

Deep Learning for Automated Segmentation of the Hip Joint in X-ray Images

A study of the accuracy of a ResUNet-based approach for predicting the minimum joint space width along the weight-bearing part of the hip joint in a 2D image, in comparison to BoneFinder ground-truth data

By: Dragoş Ileana¹
 Supervisor(s): Jesse Krijthe¹, Gijs van Tulder¹, Myrthe van den Berg¹
¹EEMCS, Delft University of Technology, The Netherlands

June 24, 2024

Introduction

- **Hip osteoarthritis (HOA)** is a chronic disease affecting the hip joint cartilage and bones, causing pain and stiffness, and with no permanent cure [6].
- **Joint space width (JSW)** is one key parameter for grading OA severity on radiographs. It is estimated manually by measuring the narrowest point across the articulation as the minimum JSW (mJSW) [5][2][9]. However, there can be clinically significant inconsistencies between readers when measuring the JSW for the same patient [5][1].
- **Automated tools** for estimating JSW represent one solution to the reader variability problem, thanks to the expected increase in measurement precision and consistency with this approach [2][1].

Research Question

How accurate is a ResUNet based deep learning approach for predicting the minimum joint space width along the weight-bearing part of the hip joint in a 2D image, in comparison to ground-truth data generated by the BoneFinder algorithm?

In other words, this research aims to develop a pipeline that:

1. preprocesses X-ray images and automatically generates true labels (i.e., segmentation masks);
2. trains a Residual U-Net (ResUNet) deep learning architecture to segment pelvic radiographs;
3. uses the ResUNet segmentations to estimate the mJSW;
4. compares predicted segmentation masks and mJSW estimates to the ones extracted from the ground-truth labels.

