A Helmholtzian Deep Learning Approach to Glomeruli Segmentation using Energy-based Models for Uncertainty Estimation

Omar Abul-Hassan

Abstract

Histologic analyses of glomeruli in renal tissues yield a substantial amount of information about overall health and conditions or diseases stemming from glomerular structural damage. Current medical practices of standard glomeruli assessment are completed through a manual evaluation of the quantification of normal and abnormal glomeruli in kidney biopsies, which are laborious, costly, and function with high degrees of error from missed or misclassified glomeruli. Moreover, the input of an in-house trained nephropathologist is needed for a kidney biopsy analysis — labor that is not readily available at medical centers. To address this, we propose a human-in-the-loop deep learning framework to fully automatically segment glomeruli, incorporating a novel energy-based model inspired by Helmholtz' Free Energy to compute prediction uncertainty. Because our energy approach is theoretically aligned with the probability density of inputs, we demonstrate both empirically and mathematically that our energy-based model outperforms the traditional softmax confidence score in our settings. Our framework was trained on data provided by HuBMAP and AIDPATH and achieved a validation F1-score of 94.7% on an unseen dataset. These results were shown to perform significantly better than previously related works. Overall, our approach to segment glomeruli with a confidence score maximizes the efficiency of nephropathologists with a drastically expedited process of extracting glomerular features from large histopathological images.

1. Introduction

The kidneys are organs responsible for the excretion and filtration of the body's waste, as well as the maintenance of a healthy balance of water, salts, and minerals (sodium, potassium, etc.) in blood. Kidneys are composed of the medulla and the cortex. The cortex comprises roughly one million microscopic filtration units, called nephrons. Each nephron contains one glomerulus, a mass of capillaries enclosed by the Bowman's capsule, and a tubule. The function of the glomerulus is to filter for blood components, either through mechanically filtering molecules larger than albumin (3.6 nm) or preventing negatively charged molecules, such as plasma proteins, across the filtration barrier, serving as the first stage in the formation of urine (1).

Renal filtration is largely dependent on the glomerular filtration barrier. Considering that the creation of glomeruli occurs only before term birth (2), it is crucial to monitor the structural development and progression of glomeruli to obtain invaluable information about overall kidney health. For example, indices of glomerular structural damage serve as a diagnostic tool for kidney-related ailments, such as diabetes, CKD, edema, albuminuria, or IgA nephropathy (3). Kidney disease affects around forty million adults per year, not including diabetes (3). Given the relevance of glomeruli health, protocols have been developed for nephropathologists to establish the standard for the assessment of nephrons. Specifically, medical professionals are trained to quantitatively examine renal parameters including but not limited to:

- Size
  - Healthy glomeruli have a fixed size of approximately 100 m to 350 m (4). The presence of diabetes conditions or hypertension will cause glomeruli to grow abnormally in size (4).
- Shape
  - Glomeruli present spherical shapes when healthy and can be deformed or irregularly shaped under the presence of types of nephrosclerosis (5). Abnormal glomeruli shape can also be indicative of

diabetic nephropathy (6).

- Color
  - The data used in our framework are images with PAS (Periodic Acid Schiff) stain in tissues, that gives a purple color to slides. The presence of various diseases can vary the intensity of purple color in the glomeruli (5).

- Proximal Structures
  - Nephropathologists can also look for structures that are indicative of a glomeruli. For example, Bowman's space or tubules can be easier to identify given the inherent variability in stained slides. In the case of a Bowman's space, identifying empty white space that surrounds the glomeruli may be easier to identify than looking for glomeruli. For proximal tubules, identifying the locations of the lighter-colored brush borders can help in recognizing glomeruli.

These variables are evaluated in conjunction with stereological techniques to normalize the quality of information received from renal tissues (7). These statistical techniques form the foundation for the estimation of renal parameters using two-dimensional whole slide images as a representation of the three-dimensional kidney (7).

Recent years have seen improvements in digital imaging with the advent of the whole-slide scanner. Tissue slides can be evaluated digitally in the form of high-resolution scanned images, paving the way for image processing techniques to enhance a pathologist's flow of work. Trends in digital pathology have become necessary tools for pathologists. Diagnoses from high-resolution images or quantitative evaluation of renal parameters can be performed entirely digitally with whole slide images (8).

Techniques in artificial intelligence and deep learning have rapidly gained importance in medical imaging fields. Specifically, the use of convolutional neural networks (CNN) is becoming central in the diagnosis of diseases that have most information encoded in images. For example, CNNs have shown better results in automatically diagnosing differing levels of diabetic retinopathy than ophthalmologists (9). The performance of specific CNN algorithms has also been shown to perform better than their human counterparts — CheXNet, a CNN created from chest x-rays, outperforms radiologists significantly in the detection of pneumonia (10). However, the potential implications of CNN in digital pathology have yet to be realized.

2. Previous Works

Many researchers have been working on developing techniques to automatically analyze glomeruli from whole slide images. Many of the recent methods can be classified into two groups: classification approaches and segmentation approaches.

2.1. Classification

Regarding classification, researchers have focused on developing techniques to automatically group pre-annotated glomeruli under varying degrees of damage. Marsh et. al reported an F1 score of 0.865 in the classification of glomeruli in a transplanted kidney using deep learning into sclerotic and nonsclerotic groups (19). Barros et. al reported a 92.3% precision with their k-nearest neighbors algorithm that classified whether glomeruli contained proliferative lesions (11). Yamaguchi et. al delineated the groups of glomeruli more specifically to potential diseases caused by glomerular damage, such as fibrous crescent, capillary collapse, mesangial matrix, and arrived at a final test F1 score of 0.79 using deep learning techniques (12).

2.2. Segmentation

There have been numerous attempts at proposing algorithms to effectively segment glomeruli. Kannan et. al focused-on segmenting globally sclerosed glomeruli with deep learning and created a segmentation model with a Matthew's correlation coefficient of 0.628 (14). Maree et. al detected the ellipsoidal shape of glomeruli with decision trees and reported an 87.1% F1 score on Masson trichrome stained slides (20). Davis et. al achieved a 90% and 83% Dice coefficient score after separating glomeruli into non-sclerotic and sclerotic groups, respectively (15). Some of these approaches combine classification and segmentation tasks, classifying glomeruli into diseased groups and segmenting them after this initial separation. *Table 1* summarizes current research.

