Fully-Automated Bone Marrow Lesion Segmentation Using a UNET-Based CNN Architecture

Matei Gardea, Vahid Anwari, Siwen Liu, Nima Yazdankhah, Rameen Usman, Andy Kin On Wong[1,2*]

[1]Joint Department of Medical Imaging, [2]Centre of Excellence in Skeletal Health Assessment, University Health Network, Toronto, ON

* = Corresponding author contact:
andy.wong@uhnresearch.ca
Tel: 905-399-0329, Fax: 416-340-3215

Introduction

- Bone Marrow Lesions (BMLs) are associated with pain in knee osteoarthritis.
- Semi-quantitative scoring lask reproducibility.
- Semi-automated segmentation is time consuming and requires significant amounts of human intervention.

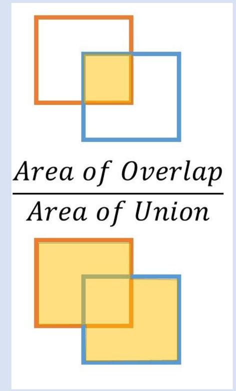
Objective

- 1) To automate the process of segmentation and quantifying bone marrow lesions in individual 2D MRI slices of the knee.
- 2) To evaluate aforementioned model using image segmentation metrics.

Methods

- A list of patients from the Osteoarthritis Initiative whose scoring in the BML component of their MOAKS, BLOKS, and WORMS evaluations was at least 1 were shortlisted.
- Of this sample, 100 randomly selected patients (each having 160 sagittal DESS slices) were used to form a subset. 2,184 BML-positive slices were split into 80% training and 20% testing.
- Mean age was 63.2 years (Std. Dev. 10.2) and mean BMI 29.67 kg/m^2 (Std. Dev. 4.78).
- KL score distributions of subset; 6% KL=1, 30% KL=2, 49% KL=3, 15% KL=4
- Annotations performed using the LabelMe tool and verified by 2 radiologist readers
- Custom-built UNET-based neural network architecture.
- Loss function: 1 - (Dice Coef.). Metric: IoU

