Muscle-Invasive Bladder Cancer Drug Response Modeling

Yujung Lin
Department of Biological Sciences
Carnegie Mellon University
Pittsburgh, PA 15213
yujunglin19@gmail.com

Nouhya Tiyal
Department of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
ntiyal@andrew.cmu.edu

Akshat Chaudhari
Department of Materials Science and Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
archaudh@andrew.cmu.edu

Abstract

This project investigates the application of expansion pathology microscopy (ExPath) images [1], a novel high-resolution imaging technique, for advancing cancer digital pathology. We hope to capture the intricate phenotypic features essential for understanding tumor behavior and predicting drug response. Using statistical methods as well as deep learning methods, we explore how different computational tools can be used to address critical challenges in oncology, such as reducing observer variability and improving the consistency and reliability of cancer diagnostics. By evaluating state-of-the-art architectures like ResNet18, DenseNet121, and SqueezeNet, as well as segmentation and feature extraction methods, this work highlights the role of image representation, model design and pre-processing steps in achieving robust classification of tumor responses.

1 Introduction

Cancer treatment relies on accurately predicting tumor behavior and response to therapy, which is a critical challenge in oncology. This project uses the novel expansion pathology microscopy images developed in the Zhao lab [1] which offer higher resolution images by physically expanding the tissue. In the field of cancer digital pathology, these images coupled with feature extraction and ML methods are instrumental in detecting phenotypic features critical for cancer diagnosis and prognosis. Previous research has demonstrated the potential of digital pathology, with studies like Campanella et al. [2] achieving pathologist-level accuracy using machine learning, and Janowczyk and Madabhushi [3] highlighting the value of feature extraction for cancer diagnostics. Building on these advancements, we compared two approaches for classifying tumor responses—feature extraction with machine learning and end-to-end deep learning—to categorize images into progressive disease (PD), partial response (RP), complete response (CR), and healthy tissue (HLTHY).

2 Dataset

Our dataset comprises 120 ExPath tumors and 10 healthy homogeneous tissue images (lymph node, squamous, and muscle) each labeled (progressive, complete response, partial response) depending on

Preprint. Under review.

their corresponding drug response. These images, originally at a resolution of approximately 10,000 x 10,000 pixels, were cropped into 512 x 512, 1024 x 1024, and 2048 x 2048 pixel patches. Each image consists of four channels: DAPI, WGA (plasma membrane), TelC (telomere), and CENPB (centromere). We mainly will be focusing on the DAPI channel that shows the cell's nuclei. Due to time and resources constraints we only use a subset of this dataset for our experiments. For approaches that use the SAM segmentation, we use a dataset of approximately 8 raw images cropped into 512 x 512 patches generating 3,200 images, uniformly distributed across our classes. For the rest of the experiments we double the dataset, generating 6,400, 1,600, and 400 images for the respective patch sizes 512, 1024, and 2048.

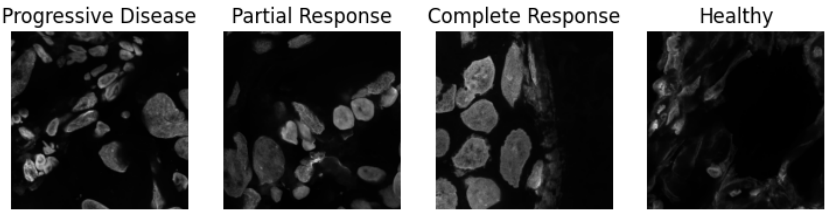

Figure 1: Dataset samples showcasing tumor and healthy tissue images labeled based on drug response (progressive, complete response, partial response). Original images (10,000 x 10,000 pixels) were cropped into patches of 512 x 512, 1024 x 1024, and 2048 x 2048 pixels, primarily focusing on the DAPI channel, which highlights cell nuclei.

3 Deep Learning Models

For this classification task, we utilized three deep learning architectures: ResNet18 [4], DenseNet121 [5], and SqueezeNet [6]. Each model was chosen for its unique strengths and suitability for handling image classification tasks.