3. Energy-based Approach

The mentioned researchers in *Table 1* all have contributed their unique approaches to segmenting or classifying glomeruli, usually with deep learning algorithms. However, there are some severe drawbacks in current research specific to the task of segmenting and classifying glomeruli.

- Current literature aspires to automate the process of detecting glomeruli completely, without any input of nephropathologists.
  - Challenges in segmenting and classifying glomeruli are still present with the usage of computer-aided software. The innate variabil-

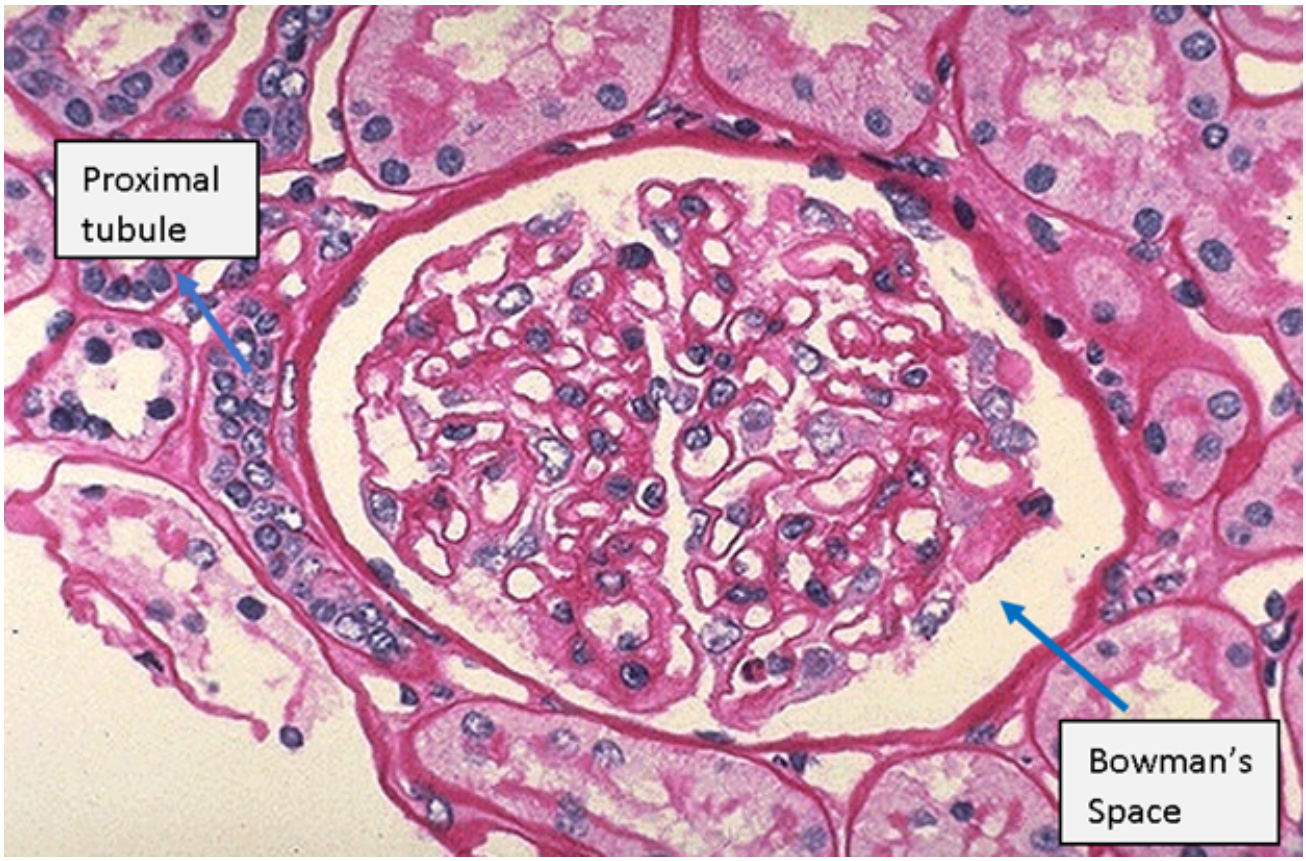

Figure 1. Illustration of Bowman's space and proximal tubules

ity between scans, heterogeneity of pathological features, and uses of multiple stains makes complete reliance on machine learning algorithms infeasible. While machine learning techniques are acquiring increasingly significant roles in renal pathology, the accuracy of image analyses is simply not enough to completely replace the role of a nephropathologist (21).

  - Because many glomeruli detection/segmentation algorithms depend on deep learning's "black box functions", the interpretability of singular predictions is nonexistent. The role of the nephropathologist is still needed to make sense of model predictions (21).

- Generalizability

  - All current research is founded upon the usage of an algorithm trained on a dataset with homogeneous structures. For example, (11) and (12) both used machine learning algorithms trained on data with one specific stain. However, renal biopsies for histopathologic evaluations use different stains: hematoxylin and eosin, periodic acid-Schiff, Masson's trichrome, Congo red, or Jones methenamine silver. The various types of stains present different benefits for specific situations, a judgment best made by the nephrologist (22). Because of this variation between stains, machine learning algorithms cannot generalize well if trained on only one specific stain.

- Uncertainty

  - Uncertainty is the largest issue that this project addresses. The errors in predictions made by machine learning models, especially black-box functions, are prone to aleatoric and epistemic uncertainty. Images used for training data can easily be corrupted by artifacts or small noise. With this said, current convolutional neural networks do not differentiate between a high-confidence prediction and a low-confidence prediction. When there is no presence of a confidence estimation, nephropathologists would treat every glomerulus prediction equally, leading to more time spent validating and invalidating predictions.