Loss Function and Metric



$$Dice = 1 - 2 \times \frac{(A \cup B + smooth)}{(A + B + smooth)}$$

$$IoU = \frac{A \times B}{A + B - (A \times B)}$$

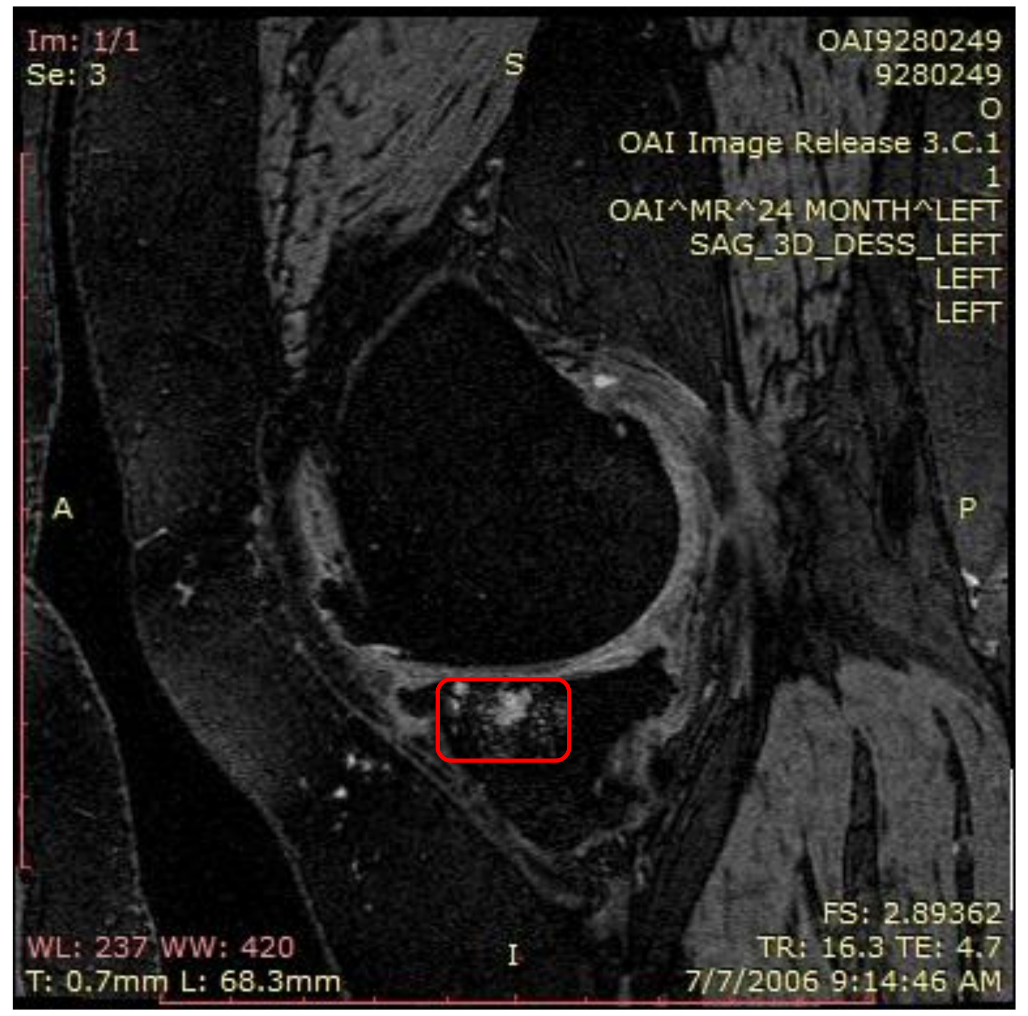


Fig 1: Sample Bone Marrow Lesion in DESS MRI scan from dataset

	Mean	Standard Deviation
Age (*years*)	63.2	10.2
BMI (*kg/m^2*)	29.67	4.78

Fig 2: Cohort Descriptives: Our dataset was generated with scans from *The Osteoarthritis Initiative* (https://nda.nih.gov/oai/)

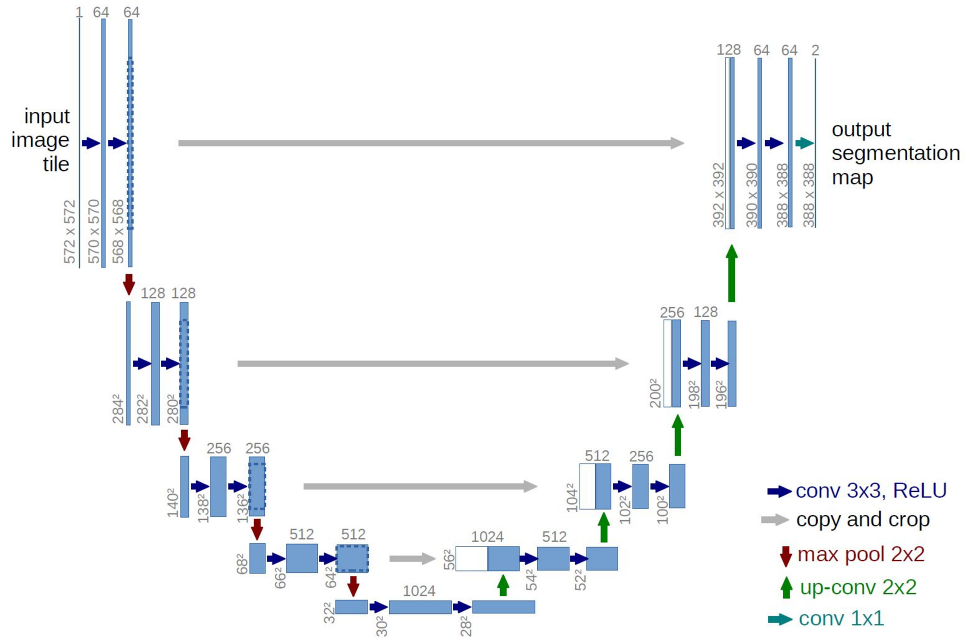


Fig 3: Standard UNET neural network architecture. Fully-convolutional neural network designed to provide high-quality image segmentations while being trained on a relatively small dataset. This structure is known as a fully-convolutional neural network because convolution operations are applied in the first and second halves of the model.