3.1 Various models used

3.1.1 ResNet18

ResNet18, part of the ResNet family, addresses challenges in training deep networks, particularly vanishing gradients, through residual connections. These shortcut connections link earlier and later layers, enabling efficient learning of identity mappings. The architecture comprises 18 layers organized into four residual blocks, each with two convolutional layers and a skip connection. ResNet18's relatively shallow depth makes it computationally efficient while maintaining strong performance through robust feature extraction and faster convergence.

3.1.2 SqueezeNet

SqueezeNet is a lightweight CNN optimized for efficiency and low memory usage. Its "fire modules," consisting of squeeze and expand layers, significantly reduce parameters. The squeeze layer employs 1x1 convolutions to compact feature maps, minimizing dimensionality and computational costs. SqueezeNet is ideal for resource-constrained scenarios or edge device deployment while maintaining competitive accuracy.

3.1.3 DenseNet121

DenseNet121 connects each layer to every other layer, promoting feature reuse and improving efficiency. Its dense blocks facilitate the flow of features and gradients, while transition layers with 1x1 convolutions and pooling manage complexity. DenseNet121 excels at capturing fine-grained details, making it well-suited for tasks requiring nuanced class distinctions.

The major differences have been compiled in 3.1.3.

Aspect	ResNet18	SqueezeNet1_0	DenseNet121
Architecture	Residual Neural Network with skip connections.	Lightweight model with Fire modules for parameter efficiency.	Dense connections where each layer connects to every other layer.
Model Size	Moderate (11.7M parameters).	Extremely small (1.2M parameters).	Moderate (7.98M parameters).
Computational Efficiency	Balanced.	Very efficient; low resource usage.	High memory and compute requirements.
Accuracy	Good for general-purpose tasks.	Lower compared to deeper models.	High due to feature reuse.
Training Speed	Fast and stable.	Fastest among the three.	Slower due to dense connections.
Inference Speed	Moderate.	Fast due to small size.	Slow due to complex architecture.
Parameter Efficiency	Moderate.	Highly efficient.	Efficient with better utilization.
Application Suitability	General-purpose; balanced performance and efficiency.	Resource-constrained environments.	Accuracy-driven tasks in high-resource settings.

Table 1: Comparison of ResNet18, SqueezeNet1_0, and DenseNet121

3.2 Performance on grayscale and binary masks

For this classification task, models were trained and evaluated on image patches of varying sizes: 512×512, 1024×1024, and 2048×2048. Performance metrics, measured as accuracy percentages, revealed a consistent trend in which grayscale images outperformed binary masks in all models and patch sizes as shown in 2.

Observations:

- Across all patch sizes and models, grayscale images yielded higher accuracy compared to their binary counterparts.
- For example, ResNet achieved an accuracy of 90.11% on 512×512 grayscale patches, compared to 60.13% on binary patches. Similarly, for 2048×2048 patches, ResNet reached 100.00% accuracy on grayscale images, but only 78.79% on binary masks.
- SqueezeNet and DenseNet also followed a similar pattern, with grayscale accuracies consistently exceeding binary results.

Possible Reasons for the Performance Gap:

1. **Loss of Information in Binary Masks:** Binary masks reduce the image to a two-tone representation, losing critical intensity and texture details present in grayscale images. These features are particularly important for models to learn complex patterns, such as subtle variations in cell structure and staining.
2. **Feature Extraction:** Grayscale images retain more values of pixel intensity, providing richer data for feature extraction. Deep learning models like ResNet, DenseNet, and SqueezeNet leverage these nuanced differences to better distinguish between classes.
3. **Model Sensitivity to Texture and Gradient:** The architectures used in this study are highly sensitive to texture, edges, and gradients. Grayscale images, by preserving these characteristics, align better with the ability of the convolutional layers to extract spatial features.

4. **Binary Mask Noise:** Binary masks may introduce artifacts or abrupt transitions at object boundaries, which may possibly mislead the model during training. Grayscale images offer smoother transitions and a more accurate representation of object boundaries.
5. **Data Representation Limitations:** Binary masks lack the ability to represent intermediate intensities, which might convey crucial information about cell morphology or the response to staining in this dataset.

