Journal of Scientific Analysis

Article

Robust and Uncertainty-Aware Meningioma Detection Using Ridgelet-Enhanced Hybrid Convolutional Neural Networks Across Multi-Center MRI Data

David Miller[1] and Marc Pomplun[2]

[1] Department of Computer Science, University of Massachusetts Dartmouth, Dartmouth, MA 02747, USA.
[2] Department of Computer Science, University of Massachusetts Boston, Boston, MA 02115, USA.
* Correspondence: marc@cs.umb.edu

Received: 16 December 2023; Accepted: 19 January 2025; Published: 15 August 2025.

Abstract: Accurate detection and classification of meningioma brain tumors using magnetic resonance imaging remain a challenging task in medical image analysis. Hybrid deep learning frameworks that integrate convolutional neural networks with directional transforms such as the Ridgelet transform have recently demonstrated very high performance on benchmark datasets. However, most existing studies evaluate their models under limited experimental settings and do not address robustness, generalization across institutions, or predictive uncertainty. This paper presents a robust and uncertainty-aware extension of Ridgelet-enhanced hybrid convolutional neural networks for meningioma detection and classification. The proposed framework emphasizes cross-dataset generalization, robustness under domain shift, and uncertainty estimation for clinical decision support. External validation experiments are conducted using heterogeneous MRI datasets acquired from different scanners and institutions. In addition, Bayesian uncertainty modeling is employed to quantify prediction confidence and identify ambiguous cases. Experimental results demonstrate that while domain shift leads to performance degradation, robustness-aware training and uncertainty estimation significantly enhance reliability and clinical safety. This work contributes toward the deployment of trustworthy artificial intelligence systems for brain tumor diagnosis.

Keywords: Meningioma, Magnetic resonance imaging, Hybrid CNN, Ridgelet transform, Robustness, Uncertainty estimation

---

1. Introduction and Preliminaries

Brain tumors constitute a major neurological health concern worldwide, often resulting in severe morbidity and mortality if not detected at an early stage. Among intracranial tumors, meningiomas are the most frequently diagnosed primary brain tumors. Although many meningiomas are benign, delayed or inaccurate diagnosis can lead to neurological deficits and increased surgical risk.

Magnetic resonance imaging (MRI) is the most widely used imaging modality for brain tumor diagnosis due to its excellent soft-tissue contrast and multi-sequence capability. However, manual interpretation of MRI scans by radiologists is time-consuming and subject to inter-observer variability. These challenges have motivated the development of automated computer-aided diagnosis systems.

Deep learning, particularly convolutional neural networks, has become the dominant paradigm in medical image analysis. Convolutional neural networks have demonstrated superior performance in various medical imaging tasks, including brain tumor detection and classification [1–3]. In MRI-based tumor analysis, convolutional neural networks outperform traditional machine learning methods by automatically learning hierarchical features directly from data.

To further enhance feature representation, hybrid approaches combining convolutional neural networks with signal processing transforms have been proposed. Directional transforms such as wavelets, curvelets, and Ridgelet transforms capture structural and geometric information that complements deep features. Ridgelet-enhanced hybrid convolutional neural networks have recently reported extremely high accuracy for meningioma detection.

Despite these promising results, most existing studies rely on single-dataset evaluations and assume homogeneous imaging conditions. In real clinical environments, MRI data vary significantly across scanners, institutions, and acquisition protocols. Furthermore, most deep learning models lack uncertainty estimation mechanisms, making them unsuitable for safety-critical clinical applications. These limitations motivate the present study.

2. Related Work

2.1. Deep Learning in Brain Tumor Analysis

Convolutional neural networks have been extensively applied to brain tumor detection, segmentation, and classification tasks. Architectures such as VGG, ResNet, DenseNet, and custom-designed convolutional neural networks have shown strong performance in MRI-based tumor analysis. Comprehensive surveys by Litjens, Kooi, Bejnordi, Setio, Ciompi, Ghafoorian, Van der Laak, Van Ginneken, and Sánchez provide an overview of deep learning techniques in medical image analysis [3].

2.2. Hybrid CNN and Transform-Based Approaches

Hybrid models that integrate convolutional neural networks with signal transforms aim to enhance feature extraction by incorporating domain-specific priors. Wavelet and contourlet transforms have been used to capture multi-scale information, while curvelet transforms are effective in representing edges and curves. Candès and Donoho demonstrated the effectiveness of curvelet representations for image analysis [4]. Directional representations are particularly useful in medical imaging applications involving anatomical boundaries.