Year	Dataset	Techniques	Task	Performance
2017 (11)	811 H&E-stained ROIs from IGM-FIOCRUZ	kNN Classifier	Classification	0.88 (Recall) 0.88 (Precision) 0.85 (Accuracy)
2020 (12)	293 PAS-stained WSIs from 3 hospitals (manual annotations)	ResNet50 CNN	Classification	0.476 (average F1-score)
2020 (13)	15888 PAS- and PAM-stained images from 283 renal biopsies from Kyoto University Hospital	InceptionV3 CNN	Classification	0.986 (AUROC)
2019 (14)	275 Trichrome-stained images from Boston Medical Center	InceptionV3 CNN	Segmentation	0.927 (Accuracy) 0.868 (Cohen's Kappa)
2021 (15)	268 H&E-stained WSIs from Duke University Medical Center	U-Net CNN (with dilated convolutions)	Segmentation	0.9 (average F1-score) 0.945 (Recall) 0.855 (Precision)
2021 (16)	348 PASM, PAS, and Masson-stained WSIs from Peking University People's Hospital	Cascade Mask R-CNN	Segmentation	0.835 (F1-score)
2019 (17)	40 PAS-stained WSIs from Radboud University Medical Center	U-Net CNN Ensemble	Segmentation	0.79 (F1-score)
2020 (18)	382 PAS-stained WSIs from National Clinical Research Center of Kidney Diseases	U-Net CNN + Watershed	Segmentation	0.94 (F1-score) 0.949 (Recall) 0.931 (Precision)
Proposed	108 PAS-stained WSIs from various (refer to 4.1) sources	U-Net CNN with EfficientNetB3 encoder	Segmentation	0.947 (F1-score) 0.954 (Recall) 0.941 (Precision)

Table 1. Summary of past related works in the field

To address all the issues mentioned above with current research, I present a novel approach to segmenting and classifying glomeruli. My main unique contributions are summarized below:

- To the best of my knowledge, I introduce the first application of an energy-based out-of- distribution model inspired by Helmholtz free energy and Liu et. al (23) to predict the uncertainty of a black-box model's predictions.
- I present an ensemble of convolutional neural networks to accurately segment glomeruli, outperforming previously related research and other state-of-the-art machine learning techniques.
- To the best of my knowledge, my work is the one of the only projects that assesses generalizability to other stains. I incorporated and pre-processed training data with special augmentations that makes my model robust to other stains.
- Finally, my work is the only research that provides a currently practical framework to aid nephropathologists. Combined with my CNN, I present a human-in-the-loop pipeline for expedited glomeruli classification and segmentation.

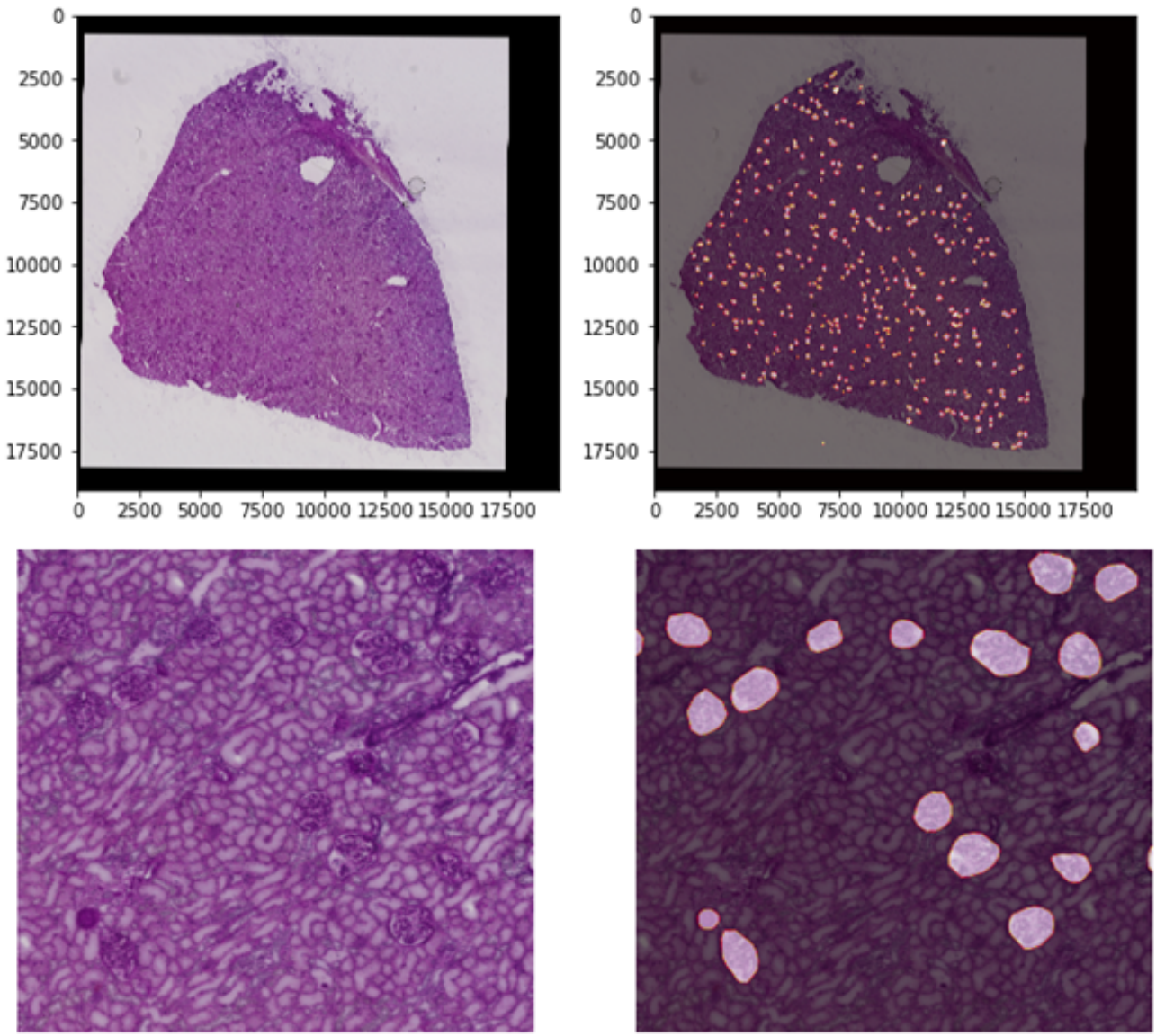

Figure 2. Processing of whole slide images and masks (top) into magnified .TIFF files and masks (bottom)

4. Methodology

4.1. Datasets

This study involved digital slides from the HuBMAP Consortium (25) and the AIDPATH kidney database (26).

