

Article

Fusion of Global and Local Features for Bone Marrow Lesion Segmentation Using a Hybrid Deep Learning Model

Devi Sowjanya Padala ¹, Lin Li ^{2,*}, Hetali Tank ³ , Ming Zhang ^{4,*}, Jeffrey B. Driban ⁵ , Timothy McAlindon ⁶ and Juan Shan ^{3,*} 

¹ Department of Mathematics, College of Science and Engineering, Seattle University, Seattle, WA 98122, USA; dpadala@seattleu.edu

² Department of Computer Science, College of Science and Engineering, Seattle University, Seattle, WA 98122, USA

³ Department of Computer Science, Seidenberg School of CSIS, Pace University, New York City, NY 10038, USA; hc18137n@pace.edu

⁴ Department of Computer Science, Boston University, Boston, MA 02215, USA

⁵ Department of Population and Quantitative Health Sciences, UMass Chan Medical School, Worcester, MA 01655, USA; jeffrey.driban@umassmed.edu

⁶ Division of Rheumatology, Department of Medicine, UMass Chan Medical School, Worcester, MA 01655, USA; timothy.mcalindon@umassmed.edu

* Correspondence: lil@seattleu.edu (L.L.); mzhang2@bu.edu (M.Z.); jshan@pace.edu (J.S.)

Abstract

Segmentation of bone marrow lesions (BML) is vital for quantifying lesion volume, a biomarker associated with cartilage damage and pain in knee osteoarthritis (KOA). Manual segmentation is challenging due to low contrast and variable lesion locations. We propose an automated deep learning method featuring a dual-stream encoder that integrates global image-level and local patch-level features. The study included 300 participants from the Osteoarthritis Initiative (OAI) database, each with approximately 36 intermediate-weighted fat-suppressed (IWFS) magnetic resonance (MR) images. The ground truth masks were manually annotated by trained research staff. With physical batch size 32, the model achieved a 2D Dice similarity coefficient (DSC) of 0.68, 3D DSC of 0.62, Intersection over Union (IoU) of 0.51, precision of 0.76, sensitivity of 0.62, and Pearson's correlation coefficient (r) of 0.85 between manually labelled and automatically generated volumes. Using an effective batch size of 64 via gradient accumulation, the model achieved 2D DSC of 0.63, 3D DSC of 0.65, IoU of 0.48, precision of 0.75, sensitivity of 0.6, and r of 0.98 for volume correlation. The model outperformed baselines at batch size 32 across almost all evaluated metrics and remained robust at batch size 64, with strong volumetric correlation and improved 3D DSC, IoU, and sensitivity.

Keywords: knee osteoarthritis; bone marrow lesions; deep learning; fusion; segmentation; global and local features; computer-aided diagnosis



Academic Editor: Ignacio Bosch Roig

Received: 22 March 2026

Revised: 7 June 2026

Accepted: 10 June 2026

Published: 8 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Osteoarthritis (OA) is the most common form of arthritis and a leading cause of pain and work disability. Its prevalence is significant and growing; as of the CDC's latest reporting, about 33 million U.S. adults have OA [1]. On a global scale, in 2019, approximately 528 million people were living with OA, representing a 113% increase in age-standardized prevalence rates since 1990 [2]. OA frequently affects weight-bearing

joints, with the knee being the most common site, often causing severe pain and substantial socioeconomic impact.

A key feature in understanding knee OA is the bone marrow lesion (BML), an area of altered subchondral bone structure (e.g., edema, fibrosis, decreased mineral content, altered trabecular morphometry [3–9]) visible on Magnetic Resonance Imaging (MRI). Research has established a strong relationship between BMLs and pain and structural progression of knee OA [10–12]. Consequently, BML volume emerges as a critical biomarker for evaluating the disease status, and reducing BML size is now considered an important therapeutic goal for managing the disease [13,14]. To quantify the total volume of BMLs in a knee, it is critical to segment BMLs in each slice of a knee MRI sequence.