Table 2: Performance of Models on Grayscale and Binary Image Patches

Patches	**512 x 512**		**1024 x 1024**		**2048 x 2048**	
Models	**Grayscale**	**Binary**	**Grayscale**	**Binary**	**Grayscale**	**Binary**
ResNet	90.11	60.13	92.62	76.51	100.00	78.79
SqueezeNet	80.86	59.65	88.59	73.83	88.00	78.79
DenseNet	86.76	62.04	92.62	75.84	96.00	78.79

3.3 Performance on various patch sizes

In this section of classification task, we look into the performance of ResNet, DenseNet, and SqueezeNet on images across three patch sizes: 512×512, 1024×1024, and 2048×2048. The results, as shown in Table 3, reveal distinct trend for each model and patch size.

3.3.1 Observation

- For PD, accuracy consistently decreases as the patch size increases across all three models.
- For PR, accuracy increases for DenseNet and SqueezeNet but decreases for ResNet.
- CR shows a general upward trend inaccuracy across all models as the patch size increases, with DenseNet achieving 100% accuracy at 2048×2048.
- HLTHY tissue classification accuracy is consistently high and achieved 100% accuracy for ResNet and DenseNet at 2048×2048. The accuracy of SqueezeNet for HLTHY, however, decreases as the patch size increases.
- Notably, ResNet with a patch size of 512×512 demonstrated the best overall performance, achieving an accuracy of 80.69%.

3.3.2 Discussion

This section highlights the interplay between patch size and model architecture in influencing classification accuracy. Smaller patch sizes preserve localized details, which is crucial for detecting subtle phenotypic differences for models such as ResNet. The accuracy of CR increases as the patch size increases to 2048×2048 is likely due to the broader contextual information one patch contains. The drop in PD accuracy with increasing patch size suggests that important details such as cell density might be lost or indistinguishable between classes at higher resolutions.

Table 3: Performance of Models on Different Patch Sizes

	ResNet			**DenseNet**			**SqueezeNet**		
	512	**1024**	**2048**	**512**	**1024**	**2048**	**512**	**1024**	**2048**
PD	90.10	62.22	50.00	73.96	77.78	40.00	72.40	68.89	40.00
PR	78.08	79.41	71.43	78.77	73.53	85.71	71.23	79.41	85.71
CR	67.02	80.85	90.01	73.82	63.83	100.00	70.16	65.96	90.91
HLTHY	91.67	96.15	100.00	89.81	96.15	100.00	88.89	84.62	83.33
Overall	80.69	77.63	76.47	77.71	75.66	79.41	74.25	73.03	73.53