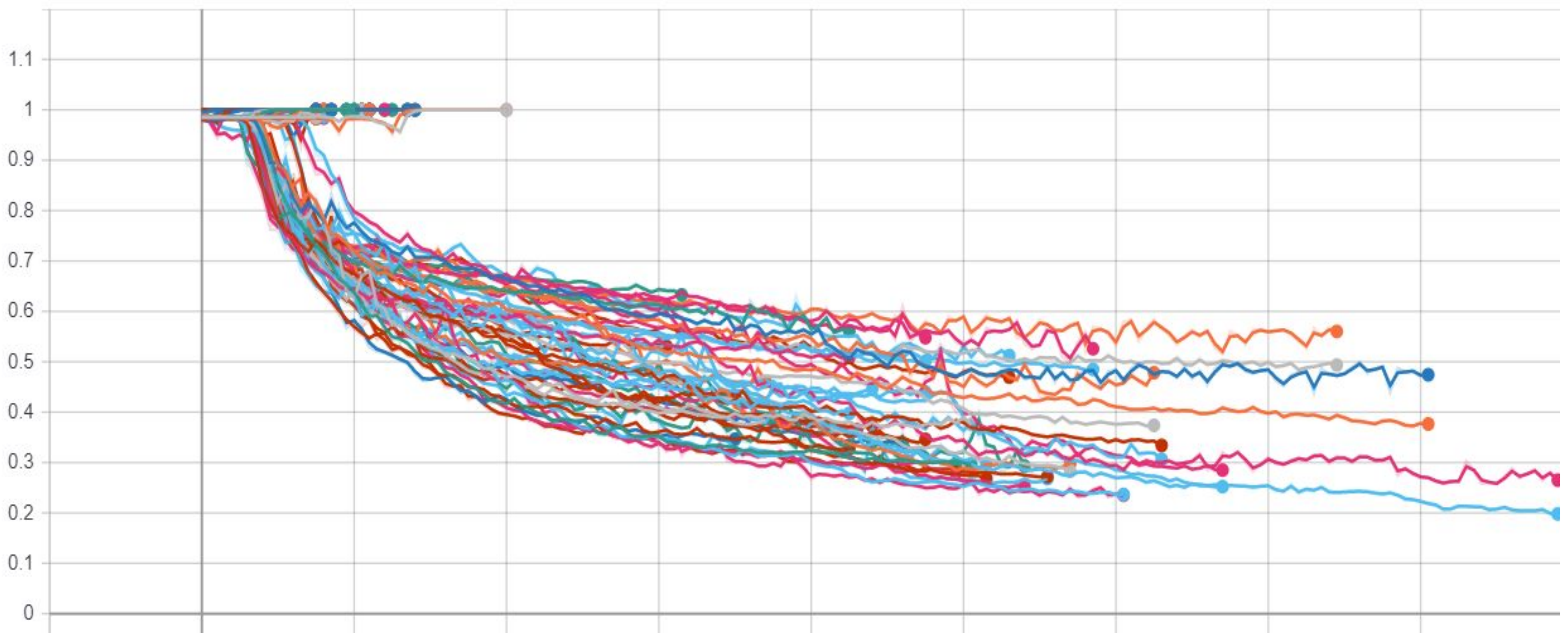


Fig 4: Graph of the loss functions during training of 80 different model configurations. Hyperparameter variables include the learning rate, the number of starting filters, the kernel size, and the dropout rate for neurons in hidden layers. Certain hyperparameter configurations failed to locate a local minima, thus converging at 1 (maximum possible loss). The early-stopping callback was used to stop a specific model's training once there has been no significant improvement (delta 0.001 / 15 epochs for validation IoU).