Manual segmentation of BML remains a time-consuming and labor-intensive process, placing a heavy burden on radiologists and research staff. Moreover, inherent challenges, such as low contrast between lesion and surrounding tissues, fuzzy lesion boundaries, and inter-observer variability, often lead to inconsistent annotations. In the field of BML segmentation, most methods are semi-automated [15–17]. In 2010, Dijkstra et al. introduced a KNN-based approach for semi-automated segmentation of BMLs [15]. On T2w SPIR MRIs, the method achieved a mean Dice Similarity Index (DSI) of 0.702 (± 0.202). This method required users to select regions of interest (ROI) prior to classification, after which each voxel within those regions was treated as an individual sample. Voxels were assigned labels of lesion (1) or non-lesion (0) based on expert manual segmentations, and the KNN classifier then produced the final lesion maps within the predefined areas. The manual step of selecting ROI limits automation and might introduce operator bias and inconsistency between users. In 2013, Pang et al. [16] proposed a semi-automated segmentation approach that detects, extracts, and quantifies BML structures. In this work, they used a custom graphical user interface to identify the boundaries of the tibia and femur in each MRI slice by marking multiple points along the articular surface. To determine a threshold between bright and normal areas, they used the false discovery rate (FDR) to construct initial BML boundaries curves, and then a region-based curve evolution method was used to refine them to obtain an accurate BML border. The efficiency of this method was evaluated by performing intra- and inter-tester reliability tests in addition to quantifying BML volume. Although curve evolution refines boundaries of BML, this method remains operator-dependent and manual-intensive. In 2022, Preiswerk et al. [17] developed and validated a deep learning method using a U-Net model to segment BMLs from knee MRI datasets. This method achieved an average 2D Dice Similarity Score of 0.70 and Pearson's correlation coefficient of 0.94. However, this work still requires initial BML identification, performed by a human reader with a mouse click. These semi-automated approaches are often limited to predefined areas, which require human intervention and introduce variability.

To address the limitations in semi-automated approaches, fully automated BML segmentation methods have been developed. In 2013, Dodin et al. [18] proposed a fully automated system for quantifying knee BML using MRI in the Osteoarthritis Initiative cohort. In this approach, BML was automatically detected using mathematical morphology operators and filtering to remove unrelated structures. They achieved DSC values of 0.4–0.6. In recent years, deep learning has become the dominant paradigm for medical image segmentation due to its ability to learn hierarchical feature representations directly from data. Transformer-based methods (e.g., TransUNet or Vision Transformers) [19,20] and other emerging architectures, such as Mamba-based state-space models [21] and diffusion-based generative models [22], have also demonstrated strong performance in general medical image segmentation tasks. However, in the context of BML segmentation, convolutional neural network (CNN)-based methods remain the predominant and most extensively

studied approach in the existing literature, largely due to their robustness in limited-data settings and their effectiveness in capturing fine-grained spatial details. One of the most successful CNN-based architectures is the U-Net proposed by Ronneberger et al. [23]. U-Net is a symmetric U-shaped encoder–decoder framework with skip connections that preserve spatial detail while capturing high-level semantics. Its significance lies in its specialization for image segmentation and its ability to achieve accurate, efficient segmentation. In 2024, Ponnusamy et al. proposed a fully automated segmentation approach using a deep learning method on a dataset of 300 subjects [24]. In this paper, the authors proposed a U-Net + InceptionResNetV2 model that has achieved a Pearson correlation coefficient of 0.98, a 2D Dice similarity coefficient (DSC) of 0.68, and a 3D DSC of 0.60. Stirling et al. proposed an approach for automated detection and segmentation of bone and traumatic BML from MRI [25]. This approach addresses the labor-intensive manual assessment of BMLs by using a 3D U-Net with morphological post-processing. They achieved a DSC of 0.76 (± 0.13). In all these approaches, whole images are used as input to the U-Net, which has notable limitations for BML segmentation. Since lesions are often small and irregularly shaped, downsampling operations in the encoder can dilute or completely lose these fine details, reducing sensitivity to subtle lesions. Moreover, when the lesion occupies only a very small portion of the image, the overwhelming background context can bias the network toward missed detections.

Multi-scale, dual-stream, and patch-aware architectures [26,27] have been explored in medical image segmentation to improve the integration of local and global contextual information. These approaches often combine whole-image representations with patch-level or multi-resolution features to enhance segmentation performance, particularly for small targets. However, such architectures have not been investigated in the context of BML segmentation. To address the limitations of using images solely as inputs, this study proposed a hybrid deep learning model that fuses image and patch features for fully automated segmentation of BML. By incorporating patch-level information alongside the whole image, the model can capture local context and fine-grained texture patterns, improving sensitivity to subtle lesions while preserving global information and overall bone structure. The rest of the paper is organized as follows: in Section 2, we describe the dataset and discuss the hybrid deep learning model. Section 3 presents the experiment setup and results. The comparison of the proposed method with several other state-of-the-art methods is also included in Section 3. In Section 4, we discuss the contributions and limitations. Section 5 concludes the work.