HuBMAP is a global project that aims to create an open map of the human body at the cellular level, an endeavor of similar importance and level of implications as the Human Genome Project. There are around six hundred datasets that HuBMAP has made publicly available to increase outside involvement and encouragement from the global community. The dataset used consists of twenty tissue samples with glomeruli labels/masks stained by Periodic Acid Schiff (PAS) staining, eleven of which are fresh frozen and nine of which are Formalin-fixed Paraffin-embedded (FFPE) images. Metadata along with each separate kidney tissue was included, consisting of weight, height, age, sexuality, ethnicity, and race, of the kidney owner.

The AIDPATH database gathered kidney tissue samples from several institutions around Europe. It was created with the intent to encourage AI research in digital pathology (26). The dataset consists of 108 whole slide images stained by PAS staining. Glomeruli annotations exist for all images. Slides were prepared using FFPE tissue sections, according

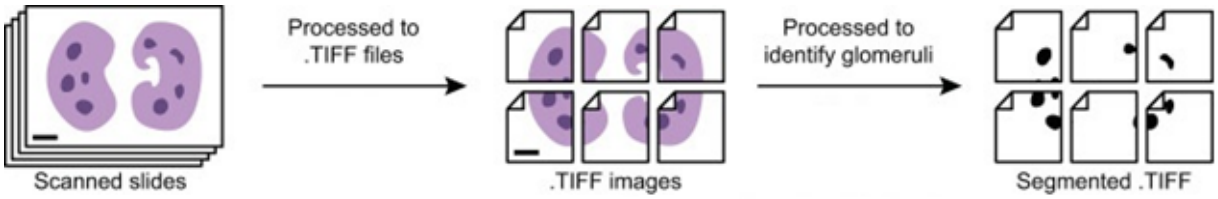

Figure 3. Processing of scanned slide images to .TIFF files (adapted from (24))

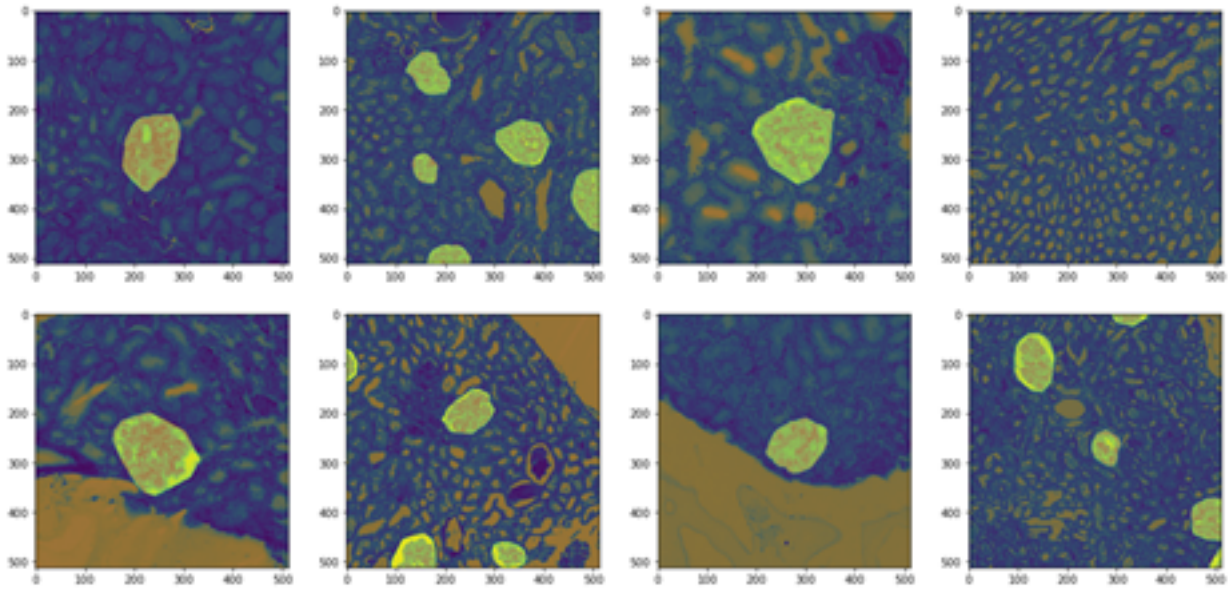

Figure 4. .TIFF images after augmentation

to the guidelines outlined in (27). Datasets with the PAS stain were specifically chosen since the aim of this study was to identify groups of glomeruli that potentially suffer from an underlying pathological disorder. PAS stains structures with carbohydrate macromolecules, typically found in connective tissues. In kidney slides, PAS is exceptional in highlighting basal membranes, where kidney glomerular disease can be reflected especially in the examination of basal membranes of glomeruli (22).

4.2. Preprocessing

There are few common approaches to extracting information from whole slide images. These digital scans are commonly captured in TIFF files, graphics files that contain raster images. Training or validating a neural network algorithm directly on these files would be hugely impractical: TIFF files that encode kidney scans are usually several gigabytes large. In order to reduce the computational power and time spent in the training of a neural network, I sampled from all files that contained the center point of a glomeruli, decomposing a large kidney scan into smaller TIFF images. This approach ensures that at least one glomerulus is seen by the model in every epoch of training, but not to the extent that oversampling from our training data occurs on a significant scale. Ultimately, this decomposition into TIFF images led to an overall faster training of my neural network as well as an easier implementation of data augmentation techniques, explained in further sections. Inspiration for this approach came mostly from (4), (14), and (28).

4.3. Data Augmentation

Data augmentation was created in a manner to create a model that is robust to differently stained scans. Thus, for the training set, data augmentation was used to increase the amount of training data and differentiate the number of examples that my model would train with. The specific transformations included random contrasts, gammas, brightnesses, and blurs, random flips, Gaussian noise, and rotations.