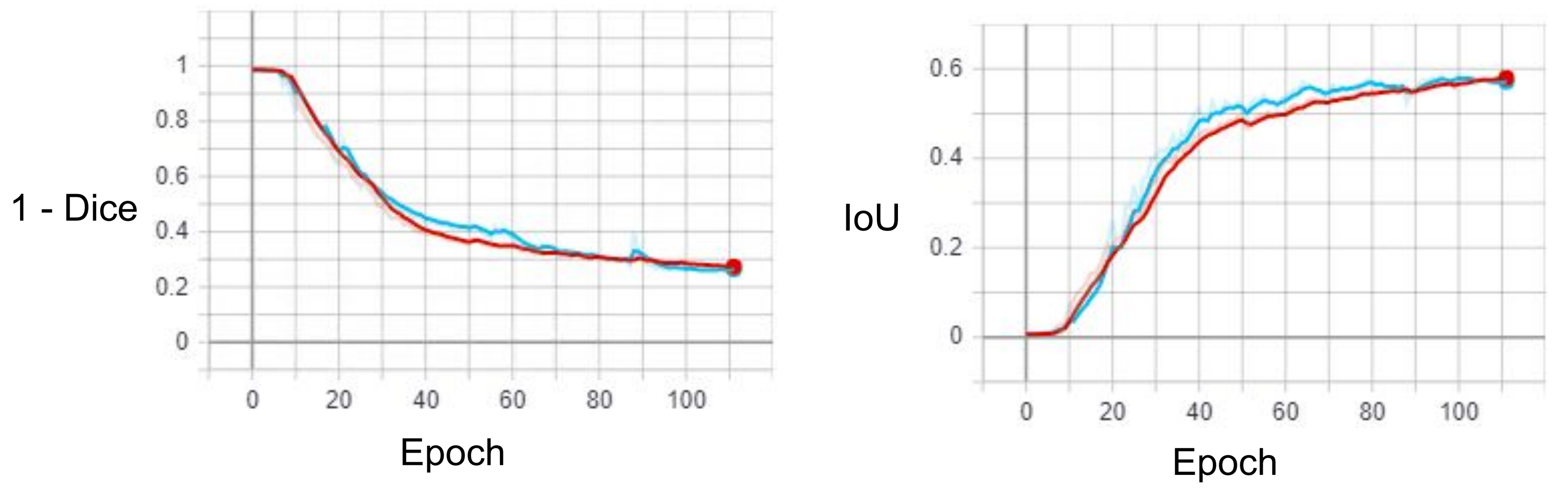


Fig 5: Loss and Metric graphs of best performing models. The best performing model, selected from the 80 aforementioned, achieved a minimum 1-Dice score of 0.2907 and a maximum IoU score of 0.5443 on the validation set. This model scored a 1-Dice coefficient of 0.3022 and an IoU of 0.5404 on the test dataset.

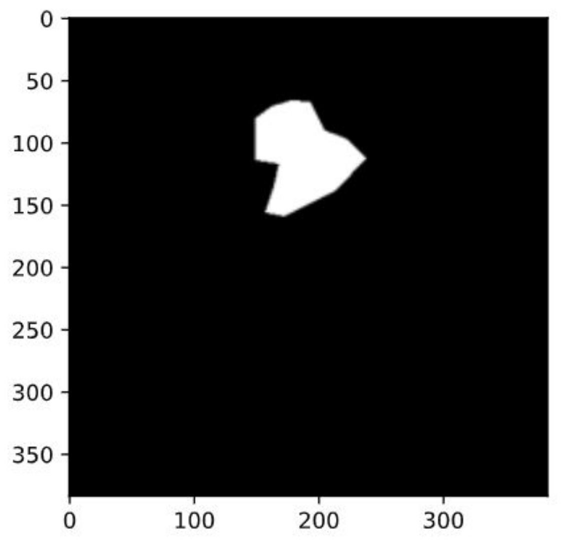


Fig 6: Severe class imbalance (136:1) in our dataset proved difficult for the BCE loss function.