2. Materials and Methods

2.1. Dataset

The dataset employed in this study is sourced from the publicly available Osteoarthritis Initiative (OAI) database [28]. OAI is a multi-center longitudinal cohort sponsored by the National Institutes of Health and private partners. Established to discover biomarkers for knee OA onset and progression, OAI recruited 4796 adults (ages 45–79 years) with or at elevated risk for symptomatic knee OA across multiple clinical centers. The initiative collected a comprehensive public dataset featuring serial MRIs and radiographs, making it a premier resource for musculoskeletal research. Multiple MRI sequences were obtained, including a sagittal intermediate-weighted turbo spin-echo with fat suppression (IWFS). All study procedures were approved by the institutional review boards of the coordinating and participating clinical centers. Written informed consent was obtained from every participant.

In this study, we used MRI scans of 300 participants' knees from the public OAI dataset. The IWFS sequence was chosen for analysis as it provides enhanced sensitivity for

visualizing subchondral BMLs, offering optimal clarity in both their size and morphological appearance. The BML ground truth was labeled using validated semi-automated software by trained readers [29]. All BML segmentation results were reviewed for quality by a second reader and adjusted when necessary. Our reader had good intra-reader reliability among 20 knees measured twice with ≥ 48 h between readings: Intraclass Correlation Coefficient (ICC) of 0.98 for baseline BML volume. Subsequently, all IWFS images were preprocessed by resizing them to a uniform 448×448 pixels and normalizing the intensity values to a consistent range of 0–255. Furthermore, Contrast Limited Adaptive Histogram Equalization (CLAHE) [30] was applied to locally enhance contrast, thereby making subtle structures such as lesions more visible. Finally, a Gaussian blur filter was used to reduce small-scale noise and smooth the image while preserving important structural edges. These preprocessing steps produce images that are both clearer and more consistent, making them suitable for training a segmentation model. Alternative preprocessing strategies were also explored during preliminary experiments; however, the selected combination of CLAHE and Gaussian filtering provided stable and consistent performance and was therefore adopted in the final pipeline.

The dataset was divided into 70% training (210 participants), 15% validation (45 participants), and 15% test (45 participants). Only those slices containing visible BMLs were included. In total, 1034, 194, and 209 images with BMLs were used in training, validation, and testing, respectively. No slices from the same participant appeared across training, validation, and testing, ensuring a fully participant-disjoint evaluation and preventing data leakage.

2.2. Model Architecture

To combine both the image and the patches of the image for segmentation, we propose a segmentation model that contains two branches: one branch takes the image as input and adopts a pretrained InceptionResNetV2 as the encoder for the image; the other branch extracts patches from the image, and a patch encoder path is built to capture localized contextual cues from patches. The outputs of the image encoder path and patch encoder path are concatenated and passed to the upsampling decoder path. The model architecture is shown in Figure 1.

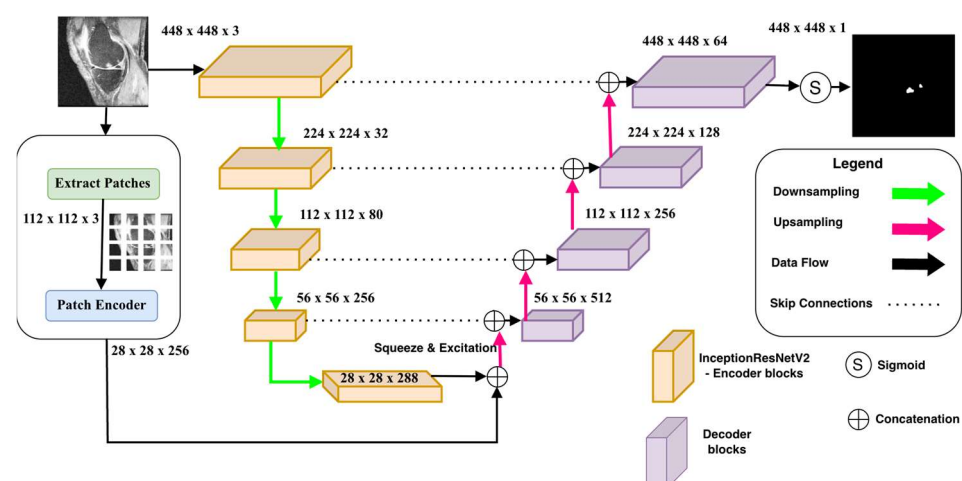


Figure 1. The proposed hybrid segmentation model architecture.