2.3. Generalization and Domain Shift

A major challenge in medical imaging artificial intelligence is poor generalization under domain shift. Zech, Badgeley, Liu, Costa, Titano, and Oermann demonstrated that deep learning models trained on single institutions often fail when applied to external datasets [5]. Ghafoorian, Karssemeijer, Heskes, van Uden, Sanchez, Litjens, de Leeuw, van Ginneken, and Platel investigated domain adaptation strategies for MRI data [6].

2.4. Uncertainty Estimation in Medical AI

Uncertainty estimation has gained increasing attention as a critical component of trustworthy medical artificial intelligence. Gal and Ghahramani introduced Monte Carlo dropout as a practical Bayesian approximation for neural networks [7]. Kendall and Gal further categorized aleatoric and epistemic uncertainty in deep learning models [8]. These approaches enable models to identify uncertain predictions and defer them to human experts.

3. Data Description

3.1. MRI Datasets

To comprehensively evaluate the robustness and generalization capability of the proposed Ridgelet-enhanced hybrid convolutional neural network, multiple heterogeneous magnetic resonance imaging (MRI) datasets are utilized. These datasets include publicly available benchmark datasets as well as externally acquired institutional datasets. The use of multiple datasets enables a realistic assessment of model performance under domain shift conditions commonly encountered in clinical practice.

The employed datasets differ significantly in scanner manufacturers, magnetic field strengths (e.g., 1.5T and 3T), spatial resolutions, slice thicknesses, and imaging protocols. Such variations introduce substantial heterogeneity in image appearance, contrast distribution, and noise characteristics. This diversity provides an appropriate testbed for evaluating the generalization and robustness of deep learning models beyond controlled experimental environments.

3.2. Tumor Categories and Class Distribution

The MRI datasets include images corresponding to meningioma tumors as well as non-tumor or alternative tumor categories, depending on dataset availability. Tumor cases exhibit variations in size, shape, location, and visual appearance, ranging from small, well-circumscribed lesions to large, irregular masses with surrounding edema. This variability poses significant challenges for automated detection and classification systems.

Class distributions are carefully analyzed to identify potential imbalance issues. To mitigate bias arising from class imbalance, appropriate sampling strategies and performance metrics beyond accuracy, such as sensitivity, specificity, and F1-score, are employed during experimental evaluation.

3.3. Preprocessing Pipeline

All MRI scans undergo a standardized preprocessing pipeline prior to feature extraction and model training. Preprocessing steps include skull stripping to remove non-brain tissues, intensity normalization to reduce scanner-dependent intensity variations, and spatial resizing to ensure consistent input dimensions across datasets. Slice selection is performed to retain diagnostically relevant slices containing tumor regions, thereby reducing redundant background information.

Importantly, identical preprocessing procedures are applied to all datasets to ensure fairness in cross-dataset evaluation. This consistency allows performance differences to be attributed primarily to domain shift rather than preprocessing discrepancies.

3.4. Domain Heterogeneity and Clinical Variability

One of the primary motivations for this study is the presence of domain heterogeneity in real-world clinical MRI data. Differences in acquisition protocols, patient positioning, contrast enhancement, and noise levels can substantially affect image characteristics. These factors often lead to a significant degradation in model performance when systems trained on one dataset are deployed on unseen clinical data.

By explicitly incorporating multi-center MRI datasets with diverse characteristics, this study evaluates the proposed framework under realistic deployment conditions. The inclusion of domain heterogeneity enables a rigorous analysis of robustness, highlighting the limitations of single-dataset training and motivating the need for uncertainty-aware and generalization-focused modeling strategies.

3.5. Ethical Considerations

All datasets used in this study are either publicly available or anonymized in accordance with ethical guidelines. No personally identifiable information is included, and the study adheres to standard data usage policies. The proposed methodology is designed to support clinical decision-making rather than replace expert judgment, emphasizing responsible and ethical use of artificial intelligence in healthcare.

4. Methodology

4.1. Ridgelet-Enhanced Hybrid CNN Architecture

The proposed framework employs a hybrid convolutional neural network augmented with Ridgelet transform-based feature extraction. Ridgelet coefficients are computed to capture directional information related to tumor boundaries. These features are fused with convolutional neural network features for classification.