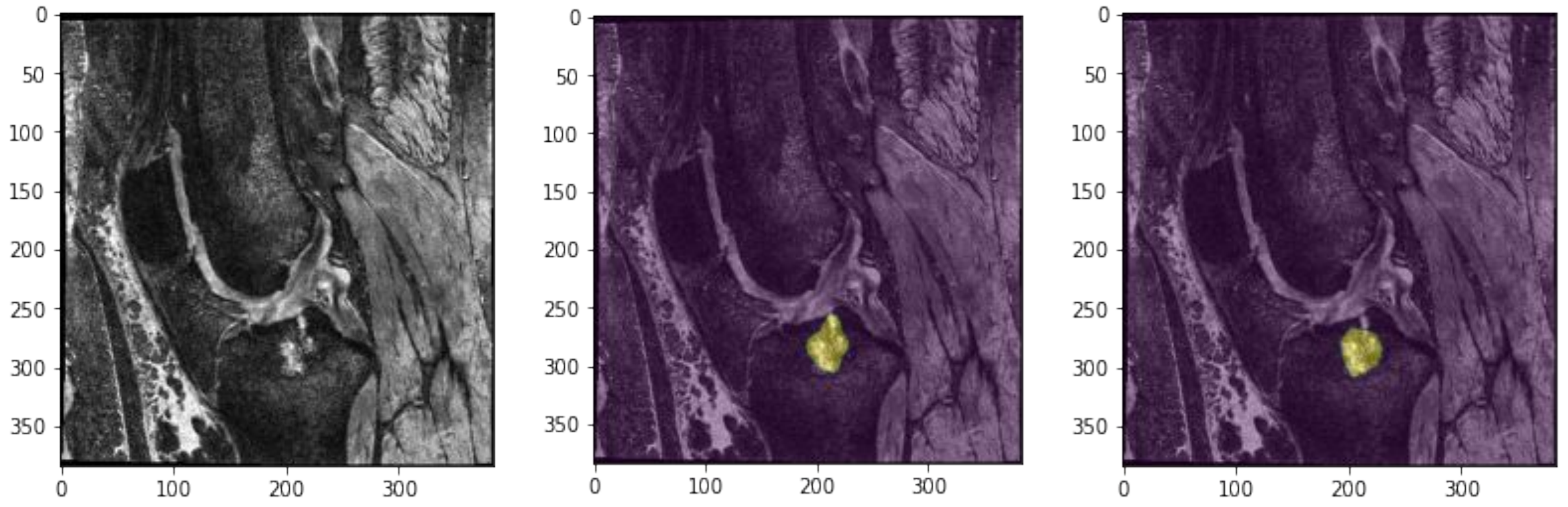


Fig 7: From left to right: ground truth (image fed into network), model prediction (notice high-fidelity), ground truth. Visual confirmation of test scans suggest the model is able to distinguish between ROI and nearby cartilage, bone, etc.)

Results

- The best performing model achieved a minimum 1-Dice score of 0.2907 and a maximum IoU of 0.5443. These results were generated from the validation dataset (3% of training - not used to backpropagate).
- This model achieved a 1-Dice coefficient of 0.3022 and an IoU of 0.5404 on the test dataset.
- The hyperparameter configuration used to obtain these scores were:
  - 10^{-4} learning rate
  - 28 starting filters (convolutional)
  - (3x3) kernel size (convolutional)
  - 50% dropout rate
- Training averaged ~90 epochs excluding models that converged at 1.
- Binary cross-entropy, a loss function typically used in deep learning, specifically convolutional neural networks, did not perform well. This was most likely because of the severe class imbalance present in our dataset (136 background for every 1 foreground pixel).
- The model performance on single-slice (2D) scans suggests that a 3D architecture (which would hypothetically examine adjacent slices or the entire stack) would not justify the significant increase in hardware memory requirement. Instead, our 2D architecture could be optimized to achieve higher accuracy.

Conclusion

- The study's objective was to fully-automate the task of segmenting bone marrow lesions in 2D MRI scans of the knee.
- The subsequent evaluation of the neural network we developed suggest this task was accomplished to a reasonable degree of accuracy.
- We've reported our findings, including network structure, hyperparameter configuration, and evaluation on unseen test data.
- Future improvements include adjusting for scanner inconsistencies and training with a larger dataset.

Acknowledgment

This research was made possible by JDMI AIF (Naraghi, Ali), UofT UofM International Training, CIHR Project Grant (PJT-156274), CIHR Project Grant (PJT-166012), CIHR Operating Grant (MOP-115094), John Lynch of OAI Imaging, and the co-authors listed above.

McMaster University

UNIVERSITY OF TORONTO
DALLA LANA SCHOOL OF PUBLIC HEALTH

CIHR IRSC

UHN COURAGE LIVES HERE

Osteoporosis Canada
Ostéoporose Canada