The encoder path of the image branch and the decoder path in our proposed model architecture are built on the model introduced by Ponnusamy et al. [24]. For completeness, we briefly summarize the key components relevant to our work. The image encoder path consists of four convolutional blocks, each with two 3×3 convolution layers. The

encoder and bottleneck layers use InceptionResNetV2, a 164-layer network pretrained on more than a million images from the ImageNet database. The pretrained InceptionResNetV2 model requires a three-channel (red, green, blue) input with a dimension of $448 \times 448 \times 3$ pixels; therefore, the input images were converted from a single-channel grayscale MRI scan to a three-channel image by replicating the single channel across all three RGB channels. The decoder path also contains four convolutional blocks. Each decoder block has a 2×2 transposed convolution layer, followed by a concatenation layer that combines the encoder-extracted feature maps, and then a convolution block with two 3×3 convolution layers. The final layer is a 1×1 convolutional layer that outputs a binary image as the lesion segmentation result. For details of encoder and decoder blocks, please refer to the original work [24].

In this work, we extended the architecture in [24] by adding a patch branch in the encoder path. In the patch branch, patches of 112×112 pixels were extracted from a 448×448 image using a non-overlapping sliding window (stride = 112), with a patch dimension of $112 \times 112 \times 3$, where 3 is the number of channels (red, green, blue). Each $112 \times 112 \times 3$ patch was passed through a patch encoder block (Figure 2), which had an InceptionResNetV2 encoder that progressively downsampled the spatial resolution while increasing the channel depth. The encoder within this patch branch utilized intermediate feature activations from progressively deeper layers of the InceptionResNet architecture. Specifically, feature maps were extracted from three successive downsampling stages of the encoder, with spatial resolutions of $56 \times 56 \times 32$, $28 \times 28 \times 80$, and $14 \times 14 \times 256$, respectively. From the $14 \times 14 \times 256$ representation, an AveragePooling2D layer with pool_size = 2 and strides = 2 was applied to obtain a $7 \times 7 \times 256$ patch bottleneck.

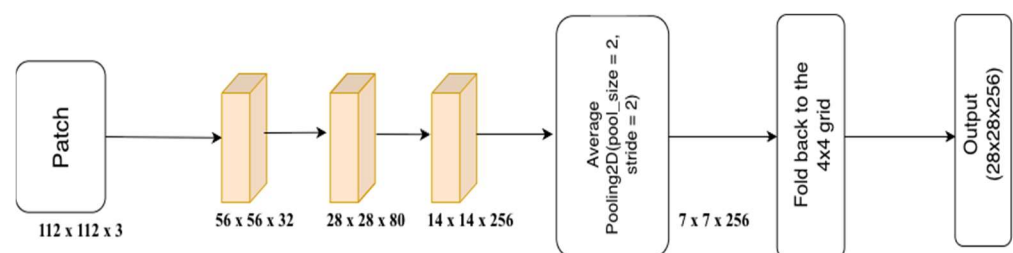


Figure 2. Architecture of the patch encoder in our hybrid model.

This average pooling step was intentionally used instead of an additional inception block to achieve a smooth spatial reduction without increasing network complexity. More importantly, pooling to a 7×7 resolution ensured that when the 16 patch embeddings (arranged in a 4×4 grid) were folded back, the resulting feature map became $28 \times 28 \times 256$, which closely aligned with the $28 \times 28 \times 288$ bottleneck feature map of the full-image InceptionResNetV2 encoder path.

The patch-based feature map ($28 \times 28 \times 256$) was then fused with the bottleneck feature map produced by the image-branch InceptionResNetV2 encoder ($28 \times 28 \times 288$). Finally, the fused representation was refined using a Squeeze-and-Excitation (SE) module [31] to activate the effective channels and suppress the less useful ones.

2.3. Loss Function

The loss function is identical to that of Ponnusamy et al. [24], defined as

$$\text{Loss} = L_{BCE} + L_{Dice} \quad (1)$$

where L_{BCE} and L_{Dice} denote binary cross-entropy loss and Dice loss, respectively. Please refer to work [24] for details.

3. Experiment and Results

3.1. Experiment Setup

We conducted all experiments using the TensorFlow framework, and model training was performed on a single-core NVIDIA H100 GPU with 70 GB of memory. In this experiment, we utilized the Adam optimizer with a learning rate of 0.001. All models were trained for 100 epochs with a batch size of 32, which is the maximum physical batch size supported by our single-GPU hardware. A dropout rate with a ratio of 0.3 was used to prevent the network from overfitting during training. The training process was halted if validation accuracy failed to improve after a certain number of epochs. To ensure a fair comparison, all baseline models, including U-Net [17], U-Net + EfficientNetB7 [32], Sharp U-Net [33], U-Net + ResNet-34 [34], and U-Net + InceptionResNetV2 [24], were reimplemented and trained from scratch on the same dataset splits described in Section 2.1 using the same GPU hardware (NVIDIA H100). For each baseline model, the hyperparameters, including learning rate, optimizer, and dropout rate, were configured according to the settings reported in their respective original publications to preserve the intended design of each architecture. All models were evaluated on the identical test set using the metrics described in Section 3.2.