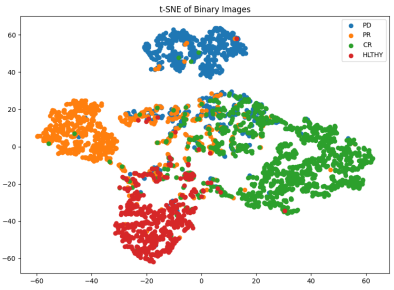

Figure 2: Binary 512x512

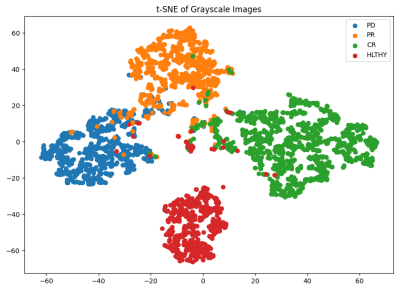

Figure 3: Grayscale 512x512

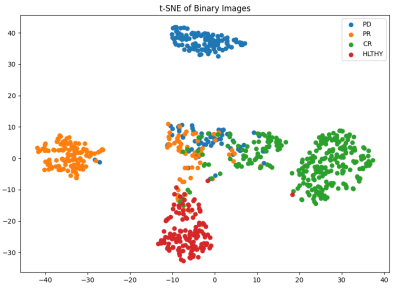

Figure 4: Binary 1024x1024

Figure 5: Grayscale 1024x1024

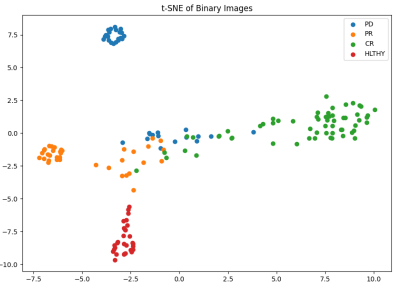

Figure 6: Binary 2048x2048

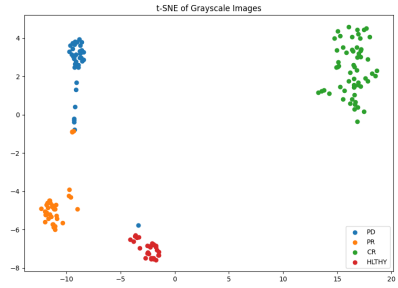

Figure 7: Grayscale 2048x2048

Figure 8: These t-SNE plots display the results of ResNet18 for various patch sizes and image types. The plots clearly show significant overlap between the four categories when using binary masks, whereas the grayscale datapoints exhibit much better separation. This suggests that the model is able to distinguish the categories more effectively when grayscale images are used.

4 Evaluating Segmentation and Feature Extraction Importance

The intuition behind investigating the role of segmentation and feature extraction lies in their ability to reveal underlying biological patterns that are not directly discernible from raw image data. By isolating regions of interest and extracting meaningful features, these methods enable models to learn class-distinguishing characteristics effectively, improving classification accuracy and interpretability. This section evaluates these aspects using two complementary approaches: traditional machine learning (ML) pipelines leveraging handcrafted features and deep learning (DL) models enhanced with segmentation outputs.

4.1 Segmentation and Feature Extraction Through Statistical Analysis and ML

4.1.1 Data Preparation

To analyze the effect of segmentation and feature extraction, 512x512 patches were generated using the SAM segmentation tool, which isolates regions of interest such as cells and tissue structures. The segmentation output allowed for the extraction of high-scoring features based on prior experimentation with various segmentation methods (details visualized in supplementary figures). Features extracted were broadly categorized into:

1. Numerical and Geometrical: Area, perimeter, aspect ratio.
2. Intensity-Based: Mean intensity, standard deviation of intensity, sum intensity.
3. Texture Features: Derived from the gray-level co-occurrence matrix (GLCM), including contrast, dissimilarity, homogeneity, energy, and correlation.

4.1.2 Methodology

A random forest classifier was trained using these features to classify the dataset into four classes: Progressive Disease (PD), Partial Response (PR), Complete Response (CR), and Healthy Control Tissue (HLTHY). This classifier provides a lightweight and interpretable benchmark for evaluating the effectiveness of the extracted features.

4.1.3 Results and Benchmarking

The random forest classifier achieved a test accuracy of 86%, with confusion matrices in figure 9 which highlight its effectiveness in distinguishing between classes. However, its lower performance on the partial response class can be attributed to this class being transitional, sharing features with both the complete response and progressive disease classes. In comparison, a ResNet model trained on the same reduced dataset without segmentation achieved 90% accuracy 10. While ResNet's accuracy was higher, the random forest model was significantly more efficient, training in seconds on a CPU, unlike the GPU-intensive ResNet that required hours.

4.1.4 Discussion

This method demonstrates the power of handcrafted features for classification tasks. Future work could focus on increasing the dataset size to improve model reliability and exploring additional feature sets to further boost performance.

4.2 Deep Learning with Fine-Tuned Segmentation Integration

4.2.1 Data Preparation

To explore the role of segmentation using the DL approach, 1024x1024 raw image patches were used to preserve spatial context. A U-Net model with a ResNet-34 encoder backbone pretrained on ImageNet was employed for segmentation.

4.2.2 Methodology

The U-Net model was adapted by replacing its decoder with an average pooling layer and a fully connected layer, producing direct class predictions. This approach leverages the segmentation capabilities of U-Net to focus on spatial and morphological features of the tumor and tissue structures relevant to drug response classification.

4.2.3 Results and Benchmarking

The adapted U-Net achieved a test accuracy of 87.66% 11, surpassing the benchmark ResNet trained on the same dataset, which achieved only 80% accuracy 12. The performance improvement highlights the importance of incorporating segmentation into the classification pipeline to extract biologically relevant features.

