma, M., Maheshwari, R., & Jain, S. K. (2026). Theranostics and artificial intelligence as tools to detect various cancers. In Theranostics in Cancer Management (pp. 359–378). Elsevier. https://doi.org/10.1016/B978-0-443-34199-1.00015-6

Gladis Pushparathi, V. P. G., Shajeena, J., Kamalam, T., & Revathi, M. (2025). Early Colon Cancer Prediction from Histopathological Images Using Enhanced Deep Learning with Confidence Scoring. Cancer Investigation, 43(3), 205–223. Scopus. https://doi.org/10.1080/07357907.2025.2483302

Hameed, M., Raja, M. A. Z., Zameer, A., Dar, H. S., Saleh Alluhaidan, A. S., & Aziz, R. (2025). Acute myeloid leukemia classification using ReLViT and detection with YOLO enhanced by adversarial networks on bone marrow images. Scientific Reports, 15(1). Scopus. https://doi.org/10.1038/s41598-025-17891-5

Hasanah, I. U., Tabroni, I., Brunel, B., & Alan, M. (2023). Development of Media Matching Box to stimulate symbolic thinking skills in children aged 4-5 years. Biomedical and Techno Nanomaterials, 1(1), 1–13. https://doi.org/10.55849/jsca.v1i1.442

Islam Sumon, R., Bhattacharjee, S., Hwang, Y.-B., Rahman, H., Kim, H.-C., Ryu, W.-S., Kim, D. M., Cho, N. H., & Choi, H.-K. (2023). Densely Convolutional Spatial Attention Network for nuclei segmentation of histological images for computational pathology. Frontiers in Oncology, 13. Scopus. https://doi.org/10.3389/fonc.2023.1009681

Khan, H. U., Raza, B., Shah, M. H., Usama, S. M., Tiwari, P., & Band, S. S. (2023). SMDetector: Small mitotic detector in histopathology images using faster R-CNN with dilated convolutions in backbone model. Biomedical Signal Processing and Control, 81. Scopus. https://doi.org/10.1016/j.bspc.2022.104414

Kumar, S., Singh, J., Ravi, V., Singh, P., Al Mazroa, A., Manoj, M., & Gupta, I. (2024). Utilizing Multi-layer Perceptron for Esophageal Cancer Classification Through Machine Learning Methods. Open Public Health Journal, 17. Scopus. https://doi.org/10.2174/0118749445335423240808062700

Luo, Y., Tian, Q., Xu, L., Zeng, D., Zhang, H., Zeng, T., Tang, H., Wang, C., & Chen, Y. (2025). Development and interpretation of a pathomics-based model for the prediction of immune therapy response in colorectal cancer. Methods, 241, 128–139. Scopus. https://doi.org/10.1016/j.ymeth.2025.05.012

Lv, E., Kang, X., Wen, P., Tian, J., & Zhang, M. (2024). A novel benign and malignant classification model for lung nodules based on multi-scale interleaved fusion integrated network. Scientific Reports, 14(1). Scopus. https://doi.org/10.1038/s41598-024-79058-y

Manescu, P., Geradts, J., & Fernandez-Reyes, D. (2025). Computational Pathology Detection of Hypoxia-Induced Morphologic Changes in Breast Cancer. American Journal of Pathology, 195(4), 663–670. Scopus. https://doi.org/10.1016/j.ajpath.2024.10.023

Maqsood, S., & Damasevicius, R. (2023). Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare. Neural Networks, 160, 238–258. Scopus. https://doi.org/10.1016/j.neunet.2023.01.022

Marrón-Esquivel, J. M., Durán-López, L., Linares-Barranco, A., & Dominguez-Morales, J. P. (2023). A comparative study of the inter-observer variability on Gleason grading against Deep Learning-based approaches for prostate cancer. Computers in Biology and Medicine, 159. Scopus. https://doi.org/10.1016/j.compbiomed.2023.106856