Due to GPU memory constraints, it was not feasible to train our model with a physical batch size of 64. To enable a fair comparison with the results reported in [24], we adopted a gradient accumulation-based batch processing strategy to simulate a batch size of 64 without exceeding GPU memory limits. Specifically, the model processed 8 images at a time (physical batch size = 8), computed the gradients for each mini-batch, and accumulated them over 8 steps before performing a single optimizer update. During each step, the loss was scaled by the number of accumulation steps to pre-average the gradients, ensuring that the accumulated gradient was mathematically equivalent to the average gradient computed from a true batch of 64 images. After 8 such steps, the Adam optimizer applied the accumulated gradient to update the model weights. In addition, the data pipeline was configured to load data batch-wise on the fly rather than loading the entire dataset into memory, further reducing memory usage.

The implementation of the proposed framework, including model architecture, training protocols, and data preprocessing, is publicly available at: <https://github.com/Sowjanya2502/Bone-Marrow-Lesions-Segmentations-on-MRI-Using-Deep-Learning> (accessed on 1 June 2026).

3.2. Evaluation Metrics

To evaluate the performance of the proposed model and validate its segmentation accuracy, we used the following metrics: 2D and 3D Dice Similarity Coefficients (DSC), Intersection over Union (IoU), Pearson correlation coefficient (r), precision, and sensitivity.

The 2D DSC is the most used metric in medical image segmentation tasks. It quantifies the per-slice spatial overlap between the predicted lesion regions and the manually annotated ground truth masks. The 2D DSC score ranges from 0 to 1, where 1 indicates perfect overlap, and 0 signifies no overlap. The 2D Dice formula is provided below, where true positive (TP) is the number of pixels correctly predicted as positive, false positive (FP) is the number of pixels incorrectly classified as positive, and false negative (FN) is the number of pixels incorrectly predicted as negative.

$$2D \text{ Dice} = \frac{2TP}{2TP + FP + FN} \quad (2)$$

The 3D DSC was proposed by Ponnusamy et al. [24] to evaluate volumetric overlap and volume-based segmentation accuracy at the subject level. The formula for 3D DSC is listed below.

$$3D\ Dice = \frac{2\sum_{i=1}^n |X_i \cap Y_i|}{\sum_{i=1}^n |X_i| + \sum_{i=1}^n |Y_i|} \quad (3)$$

where X_i is the predicted BML mask of slice i of a subject and Y_i is the corresponding ground truth of slice i , with the slice number starting from $i = 1$ to n , where n is the total number of slices with BML for that subject, and $|X|$ means the total number of non-zero pixels in image X [24].

IoU is another common metric for evaluating segmentation models. It measures the overlap between the predicted segmentation map and the ground-truth mask. A higher IoU value indicates greater overlap between the two, reflecting stronger similarity and more accurate segmentation performance. The IoU formula is listed below.

$$IoU = \frac{TP}{TP + FP + FN} \quad (4)$$

The Pearson correlation coefficient (r) is used to assess the linear relationship between the volumes of BMLs from segmentation results and ground truth masks, reflecting how well the volumes of BMLs across the two sequences align. The volume of BML is computed at the patient level by adding all the BML pixels from all slices in the 3D scan. A value of 1 represents a perfect positive correlation, -1 means a perfect negative correlation, and 0 indicates no linear correlation. The formula for calculating r is given in Equation (5) below.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (5)$$

In Formula (5), n is the number of subjects in the testing data set, x denotes the BML volume produced by the deep learning model, and y denotes the ground truth of the BML volume [24].

Precision indicates the fraction of predicted lesion pixels that are truly lesion pixels. It is calculated as the ratio of correctly predicted target pixels to the total number of target pixels classified in the prediction. The formula is provided in Equation (6).

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Sensitivity measures the proportion of actual lesion pixels correctly predicted. A higher sensitivity value indicates fewer missed detections and better coverage of the actual target regions. The formula is given in Equation (7).

$$Sensitivity = \frac{TP}{TP + FN} \quad (7)$$

3.3. Results

With a physical batch size of 32—the maximum physical batch size supported by our single-GPU hardware—our hybrid model achieved a 2D Dice similarity of 0.68, 3D DSC of 0.62, IoU of 0.51, a Pearson correlation coefficient of 0.85, a precision of 0.76, and a sensitivity of 0.62. We compared our method with several other segmentation models, including U-Net [17], U-Net + EfficientNetB7 [32], Sharp U-Net [33], U-Net + Resnet-34 [34], and U-Net + InceptionResNetV2 [24]. All the models were trained, validated, and tested on the same training, validation, and testing sets specified in Section 2.1. The batch size used for all the models was 32. The comparison result is summarized in Table 1.