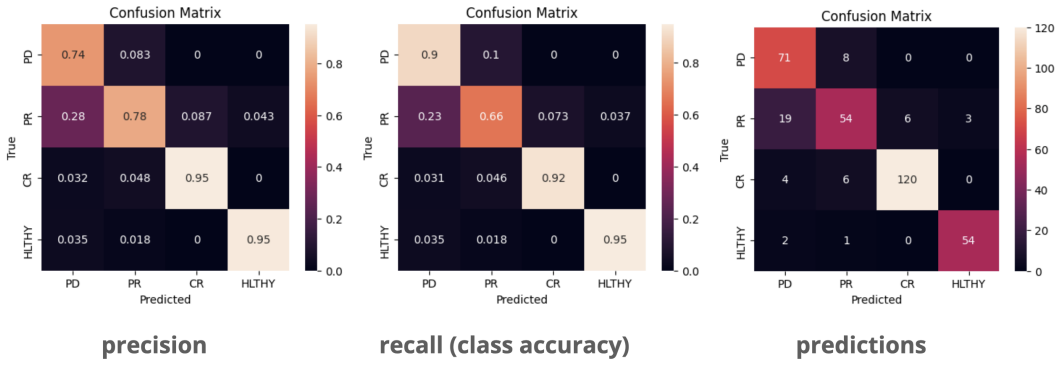

Figure 9: ML Confusion Matrices

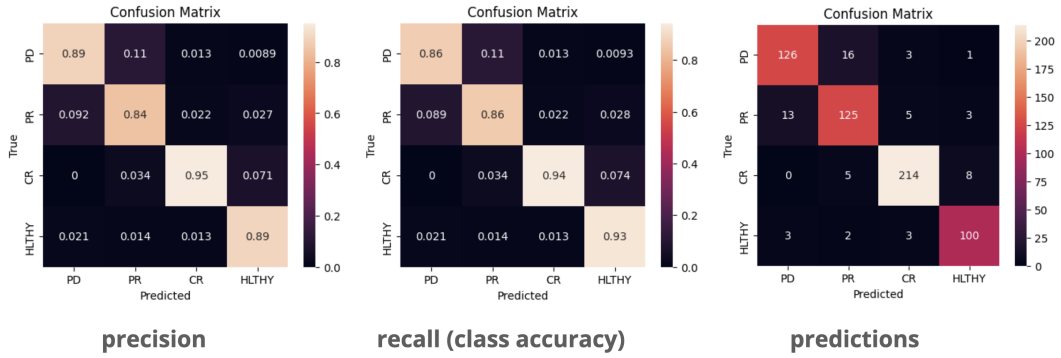

Figure 10: ResNet Confusion Matrices

4.2.4 Discussion

The U-Net-based model effectively demonstrates the value of combining segmentation with classification. However, further improvements are possible by:

1. Fine-tuning segmentation models for this specific dataset using SAM results as ground truths.
2. Employing state-of-the-art (SOTA) segmentation architectures.
3. Using domain-specific pretrained models for tasks like cell segmentation instead of generic ImageNet models.

4.3 Future Work

Future directions for advancing this work include:

- Fine-tuning segmentation models for domain-specific tasks to achieve more precise feature extraction.
- Exploring embeddings from fine-tuned models to train lightweight ML classifiers, evaluating the trade-offs between fine-tuning cost and classification performance.
- Expanding the dataset to improve generalizability and robustness across diverse tissue types and drug response scenarios.

5 Conclusion

In conclusion, this project demonstrates the superior performance of grayscale images over binary masks across all models and patch sizes tested, highlighting their ability to retain critical intensity,