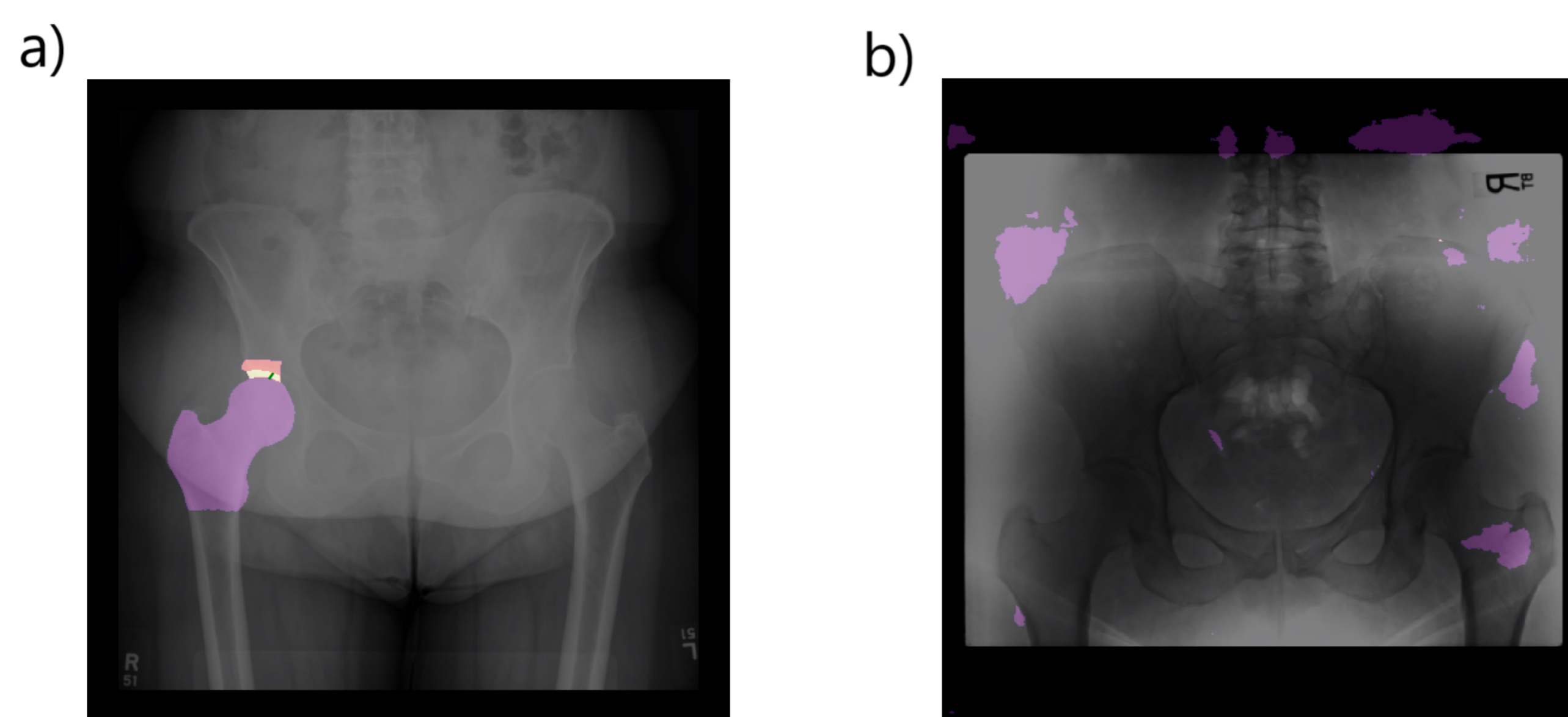


Figure 1. Successfully predicted mask (left) versus failed prediction (right).

Methodology

This research employs deep learning approach for image segmentation, which is trained on automatically generated labels to highlight the hip joint components in X-ray images. Then, an additional algorithm identifies the contours of the segmented joint bones and computes the mJSW.

- **The BoneFinder algorithm** [8][7] produces a set of points outlining the various objects to be segmented in the image. Ground-truth labels (i.e., segmentation masks) are generated using these landmarks.
- **The Residual U-Net (ResUNet)** deep network used for highlighting objects in radiographs is a modified version of U-Net, an encoder-decoder CNN-based architecture widely adopted for medical image segmentation [10]. Residual blocks replace the convolution units in the original U-Net, addressing the "degradation problem" associated with very deep architectures (i.e., decrease in performance for networks with larger number of layers) [4].
- **Minimum JSW** is estimated as the smallest point-to-point distance between the lower and upper borders of the joint space (i.e., between the femoral head and acetabular roof) identified in the segmentation masks predicted by ResUNet.
- **The X-ray data** used in this research was acquired from the CHECK and OAI initiatives [11][3].

Two sets of experiments were conducted, (1) for analysing the ResUNet performance using a selected set of model configurations and (2) for comparing the segmentation masks and mJSW estimates predicted using ResUNet to those extracted from the BoneFinder ground-truth labels.