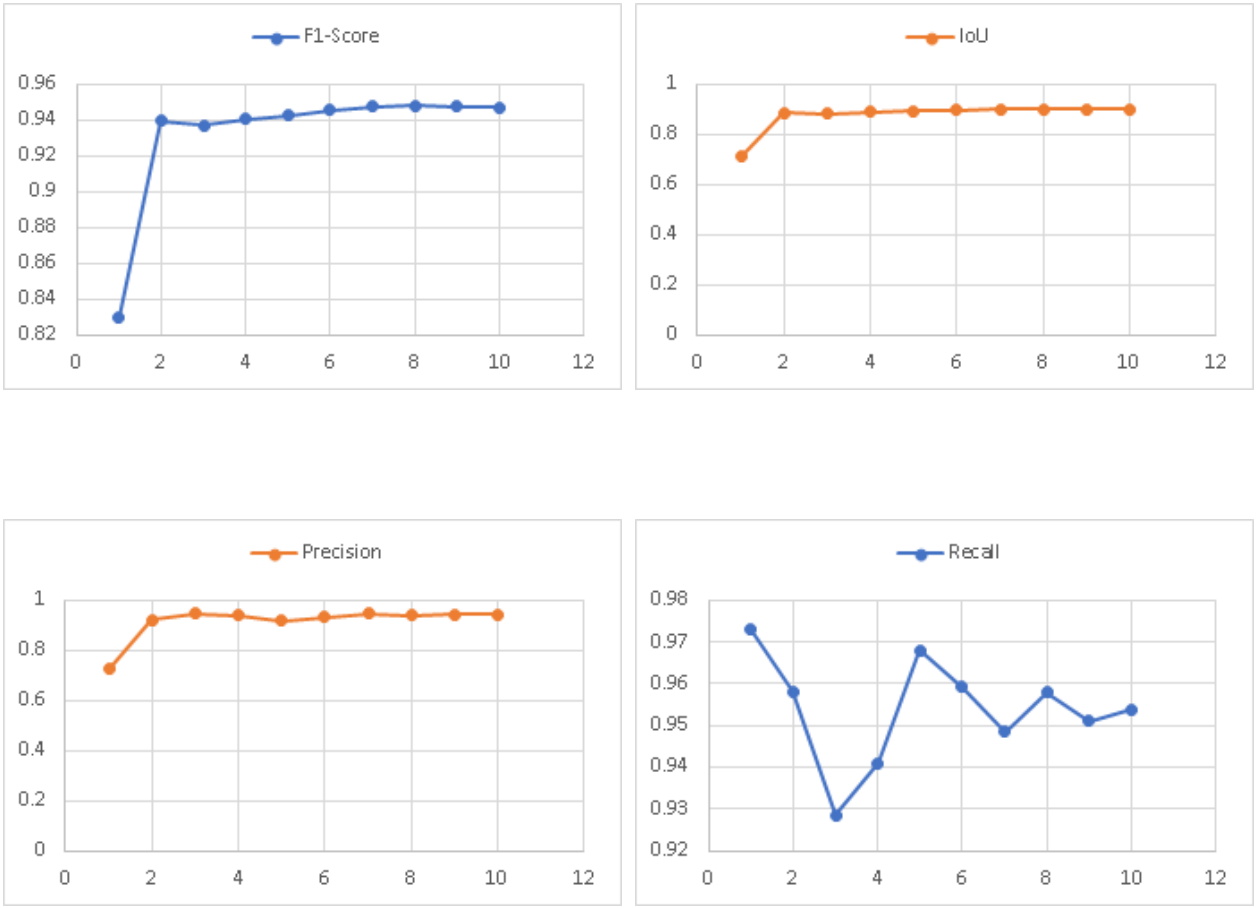

Figure 5. Overview of training performance metrics

4.4. Training

Recent advancements in convolutional neural networks have paved the way for digital image analysis and even situations with one-dimensional data. From handwritten digit recognition (29) to brain tumor segmentation (30), convolutional neural networks have a potential that is yet to be realized in imaging fields.

Semantic segmentation is a type of task in machine learning aimed to classify all pixels belonging to an input image. Usually, semantic segmentation is accomplished through convolutional neural network encoder-decoder architectures. The encoder is typically a neural network that extracts features from the training data and outputs a feature vector. The decoder is another neural network, in the opposite direction as the encoder, that "upsamples" the intermediary feature map and produces an output closest to the input or intended output. In this project, my final model uses U-Net, and a small EfficientNetB2 (ImageNet weights) as its encoder.

U-Net is a CNN architecture commonly used for biomedical applications of semantic segmentation. Developed by Ronneberger et. al, U-Net is a symmetric end-to-end fully developed convolutional neural network developed for images of any size (31). It consists of sequential upsampling and convolution layers.

The last layer of U-Net was replaced with an output classification layer. The final training framework was trained with the Dice Cross-Entropy cost function and Adam optimizer with a Cosine Annealing Learning Rate Scheduler for 100 epochs.

4.5. Validation

To evaluate the performance of my proposed model, 5-fold cross validation has been applied. Four folds were reserved for training, and the remaining one-fold in each iteration for validation. The whole dataset is validated after joining all the validation folds.

In order to be able to compare performance with related works, a combination of common segmentation metrics was used. The F1-score, or Dice coefficient (in image segmentation tasks they are mathematically equal), precision, recall, and mean Intersection over Union (IoU) were all used in assessing the performance of my model. All metrics and final testing results are mathematically represented below:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2(Precision)(Recall)}{Precision + Recall}$$

$$IoU = \frac{P \cap T}{P \cup T}$$

where *P* represents the model prediction and *T* represents ground truth.

5. Uncertainty Estimation using Helmholtzian Free Energy-based Models

To address the lack of prediction confidence in all current glomeruli detection and classification research, I investigated uncertainty estimation techniques for black-box models. Black-box models are simply functions that do not provide mathematical reasoning for predictions. A neural network can approximate any function given training data but studying the mathematical concepts behind it will not provide any practical sense. Backpropagation and feed-forward algorithms are methods to describe the inner workings of a neural network, but studying the actual variables or numbers used in these algorithms will not give insight into relationships between data, as opposed to regression techniques that can suggest correlations between variables. Neural networks are non-interpretable: we have access to their predictions, but do not know why a network predicted a certain value for a certain input.