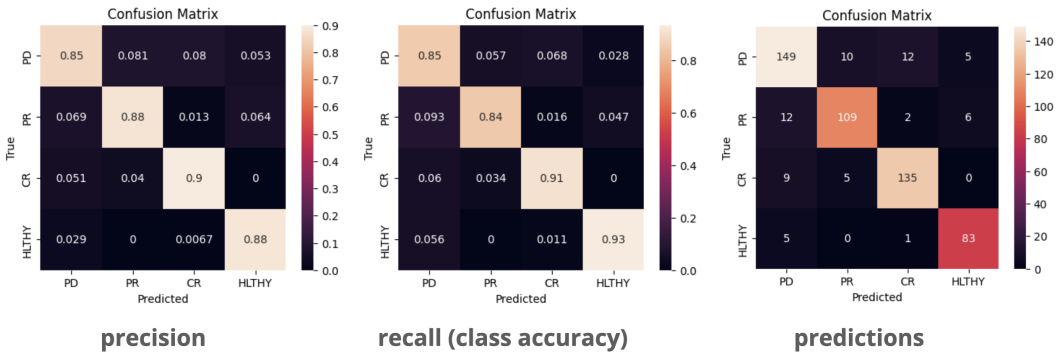

Figure 11: U-Net Confusion Matrices

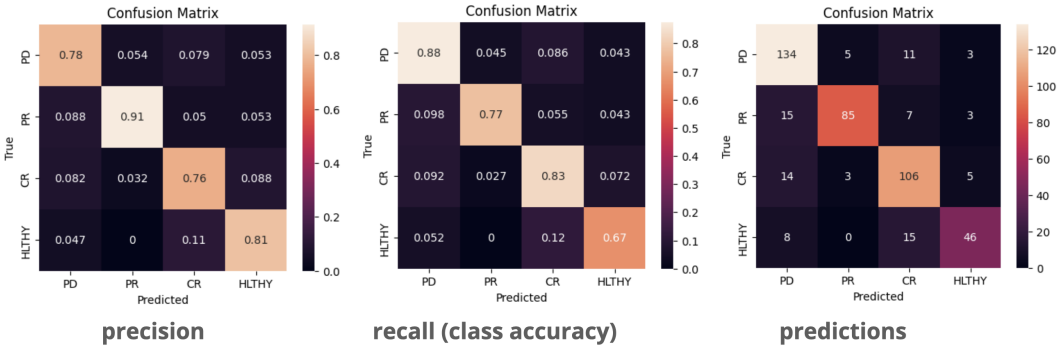

Figure 12: ResNet Confusion Matrices

texture, and gradient information essential for effective feature extraction. Binary masks, with their reduced representation, often introduced artifacts and failed to capture subtle variations, leading to a noticeable performance gap. Additionally, the findings reveal the significant impact of patch size on classification accuracy, with the need to tailor patch size and model selection to specific classification tasks and target classes. The study also underscores the importance of segmentation and feature extraction in drug response modeling, highlighting the trade-off between the computational efficiency of traditional machine learning methods and the superior accuracy of deep learning models utilizing segmentation outputs. These results collectively emphasize the value of rich data representations, tailored methodologies, and a balanced trade-off between efficiency and accuracy, paving the way for more precise and impactful applications in cancer diagnostics and treatment planning.

6 Contributions

All members contributed equally to the project. We worked collectively on the problem statement, devising strategies for getting the solution and brainstorming methods, writing code for patching, masking, model training, and the extensive analysis that followed. The following is a approximate breakdown of our respective contributions:

- **Nouha**
  - Experimented with segmentation methods and feature extraction
  - Explored the U-Net method for segmentation-based classification
  - Developed an ML classifier based on the extracted features
- **Akshat**
  - Generated binary masks for dataset after patching using Meta's SAM model.
  - Trained selected DL models and obtained performance metrics for original patches and binary masks.
  - Investigated results to find which data type is better and potential reasons.

- **Yujung**
  - Generated patches of different sizes for dataset.
  - Trained selected DL models and obtained performance metrics for different patch sizes.
  - Explore the reasons why different patch sizes give different performances

References

[1] Zhao, Yongxin, et al. (2017). Nanoscale imaging of clinical specimens using pathology-optimized expansion microscopy. *Nature biotechnology*, *35*(8), 757-764.

[2] Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., ... & Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, *25*(8), 1301–1309.

[3] Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, *7*, 29.

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. `https://doi.org/10.1109/CVPR.2016.90`.

[5] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269. `https://doi.org/10.1109/CVPR.2017.243`.

[6] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv preprint arXiv:1602.07360*. `https://doi.org/10.48550/arXiv.1602.07360`.