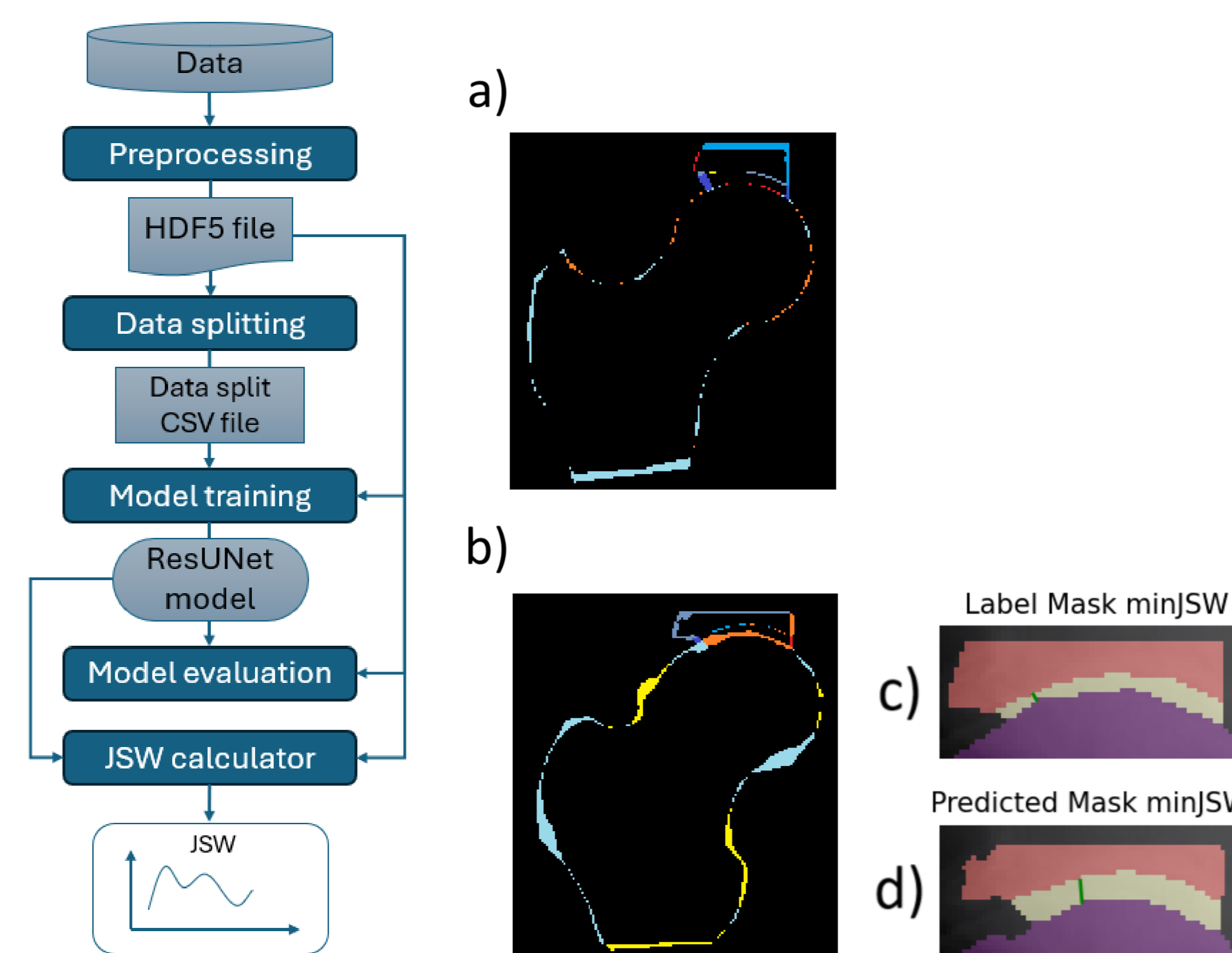


Figure 2. **Left:** experiment pipeline. **Middle:** pixel-by-pixel contrasts between predicted mask (warm colours) and ground-truth mask (cold colours) for (a) a good segmentation and (b) a worse segmentation. **Right:** differences in mJSW measurements estimated from (c) ground-truth mask ($mJSW = 0.2986\text{ mm}$) and (d) predicted mask ($mJSW = 0.9658\text{ mm}$).

Experiment Results

The first batch of experiments proceeded with establishing a baseline ResUNet configuration, using softmax output layer activation, Dice loss function, Dice evaluation metric, training and validation datasets with 1600 and 200 samples, respectively, and batches of size 20.

Experimenting with different modifications of the baseline model revealed the following:

1. Using a large training dataset does not produce a significant improvement in the test score (see table 1). This lack of improvement when using a large training dataset might be caused by the ground-truth labels do not cover the entire bone structures, wrongfully penalizing the loss during training. This problem can be addressed in future work by generating *weight maps*, clamping loss values to zero for specified regions of the masks.
2. Training with a small training dataset for the same number of steps produces a considerably smaller test metric, likely due to overfitting;
3. The softmax output layer activation is preferred over the sigmoid, as the latter may assign more than a single label to each pixel;
4. Both the Dice and Dice-Cross-Entropy loss functions performed well, with the Cross-Entropy having a more unstable validation curve.