With all of this said, research needs to address the lack of a confidence estimate in a neural network's predictions. An input image in which a glomerulus is not clearly defined (out-of- distribution data — data that does not resemble training data) should not be completely automatically predicted by a neural network but should also be confirmed by a nephropathologist. On the other hand, an input image that clearly outlines a glomerulus should be easily segmented and classified by my neural network, in which a pathologist does not have to spend as much time examining. An uncertainty estimation aims to discern between those unsure and sure predictions, which can help guide a pathologist on specific locations of a kidney tissue to spend time assessing.

I created an energy-based approach to assess my final model uncertainty, inspired by Helmholtz Free Energy equation presented in Liu et al. Using a custom energy score, Liu et. al reduced the FPR (False Positive Rate) by over 15% on classifying images of the CIFAR-10 dataset (23). Inspired by the Gibbs distribution probability density function and Helmholtz free energy of a data point, the authors created an energy-based function. The Helmholtz Free Energy E(X) of a given data point is represented as the negative log of the partition function, or the denominator of Gibbs distribution:

$$E(x) = -T \cdot \log \int_{y'} e^{-E(x,y')/T} \, dx$$

where K is the number of classes, X is the vector of predictions (logit vector), and F is my neural network. Liu et. al suggested that the temperature parameter, T, should be set equal to 1 to allow for more distinguishable energy scores between in- and out-of-distribution examples (23). This can be easily seen, as setting a temperature parameter T ¿ 1 rescales the logit vector by 1/T.

Using this formula, energy scores were calculated for all glomeruli predictions by the neural network. Higher energy scores indicate higher levels of confidence and lower levels of uncertainty, while lower scores indicate more uncertainty and less confidence in predictions.

6. Next Steps & Limitations

Below, I identify three main limitations that I plan to directly address in next steps:

- **Usage of energy score as loss function**: I adapted the Helmholtz Free Energy equation in (23) as a method of assessing model uncertainty, calculating the energy score for each glomeruli prediction. In the same paper, Liu et. al also identifies that their energy-based formula is differentiable, meaning that it could potentially be used as a loss function to optimize the neural network. Currently, my model is optimized by DiceCrossEntropy, a function that maximizes Dice score and similar metrics. However, I theorize that using the energy-based formula that Liu et. al introduced as an alternative, custom, loss function would provide a slightly faster convergence when training a neural network, due to less computationally expensive gradients used in backpropagation.
- **Stacking Architectures**: During model training, I arrived at a final model of U-Net with an EfficientNet

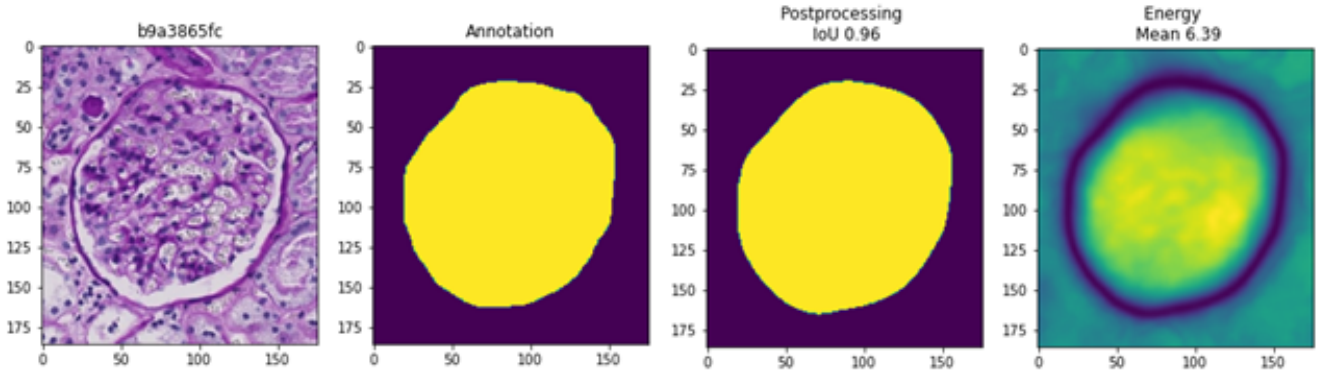

Figure 6. True positive glomerulus. High energy score (6.39) indicates high model confidence in prediction.

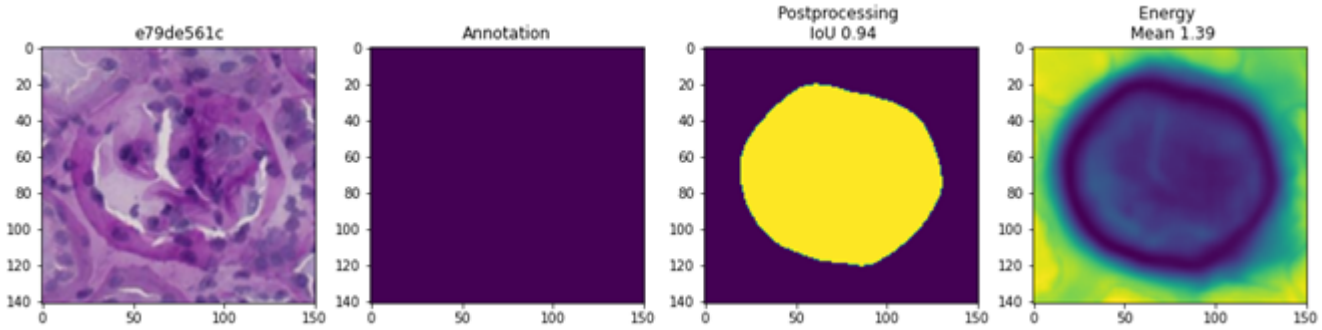

Figure 7. False positive glomerulus. Although my model incorrectly predicts a glomerulus when no annotation exists, the calculated low energy score (1.39) indicates low levels of prediction confidence.

encoder. At times, the computational power needed to execute certain combinations of neural network architectures exceeded the limitations of my home desktop. I think that I have not yet explored the power of ensembling different architectures. Especially since the importance of upsampling layers was demonstrated by Gallego et. al (4), I believe that using an ensemble of larger neural network architectures with upsampling layers would provide a significant advantage to my current model.