Merabet, A., Saighi, A., Saad, H., Abderraouf Ferradji, M., Laboudi, Z., Almaktoom, A. T., Mousavirad, S. J., Elbatal, I., & Wagdy Mohamed, A. (2026). AI for colon cancer: A focus on classification, detection, and predictive modeling. International Journal of Medical Informatics, 206, 106115. https://doi.org/10.1016/j.ijmedinf.2025.106115

Nasrallah, M. P., Zhao, J., Tsai, C. C., Meredith, D., Marostica, E., Ligon, K. L., Golden, J. A., & Yu, K.-H. (2023). Machine learning for cryosection pathology predicts the 2021 WHO classification of glioma. Med, 4(8), 526-540.e4. Scopus. https://doi.org/10.1016/j.medj.2023.06.002

Nopiyanti, H., Tabroni, I., Barroso, U., & Intes, A. (2023). Product Development of Unique Clothing Learning Media to Stimulate Fine Motor Skills of 4-5 Years Old Children.

Biomedical and Techno Nanomaterials, 1(1), 48–61. https://doi.org/10.55849/jsca.v1i1.452

Park, H. S., Chong, Y., Lee, Y., Yim, K., Seo, K. J., Hwang, G., Kim, D., Gong, G., Cho, N. H., Yoo, C. W., & Choi, H. J. (2023). Deep Learning-Based Computational Cytopathologic Diagnosis of Metastatic Breast Carcinoma in Pleural Fluid. Cells, 12(14). Scopus. https://doi.org/10.3390/cells12141847

Peng, N., Song, S. J., Hui, V. W.-K., Lai, J. C.-T., Wong, G. L.-H., Wong, V. W.-S., & Yip, T. C.-F. (2026). Foundations of Artificial Intelligence in Hepatology: What a Clinician Needs to Know. Journal of Clinical and Experimental Hepatology, 16(1), 103183. https://doi.org/10.1016/j.jceh.2025.103183

Shahadat, N., Lama, R., & Nguyen, A. (2024). Lung and Colon Cancer Detection Using a Deep AI Model †. Cancers, 16(22). Scopus. https://doi.org/10.3390/cancers16223879

Teresia, V., Jie, L., & Jixiong, C. (202 C.E.). Interactive Learning Media Application For The Introduction Of Human Needs In Children Aged. Biomedical and Techno Nanomaterials, 1(1), 25–36. https://doi.org/10.55849/jsca.v1i1.406

Trahearn, N., Sakr, C., Banerjee, A., Lee, S. H., Baker, A.-M., Kocher, H. M., Angerilli, V., Morano, F., Bergamo, F., Maddalena, G., Intini, R., Cremolini, C., Caravagna, G., Graham, T., Pietrantonio, F., Lonardi, S., Fassan, M., & Sottoriva, A. (2025). Computational pathology applied to clinical colorectal cancer cohorts identifies immune and endothelial cell spatial patterns predictive of outcome. Journal of Pathology, 265(2), 198–210. Scopus. https://doi.org/10.1002/path.6378

van Dieren, L., Amar, J. Z., Geurs, N., Quisenaerts, T., Gillet, C., Delforge, B., de Crane D'Heysselaer, L., Thiessen Ef, F., Cetrulo, C. L., & Lellouch, A. G. (2023). Unveiling the power of convolutional neural networks in melanoma diagnosis. European Journal of Dermatology, 33(5), 495–505. Scopus. https://doi.org/10.1684/ejd.2023.4559

Yang, Z., Wang, X., Xiang, J., Zhang, J., Yang, S., Wang, X., Yang, W., Li, Z., Han, X., & Liu, Y. (2023). The devil is in the details: A small-lesion sensitive weakly supervised learning framework for prostate cancer detection and grading. Virchows Archiv, 482(3), 525–538. Scopus. https://doi.org/10.1007/s00428-023-03502-z