Experiment Results

The second set of experiments trained and evaluated the baseline configuration on all the data available, giving a high Dice score. Using this final model, relatively small differences in mJSW estimations between predicted masks and real labels were observed (see table 5).

Testing dataset size	Validation dataset size	Test mean Dice metric
23544	2944	0.9136
1600	200	0.8877

Table 1.

Testing dataset size	Validation dataset size	Test mean Dice metric
1600	200	0.8877
160	200	0.7717

Table 2.

Loss function	Test mean Dice metric
Dice	0.8877
Dice + Cross-Entropy	0.8790
Cross-Entropy	0.8672

Table 3.

Output layer activation	Test mean Dice metric
Softmax	0.8877
Sigmoid	0.8859

Table 4.

Mean mJSW	mJSW standard deviation
0.0763	0.0874

Table 5.

Conclusion

The automated tool for mJSW estimation, based on the ResUNet model and using automatically generated ground-truth labels, produced accurate predictions for the segmentation masks. At the same time, relatively small differences were observed for the mJSW estimations between the ground-truth labels and predicted masks. One advantage of this approach is that it offers the possibility of visualizing where the mJSW was measured, as opposed to a less explainable model. Nevertheless, this approach presents some limitations, such as false penalization to the loss function unrelated to the inherent structure of the X-ray bones, large contribution to the loss function from the central part of the object masks and less from the bordering regions.

References

- [1] Anne Mathilde Andersen, Benjamin S. B. Rasmussen, Ole Graumann, Søren Overgaard, Michael Lundemann, Martin Haagen Haubro, Claus Varnum, Janne Rasmussen, and Janni Jensen. Minimal hip joint space width measured on x-rays by an artificial intelligence algorithm—a study of reliability and agreement. *BioMedInformatics*, 3(3):714–723, 2023.
- [2] P.G. Conaghan, D.J. Hunter, J.F. Maillefert, William Reichmann, and Elena Losina. Summary and recommendations of the oarsi fda osteoarthritis assessment of structural change working group. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*, 19:606–10, 03 2011.
- [3] F. Eckstein, W. Wirth, and M. Nevitt. Recent advances in osteoarthritis imaging—the osteoarthritis initiative. *Nat Rev Rheumatol*, 8, page 622–630, 2012.
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. pages 770–778, 06 2016.
- [5] J. H. Kellgren and J. S. Lawrence. Radiological assessment of osteo-arthrosis. *Annals of the Rheumatic Diseases*, 16(4):494–502, 1957.
- [6] Michelle J Lespasio, Assem A Sultan, Nicolas S Piuze, Nicolas S Khlopas, M Elaine Husni, George F Mutschler, and Michael A Mont. Hip osteoarthritis: A primer. *The Permanente Journal*, 22(1):17–084, 2018.
- [7] Claudia Linder. Bonefinder, 2024. Last accessed 23 June 2024.
- [8] C. Lindner, S. Thiagarajah, J.M. Wilkinson, G.A. Wallis, and T.F. Cootes. Fully automatic segmentation of the proximal femur using random forest regression voting. *IEEE Trans Med Imaging*, 08 2013.
- [9] G Neumann, D Hunter, M Nevitt, LB Chibnik, K Kwok, H Chen, T Harris, S Satterfield, and J Duryea. Health abc study. location specific radiographic joint space width for osteoarthritis progression. *Osteoarthritis Cartilage*, 2009.
- [10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [11] Janet Wesseling, Maarten Boers, Max A Viergever, Wim KHA Hilberdink, Floris PJG Lafeber, Joost Dekker, and Johannes WJ Bijlsma. Cohort Profile: Cohort Hip and Cohort Knee (CHECK) study. *International Journal of Epidemiology*, 45(1):36–44, 08 2014.