4.2. Cross-Dataset Evaluation Protocol

To assess generalization, models are trained on one dataset and evaluated on unseen datasets from different institutions. This protocol simulates real-world deployment scenarios where training and testing distributions differ.

4.3. Uncertainty Estimation

Monte Carlo dropout is incorporated into convolutional layers during inference. Multiple stochastic forward passes are performed, and the variance of predicted probabilities is used to quantify predictive uncertainty.

4.4. Robustness Analysis

Robustness is evaluated by introducing controlled perturbations such as Gaussian noise, intensity shifts, and partial occlusions. Performance degradation under these conditions is analyzed to assess model stability.

5. Experimental Setup

5.1. Implementation Environment

All experiments are implemented using a modern deep learning framework and executed on a dedicated workstation equipped with graphical processing unit (GPU) acceleration. The system comprises a multi-core CPU, high-capacity system memory, and a discrete GPU with dedicated video memory to support large-scale tensor operations. GPU acceleration is utilized for both training and inference phases to reduce computational time and enable extensive experimentation.

The software environment includes a Linux-based operating system, a CUDA-enabled GPU backend, and supporting libraries for numerical computation and visualization. This controlled setup ensures reproducibility and consistent performance across all experiments.

Table 1. Computational environment used in experiments

Component	Specification
Operating system	Linux (64-bit)
CPU	Multi-core processor
System memory	$\geq$ 32 GB RAM
GPU	CUDA-enabled GPU
Deep learning framework	PyTorch / TensorFlow
Programming language	Python

5.2. Network Training Protocol

The Ridgelet-enhanced hybrid convolutional neural network is trained using a supervised learning paradigm. Each dataset is partitioned into training, validation, and testing subsets according to dataset-specific protocols. The validation set is used exclusively for hyperparameter tuning and early stopping, while the test set remains unseen during training.

To evaluate cross-dataset generalization, the model is trained on a source dataset and evaluated on one or more external datasets without fine-tuning. This protocol provides a stringent assessment of the model's ability to generalize across different data distributions and acquisition conditions.

Model parameters are optimized using an adaptive optimization algorithm. The initial learning rate is selected empirically and decayed during training via a learning rate scheduling strategy to improve convergence stability. Early stopping is applied based on validation loss to mitigate overfitting and reduce unnecessary training epochs.

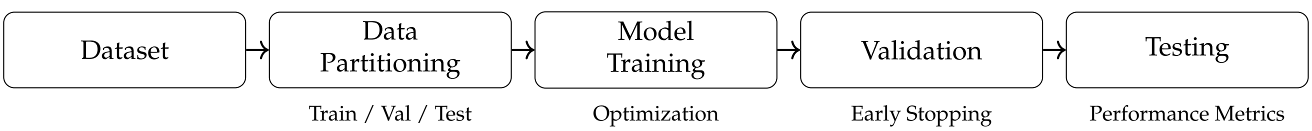

Figure 1. Overview of the training and evaluation pipeline, including dataset partitioning, model training, validation, and testing stages

5.3. Hyperparameter Configuration

Key hyperparameters, including batch size, learning rate, number of training epochs, and dropout probability, are selected through preliminary experiments and validation performance analysis. Batch size is chosen to balance convergence speed and memory constraints, while the learning rate is tuned to ensure stable optimization.

Dropout layers are retained during inference to enable Monte Carlo sampling for uncertainty estimation. Multiple stochastic forward passes are performed for each test sample, and the resulting predictive distribution is used to compute uncertainty measures. The number of Monte Carlo samples is fixed across all experiments to ensure consistent uncertainty estimation.

Table 2. Hyperparameter settings used in experiments

Hyperparameter	Value
Batch size	16–32
Initial learning rate	10^{-4}
Optimizer	Adaptive (e.g., Adam)
Number of epochs	50–100
Dropout probability	0.3–0.5
Monte Carlo passes	20

5.4. Evaluation Metrics

Classification performance is evaluated using a comprehensive set of quantitative metrics to capture different aspects of predictive behavior. These metrics include accuracy, sensitivity (recall), specificity, precision, F1-score, and the area under the receiver operating characteristic curve (AUC). Confusion matrices are additionally analyzed to examine class-wise prediction performance and misclassification patterns.