Table 1. Comparing the proposed BML segmentation model with other deep learning models using the same dataset and batch size = 32.

Model	2D DSC	3D DSC	IoU	r	Precision	Sensitivity
U-Net [17]	0.52	0.51	0.29	0.58	0.52	0.32
U-Net + EfficientNetB7 [32]	0.54	0.50	0.30	0.73	0.73	0.46
Sharp U-Net [33]	0.55	0.52	0.38	0.71	0.74	0.44
U-Net + Resnet-34 [34]	0.61	0.56	0.39	0.73	0.60	0.42
U-Net + InceptionResNetV2 [24]	0.63	0.60	0.44	0.76	0.78	0.50
The proposed hybrid model	0.68	0.62	0.51	0.85	0.76	0.62

Overall, the proposed hybrid model consistently outperformed other segmentation models across 2D DSC, 3D DSC, IoU, r, and sensitivity, with only a minor reduction in precision compared to the model used in [24]. To ensure a fair comparison with the results reported in [24], which used a batch size of 64, we also trained our model using a gradient accumulation-based batch processing strategy to effectively simulate a batch size of 64 without exceeding GPU memory limits. The comparison results are shown in Table 2 for a batch size of 64. For comparison, we also computed the metrics IoU, specificity, and sensitivity for the model in [24], though the original paper [24] did not report results for those metrics.

Table 2. Comparing U-Net + InceptionResNetV2 model [24] with the proposed hybrid model using a batch size = 64.

Model	2D DSC	3D DSC	r	IoU	Precision	Sensitivity
U-Net + InceptionResNetV2 [24]	0.68	0.60	0.98	0.46	0.77	0.53
Proposed hybrid model	0.63	0.65	0.98	0.48	0.75	0.6

As shown in Table 2, with an effective batch size of 64, the proposed hybrid model achieved a 2D DSC of 0.63, 3D DSC of 0.65, volume correlation of $r = 0.98$, IoU of 0.48, precision of 0.75, and sensitivity of 0.6. Compared to the U-Net + InceptionResNetV2 model in [24], our model achieved lower 2D Dice and precision but improved 3D Dice, IoU, and sensitivity, while maintaining the same strong Pearson correlation with ground-truth lesion volume.

Since the proposed hybrid model integrates image patches, we also tested how patch size selection affects BML segmentation. The result is summarized in Table 3, which shows that with a batch size of 32, a patch size of 112×112 yields better results across all metrics than other patch sizes.

Table 3. BML segmentation result of the proposed hybrid approach using different patch sizes with a batch size of 32.

Patch Size	2D DSC	3D DSC	IoU	r	Precision	Sensitivity
28×28	0.63	0.53	0.47	0.72	0.68	0.62
56×56	0.63	0.58	0.47	0.76	0.59	0.59
112×112	0.68	0.62	0.51	0.85	0.76	0.62
224×224	0.61	0.52	0.45	0.69	0.60	0.45

To evaluate the individual contribution of each component in the proposed hybrid model, we conducted an ablation study by systematically removing one module at a time while keeping all other training conditions identical (batch size = 32). The ablation configurations include: (1) the baseline U-Net + InceptionResNetV2 model from [24], which uses only the image stream without the patch stream or the SE module; (2) the image stream with the SE module but without the patch stream; (3) the image stream with the patch stream but without the SE module; and (4) the full proposed hybrid model with the image stream, patch stream and the SE module. The results are summarized in Table 4.

Table 4. Ablation study of the proposed hybrid model components with batch size = 32.