- **Nephropathologist access**: Access to nephropathologists would greatly improve my project's methodology and results. Uchino et. al utilized input from actual nephropathologists to confirm a prediction of glomeruli via a majority decision vote (13). Yamaguchi et. al compared his model to the review of nephropathologists, using their input as a baseline to compare to (12). Real life predictions on unseen data from nephropathologists provide invaluable comparisons to evaluate a model against. In addition, many earlier studies cite the issue of unlabeled glomeruli data as a barrier in training models. To address this, authors employ the help of pathologists to manually annotate and label data. The input of nephropathologists to produce more data would help any machine learning model tremendously. Another aspect of pathologist access that would benefit my project is with the addressal to generalizability. In this work, I aim to address current generalizability issues with heavy data augmentation, attempting to make my model robust to outliers. However, validating my model on data with different types of stains was not feasible due to a lack of access of annotated renal tissue scans with other stains. Nephropathologist access could solve this through a manual annotation of glomeruli in renal scans that are not stained by PAS, providing validation data to confirm the robustness of my model.

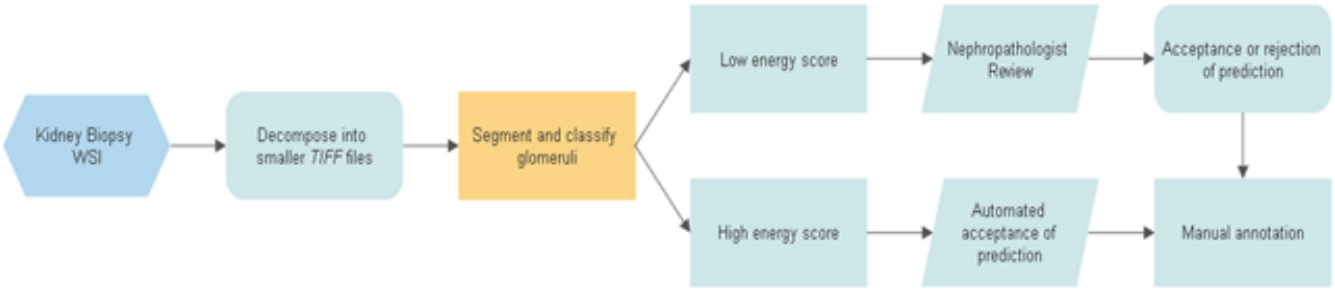

Figure 8. Rough illustration of how my energy-basd model can be implemented with a pathologist's workflow of segmenting glomeruli

7. Discussion

There are approximately one million glomeruli in each kidney. The fundamental process of renal histopathological assessment revolves around the examination of features in glomeruli. Currently, human pathologists diagnose certain pathological kidney disorders through manually localizing and assessing glomeruli.

In developed countries, this manual annotation of glomeruli delays the daily routine of a nephropathologist. In developing countries, manual annotations are virtually impossible: a delay to a trained pathologist costs life. In both, manual examinations are largely ineffective and lead to many missed glomeruli, due to sheer number of glomeruli and innate variability of nephrons.

To aid the manual process of localization and provide research towards a solution, my work provides a highly accurate neural network model to segment glomeruli.

Previous related studies choose to work with one stain, usually either PAS, Masson's Trichrome Stain, or hematoxylin and eosin. The methodology discussed in this project was created with the intent to create a model that can generalize and perform well on other renal scans with different stains through a combination of image augmentations. The proposed model was trained on augmented PAS-stained images but was also shown to perform as well on Masson-stained images.

Rather than custom architectures with few layers, my work uses transfer learning to adapt pre-made models specifically designed for biomedical image analysis. My final model consisted of a U-Net architecture with an EfficientNet encoder. The performance of my model was shown to perform better than models published in previous related works, via comparison of mean IoU, F1, recall, and precision scores.

Moreover, my work is the first to assess uncertainty of a model's segmentation predictions via an energy score inspired by the Helmholtz free energy equation. No previously related study included uncertainty estimation because they aimed to fully automate glomeruli prediction. However, because of the lack of generalizability, accuracy, and prediction interpretation, I accepted the role of nephropathologist as a confirmation figure of my model's predictions. Because of my unique approach, uncertainty estimation followed as a necessary step: the confirmation from the pathologist needed to be expedited through uncertainty estimation of the model's predictions. The model presented in this study was designed to be compatible with a pathologist's workflow. A rough outline of the implementation of my model into a pathologist's routine is depicted in Figure 8.

This work was an attempt at addressing issues with current research in glomeruli segmentation in order to expedite the workflow of a nephropathologist. I took a unique approach at understanding the problem and accepted the role of the nephropathologist as a confirmation of model prediction, an acceptance that no other current research incorporates. As a result of this, estimating the model uncertainty of predictions as an aid to nephropathologists came intuitively in my thought process, and I arrived at a final energy-based model to estimate confidence. I also addressed other issues with current research, like generalizability with data augmentation.

References

[1] R. M. Rangayyan, I. Kamenetsky, and H. Benediktsson, "Segmentation and analysis of the glomerular basement membrane in renal biopsy samples using active contours: a pilot study," *J Digit Imaging*, vol. 23, pp. 323–331, Jun 2010.

[2] M. R. Sutherland, D. Vojisavljevic, and M. J. Black, "A practical guide to the stereological assessment of glomerular number, size, and cellular composition," *The Anatomical Record*, vol. 303, no. 10, pp. 2679–2692, 2020.

[3] "Glomerular Diseases." https://www.niddk.nih.gov/health-information/kidney-disease/glomerular-diseases. Accessed September 10, 2021.

[4] J. Gallego, A. Pedraza, S. Lopez, G. Steiner, L. Gonzalez, A. Laurinavicius, and G. Bueno, "Glomerulus classification and detection based on convolutional neural networks," *Journal of Imaging*, vol. 4, p. 20, Jan 2018.

[5] M. D. Hughson, V. G. Puelles, W. E. Hoy, R. N. Douglas-Denton, S. A. Mott, and J. F. Bertram, "Hypertension, glomerular hypertrophy and nephrosclerosis: the effect of race," *Nephrol Dial Transplant*, vol. 29, pp. 1399–1409, Jul 2014.

[6] E. J. Lewis and X. Xu, "Abnormal glomerular permeability characteristics in diabetic nephropathy: implications for the therapeutic use of low-molecular weight heparin," *Diabetes Care*, vol. 31 Suppl 2, pp. S202–207, Feb 2008.