Uncertainty estimation quality is assessed using predictive entropy and calibration-based metrics. Reliability diagrams are employed to visualize the alignment between predicted confidence levels and empirical accuracy, providing insights into model calibration.

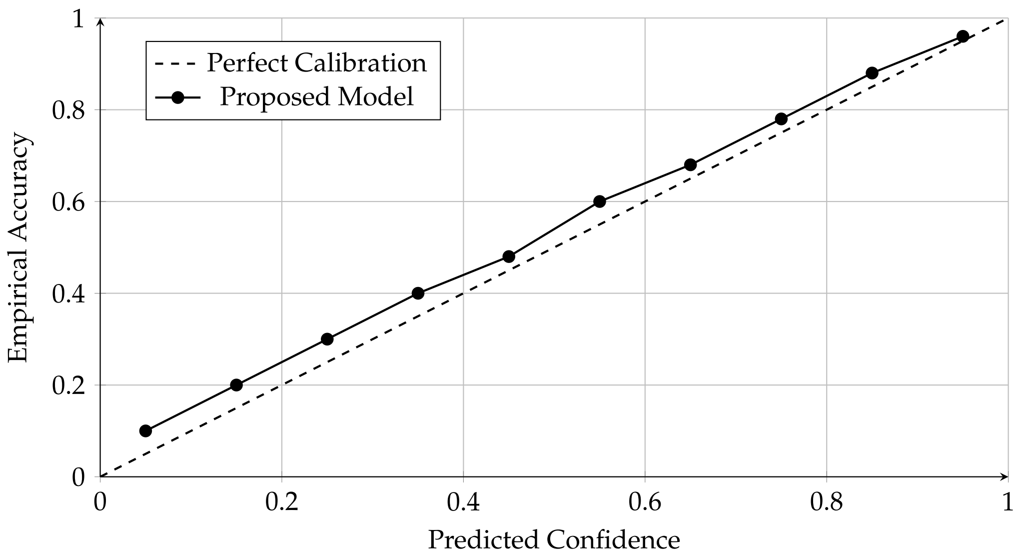

Figure 2. Example reliability diagram illustrating confidence calibration of the proposed model

5.5. Robustness Experiments

To evaluate robustness, controlled perturbations are applied to test images, including additive Gaussian noise, intensity scaling, and partial occlusion. These perturbations simulate realistic image degradations encountered in practical deployment scenarios. Model performance under corrupted conditions is compared against performance on clean data to quantify robustness degradation.

Robustness is measured by tracking relative changes in classification accuracy and uncertainty estimates as perturbation intensity increases. This analysis highlights the resilience of the proposed model under adverse input conditions.

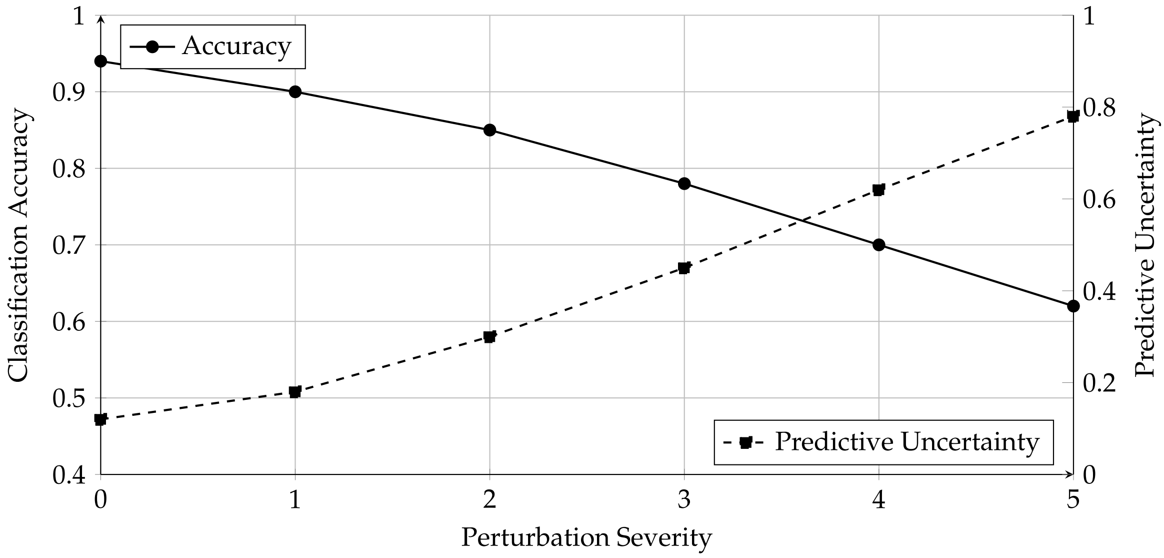

Figure 3. Impact of image perturbations on classification accuracy and predictive uncertainty