Configuration	2D DSC	3D DSC	IoU	r	Precision	Sensitivity
U-Net + InceptionResNetV2 [24]	0.63	0.60	0.44	0.76	0.78	0.50
Image Stream + SE	0.60	0.62	0.49	0.78	0.76	0.58
Image Stream + Patch Stream	0.65	0.60	0.46	0.80	0.76	0.56
The proposed hybrid model	0.68	0.62	0.51	0.85	0.76	0.62

The ablation study demonstrates that both the patch stream and the SE module contribute positively to the performance of the proposed hybrid model. Compared with the baseline U-Net + InceptionResNetV2 model [24], adding the SE module improved 3D DSC (0.60 $\rightarrow$ 0.62), IoU (0.44 $\rightarrow$ 0.49), r (0.76 $\rightarrow$ 0.78), and sensitivity (0.50 $\rightarrow$ 0.58), although the 2D DSC and precision slightly decreased. Adding the patch stream improved 2D DSC (0.63 $\rightarrow$ 0.65), IoU (0.44 $\rightarrow$ 0.46), r (0.76 $\rightarrow$ 0.80) and sensitivity (0.50 $\rightarrow$ 0.56), indicating that the patch-based features enhanced segmentation and correlation performance. The proposed hybrid model, which combines both the patch stream and the SE module, achieved the best overall performance across most evaluation metrics, including the highest 2D DSC (0.68), 3D DSC (0.62), IoU (0.51), r (0.85), and sensitivity (0.62). These results suggest that the SE module and patch stream provide complementary benefits, and their integration leads to more accurate segmentation performance.

To assess the computational cost of the proposed hybrid model, we compared it with all baseline models in terms of total number of parameters, giga floating-point operations (GFLOPs), and average inference time per image. All models were evaluated under the same hardware conditions using a single NVIDIA H100 GPU. The results are summarized in Table 5. As shown in Table 5, the proposed hybrid model has 39.20 M parameters and 210.73 GFLOPs, representing an increase of approximately 6.5% in parameters and 6.2% in GFLOPs compared to the U-Net + InceptionResNetV2 baseline [24]. The average inference time is 236.72 ms per image, approximately 30% slower than baseline. This additional cost is primarily attributable to the patch encoder branch, which processes 16 patches per image through a separate InceptionResNetV2 encoder, as well as the inclusion of the SE module. Overall, these results indicate a moderate increase in computational overhead associated with the proposed dual-branch design. This additional computational cost reflects the trade-off between improved feature representation (i.e., combining global structural context with fine-grained patch-level information) and efficiency. While the current implementation is not optimized for real-time inference, it may be more appropriate for scenarios where offline processing is feasible and accuracy is prioritized. Further optimization would be required for time-sensitive clinical deployment.

Table 5. Computational cost comparison of segmentation models.

Model	Parameters (M)	GFLOPs	Inference Time (ms)
U-Net [17]	0.48	14.24	9.32
U-Net + EfficientNetB7 [32]	10.85	168.08	102.41
Sharp U-Net [33]	7.76	74.23	13.66
U-Net + Resnet-34 [34]	24.44	46.05	39.44
U-Net + InceptionResNetV2 [24]	36.79	198.43	182.00
The proposed hybrid model	39.20	210.73	236.72

Beyond quantitative evaluation, Figure 3 presents qualitative segmentation results from the proposed hybrid model for some sample images in the test set that had a high Dice score. The proposed model effectively delineated lesion boundaries across diverse cases, including small lesions, multiple lesions in different locations, and large contiguous lesions.