[7] J. R. Nyengaard, "Stereologic methods and their application in kidney research," *J Am Soc Nephrol*, vol. 10, pp. 1100–1123, May 1999.

[8] G. Bueno, M. M. Fernández-Carrobles, O. Deniz, and M. García-Rojo, "New Trends of Emerging Technologies in Digital Pathology," *Pathobiology*, vol. 83, no. 2-3, pp. 61–69, 2016.

[9] L. Gutierrez, J. S. Lim, L. L. Foo, W. Y. Y. Ng, M. Yip, G. Y. S. Lim, M. H. Y. Wong, A. Fong, M. Rosman, J. S. Mehta, H. Lin, D. S. J. Ting, and D. S. W. Ting, "Application of artificial intelligence in cataract management: current and future directions," *Eye Vis (Lond)*, vol. 9, p. 3, Jan 2022.

[10] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Y. Ding, A. Bagul, C. P. Langlotz, K. S. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *CoRR*, vol. abs/1711.05225, 2017.

[11] G. O. Barros, B. Navarro, A. Duarte, and W. L. C. Dos-Santos, "PathoSpotter-K: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys," *Sci Rep*, vol. 7, p. 46769, 04 2017.

[12] R. Yamaguchi, ""Glomerular Classification using Convolutional Neural Networks Based on Defined Annotation Criteria and Concordance Evaluation Among Clinicians"," *Kidney International Reports*, 2020.

[13] E. Uchino, K. Suzuki, N. Sato, R. Kojima, Y. Tamada, S. Hiragi, H. Yokoi, N. Yugami, S. Minamiguchi, H. Haga, M. Yanagita, and Y. Okuno, "Classification of glomerular pathological findings using deep learning and nephrologist–ai collective intelligence approach," *International Journal of Medical Informatics*, vol. 141, p. 104231, 2020.

[14] S. Kannan, L. A. Morgan, B. Liang, M. G. Cheung, C. Q. Lin, D. Mun, R. G. Nader, M. E. Belghasem, J. M. Henderson, J. M. Francis, V. C. Chitalia, and V. B. Kolachalama, "Segmentation of glomeruli within trichrome images using deep learning," *Kidney International Reports*, vol. 4, no. 7, pp. 955–962, 2019.

[15] R. C. Davis, X. Li, Y. Xu, Z. Wang, N. Souma, G. Sotolongo, J. Bell, M. Ellis, D. Howell, X. Shen, K. Lafata, and L. Barisoni, "Deep learning segmentation of glomeruli on kidney donor frozen sections," *medRxiv*, 2021.

[16] L. Jiang, W. Chen, B. Dong, K. Mei, C. Zhu, J. Liu, M. Cai, Y. Yan, G. Wang, L. Zuo, and H. Shi, "A deep learning-based approach for glomeruli instance segmentation from multistained renal biopsy pathologic images," *The American Journal of Pathology*, vol. 191, pp. 1431–1441, Aug. 2021.

[17] M. Hermsen, T. de Bel, M. den Boer, E. J. Steenbergen, J. Kers, S. Florquin, J. J. T. H. Roelofs, M. D. Stegall, M. P. Alexander, B. H. Smith, B. Smeets, L. B. Hilbrands, and J. A. W. M. van der Laak, "Deep Learning-Based Histopathologic Assessment of Kidney Tissue," *J Am Soc Nephrol*, vol. 30, pp. 1968–1979, 10 2019.

[18] C. Zeng, Y. Nan, F. Xu, Q. Lei, F. Li, T. Chen, S. Liang, X. Hou, B. Lv, D. Liang, W. Luo, C. Lv, X. Li, G. Xie, and Z. Liu, "Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning," *J Pathol*, vol. 252, pp. 53–64, 09 2020.

[19] J. N. Marsh, M. K. Matlock, S. Kudose, T. C. Liu, T. S. Stappenbeck, J. P. Gaut, and S. J. Swamidass, "Deep Learning Global Glomerulosclerosis in Transplant Kidney Frozen Sections," *IEEE Trans Med Imaging*, vol. 37, pp. 2718–2728, 12 2018.

[20] R. Marée, S. Dallongeville, J.-C. Olivo-Marin, and V. Meas-Yedid, "An approach for detection of glomeruli in multisite digital pathology," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1033–1036, 2016.

[21] Y. Huo, R. Deng, Q. Liu, A. B. Fogo, and H. Yang, "AI applications in renal pathology," *Kidney Int*, vol. 99, pp. 1309–1320, 06 2021.

[22] S. K. Agarwal, S. Sethi, and A. K. Dinda, "Basics of kidney biopsy: A nephrologist's perspective," *Indian J Nephrol*, vol. 23, pp. 243–252, Jul 2013.

[23] W. Liu, X. Wang, J. D. Owens, and Y. Li, "Energy-based out-of-distribution detection," *CoRR*, vol. abs/2010.03759, 2020.

[24] S. M. Sheehan and R. Korstanje, "Automatic glomerular identification and quantification of histological phenotypes using image analysis and machine learning," *Am J Physiol Renal Physiol*, vol. 315, pp. F1644–F1651, 12 2018.

[25] "Datasets." https://portal.hubmapconsortium.org. Accessed January 12, 2021.

[26] G. Bueno, L. Gonzalez-Lopez, M. Garcia-Rojo, A. Laurinavicius, and O. Deniz, "Data for glomeruli characterization in histopathological images," *Data Brief*, vol. 29, p. 105314, Apr 2020.

[27] N. R. Smith and C. Womack, "A matrix approach to guide IHC-based tissue biomarker development in oncology drug discovery," *J Pathol*, vol. 232, pp. 190–198, Jan 2014.

[28] J. M. C. Rehem, W. L. C. dos Santos, A. A. Duarte, L. R. de Oliveira, and M. F. Angelo, "Automatic glomerulus detection in renal histological images," in *Medical Imaging 2021: Digital Pathology* (J. E. Tomaszewski and A. D. Ward, eds.), vol. 11603, pp. 115 – 125, International Society for Optics and Photonics, SPIE, 2021.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[30] X. Feng, N. J. Tustison, and C. H. Meyer, "Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features," *CoRR*, vol. abs/1812.01049, 2018.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.