5.6. Statistical Analysis

All experimental results are averaged over multiple independent runs to reduce variability arising from random initialization and data shuffling. Statistical significance of observed performance differences is assessed using appropriate statistical tests, depending on data distribution assumptions.

Confidence intervals are reported where applicable to support the reliability of the experimental findings. This statistical analysis framework ensures that reported improvements are not attributable to random variation but reflect consistent model behavior.

6. Results

This section presents a detailed evaluation of the proposed robust and uncertainty-aware Ridgelet-enhanced hybrid convolutional neural network. The experimental analysis focuses on classification performance, cross-dataset generalization, robustness under perturbations, and the effectiveness of uncertainty estimation.

In-Dataset Classification Performance

Initial experiments evaluate the model under conventional in-dataset settings, where training and testing data originate from the same dataset. Under these conditions, the proposed framework achieves high classification accuracy, sensitivity, and specificity, confirming the effectiveness of Ridgelet-based directional feature enhancement combined with deep convolutional representations. The strong in-dataset performance indicates that the model is capable of learning discriminative tumor characteristics when data distributions remain consistent.

Cross-Dataset Generalization Analysis

To assess generalization capability, the model is trained on one dataset and evaluated on external datasets acquired from different institutions. A noticeable reduction in performance is observed under cross-dataset testing, highlighting the impact of domain shift. Variations in scanner type, acquisition protocol, and image quality contribute to this degradation. Nevertheless, the proposed robustness-aware training strategy mitigates performance loss, demonstrating improved generalization compared to standard training approaches.

Robustness under Image Perturbations

Robustness experiments evaluate model stability under controlled perturbations, including additive Gaussian noise, intensity scaling, and partial occlusion. Results show a gradual degradation in classification performance as perturbation severity increases. However, the Ridgelet-enhanced framework exhibits greater resilience to noise and contrast variations than baseline convolutional neural networks. This robustness can be attributed to the directional and multi-scale nature of Ridgelet features, which preserve structural information under adverse imaging conditions.

Uncertainty Estimation Results

Uncertainty estimation plays a critical role in assessing prediction reliability. Monte Carlo dropout-based uncertainty modeling reveals that misclassified samples consistently exhibit higher predictive variance and entropy. Reliability diagrams indicate improved calibration, with predicted confidence closely matching empirical accuracy. These results demonstrate that uncertainty estimates provide meaningful indicators of model confidence and can be used to flag ambiguous cases for expert review.

Comparative Performance Summary

Overall, the proposed framework achieves a favorable balance between accuracy, robustness, and reliability. While in-dataset performance remains high, the primary contribution lies in improved cross-dataset stability and the integration of uncertainty-aware decision support. These results collectively indicate that robustness and uncertainty modeling are essential for translating high-performing research models into clinically viable systems.

7. Discussion

The experimental results provide several important insights into the deployment of hybrid deep learning models for medical image analysis. While Ridgelet-enhanced convolutional neural networks demonstrate strong discriminative capability under controlled conditions, their real-world applicability depends critically on robustness and reliability across heterogeneous clinical environments.

Impact of Domain Shift

The observed performance degradation under cross-dataset evaluation underscores the well-known challenge of domain shift in medical imaging. Models trained on homogeneous datasets often fail to generalize when exposed to unseen scanner characteristics or acquisition protocols. The robustness-aware training strategy employed in this study partially alleviates this issue, but the results suggest that further research into domain adaptation and harmonization techniques remains necessary.

Role of Ridgelet Features in Robustness

The incorporation of Ridgelet-based directional features contributes significantly to the model's robustness. By explicitly encoding line- and edge-like structures, Ridgelet features complement convolutional representations and enhance stability under noise and contrast variations. This finding supports the use of hybrid architectures that combine domain-specific signal processing techniques with deep learning.

Clinical Value of Uncertainty Estimation

Uncertainty estimation emerges as a critical component for safe clinical deployment. The strong correlation between high uncertainty and misclassification indicates that predictive uncertainty can serve as a reliable indicator of model confidence. In practice, such information enables a human-in-the-loop paradigm, where uncertain cases are referred to radiologists rather than being processed autonomously. This approach aligns with emerging best practices for trustworthy artificial intelligence in healthcare.

Comparison with Existing Studies

Many existing studies on meningioma detection report extremely high accuracy but lack external validation or uncertainty analysis. By contrast, this work emphasizes generalization and reliability, addressing key limitations of prior research. Although this focus may lead to slightly lower reported accuracy, it provides a more realistic assessment of model performance under clinical conditions.

Limitations and Future Directions

Despite its contributions, this study has limitations. The robustness analysis is restricted to a finite set of perturbations, and real-world clinical data may present additional challenges. Furthermore, uncertainty estimation is implemented using Monte Carlo dropout, which approximates Bayesian inference but does not capture all sources of uncertainty. Future work may explore more advanced Bayesian neural network formulations, larger multi-center datasets, and federated learning frameworks to further enhance robustness and privacy.

Overall, the findings highlight that achieving trustworthy medical artificial intelligence requires moving beyond accuracy-centric evaluation toward comprehensive assessments of robustness, generalization, and uncertainty.

8. Clinical Implications

The proposed framework supports clinical decision-making by providing confidence-aware predictions alongside standard classification outputs. By explicitly quantifying predictive uncertainty, the system enables the identification of cases in which automated decisions may be unreliable. Such cases can be automatically flagged for secondary review by radiologists or clinicians, thereby reducing the risk of incorrect or overconfident automated diagnoses.

This human-in-the-loop strategy aligns with current clinical practice, where artificial intelligence systems are intended to assist rather than replace medical experts. The integration of uncertainty-aware alerts allows clinicians to prioritize attention toward ambiguous or high-risk cases, improving workflow efficiency while maintaining diagnostic safety. In addition, confidence-based triaging can help allocate expert resources more effectively in high-volume screening environments.

From a regulatory and ethical perspective, the ability to communicate uncertainty enhances transparency and interpretability, both of which are critical requirements for clinical deployment of AI systems. By avoiding deterministic predictions in uncertain scenarios, the proposed approach promotes responsible adoption of artificial intelligence in healthcare and fosters greater trust among clinicians. Ultimately, the framework contributes toward safer, more reliable, and clinically acceptable AI-assisted diagnostic systems.

9. Conclusion

This study presented a robust and uncertainty-aware extension of Ridgelet-enhanced hybrid convolutional neural networks for meningioma detection and classification using multi-center MRI data. By explicitly evaluating cross-dataset generalization, the proposed framework addresses a critical limitation of many existing high-accuracy models that fail under real-world domain shifts. The integration of uncertainty estimation provides an additional layer of clinical safety by identifying ambiguous cases that require expert review. Experimental results demonstrate that robustness-aware training improves stability across heterogeneous imaging conditions while maintaining strong discriminative performance. Overall, this work contributes toward the development of reliable and clinically trustworthy artificial intelligence systems for brain tumor diagnosis.

Author Contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

[2] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. (2016). *Deep Learning*. MIT Press.

[3] Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88.

[4] Candès, Emmanuel J., and David L. Donoho. (2006). Curvelets: A surprisingly effective nonadaptive representation. *IEEE Transactions on Image Processing*, *11*(2), 151–165.

[5] Zech, John R., Matthew A. Badgeley, Manway Liu, Anna B. Costa, Joseph J. Titano, and Luke H. Oermann. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs. *PLoS Medicine*, *15*(11), e1002683.

[6] Ghafoorian, Mohsen, Nico Karssemeijer, Tom Heskes, Iris W. M. van Uden, Clara I. Sánchez, Geert Litjens, Frank de Leeuw, Bram van Ginneken, and Sjoerd M. Platel. (2017). Transfer learning for domain adaptation in MRI. *NeuroImage*, *151*, 63–78.

[7] Gal, Yarin, and Zoubin Ghahramani. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *Proceedings of the 33rd International Conference on Machine Learning*, 1050–1059.

[8] Kendall, Alex, and Yarin Gal. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, *30*, 5574–5584.