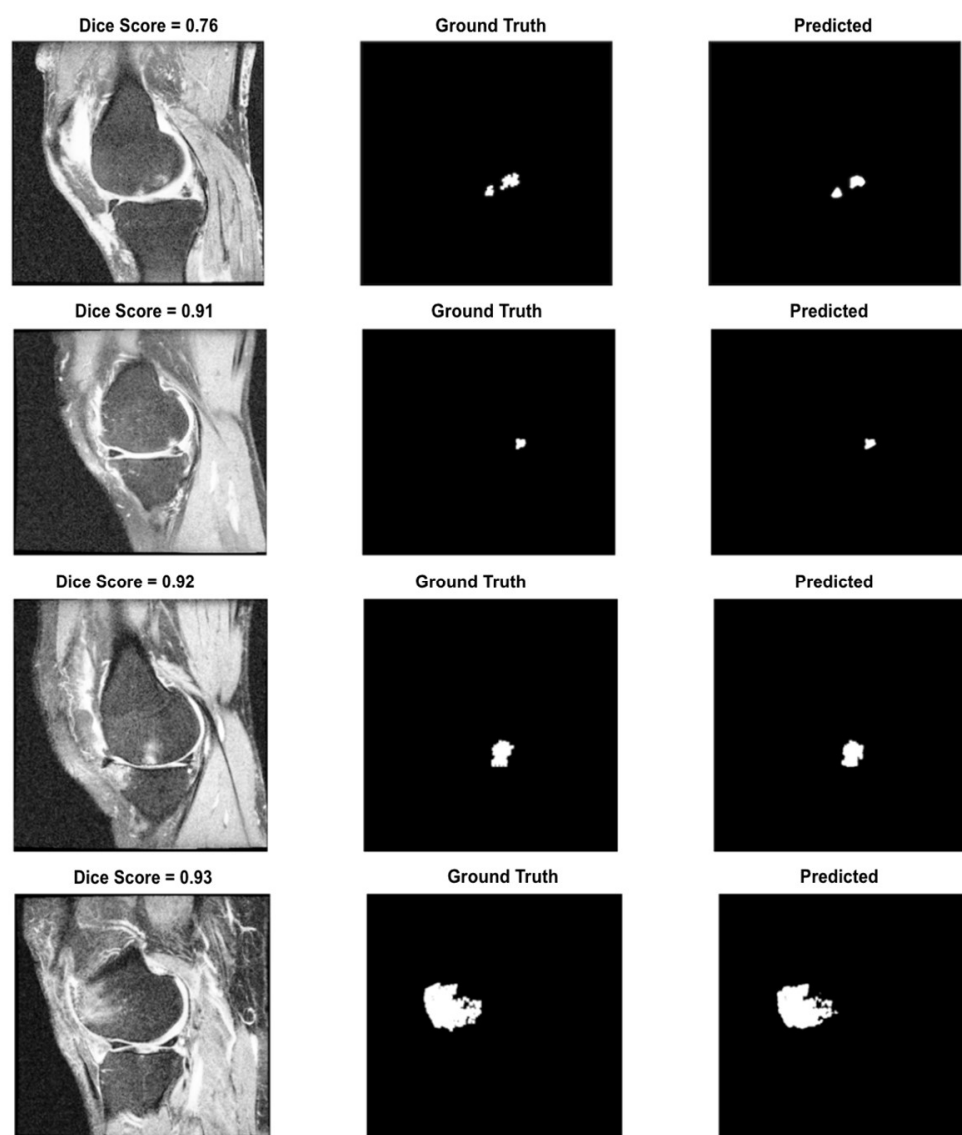


Figure 3. Sample MRI scan, their BML ground truth mask, and their BML segmentation result generated by the proposed hybrid model, showing cases with high Dice score.

3.4. Statistical Evaluation and Stability Analysis

To assess the robustness of the reported improvements, we conducted one-sided *t*-tests comparing the proposed hybrid model against the best-performing baseline model (U-Net + InceptionResNetV2 [24]) (batch size = 32). The null hypothesis states that the mean paired difference between the proposed model and the baseline (proposed–baseline) is less than or equal to zero. The analysis was performed using 209 images from 45 participants in the test set, with 95% confidence intervals (CIs) and statistical significance evaluated at $\alpha = 0.05$. The results are summarized in Table 6. In Table 6, “Mean” denotes the average performance of the proposed model across all evaluated samples, while “SD” represents the corresponding standard deviation, reflecting performance variability. The 95% CI indicates the estimated range within which the true mean performance is expected to lie with 95% confidence. “Threshold” refers to the corresponding baseline performance obtained from U-Net + InceptionResNetV2 [24], used as the reference for statistical comparison. *p*-values were not reported for *r* because *r* was computed as a single dataset-level statistic rather than a per-subject metric. Consequently, no sample distribution of *r* values was available for statistical comparison.

Table 6. Statistical significance analysis of the proposed model against U-Net + InceptionResNetV2 [24] baseline.

Metric	Mean	SD	95% CI	Threshold	<i>p</i> -Value
IoU	0.45	0.24	[0.4188, 0.4875]	0.44	0.23
Precision	0.74	0.2	[0.7117, 0.7982]	0.78	0.58
Sensitivity	0.60	0.3	[0.5474, 0.6404]	0.50	0.0004
2D DSC	0.67	0.27	[0.637, 0.715]	0.63	0.009
3D DSC	0.62	0.15	[0.5755, 0.6801]	0.60	0.02

The proposed model shows statistically significant improvements in sensitivity ($p = 0.0004$), 2D DSC ($p = 0.009$), and 3D DSC ($p = 0.02$), all below the significance threshold. IoU yielded a *p*-value of 0.23, indicating that the observed improvement in the mean over the baseline was not statistically significant. In addition, although the precision of the proposed model is lower than the baseline, no statistically significant difference was observed ($p = 0.58$). The 95% confidence intervals for sensitivity and 2D DSC lie entirely above their corresponding baseline thresholds, indicating consistent improvement over baseline performance across the evaluated samples. Although the mean 3D DSC exceeds the baseline threshold and achieves statistical significance ($p = 0.02$), its confidence interval slightly overlaps the baseline threshold, suggesting less consistent improvement across subjects. Similarly, IoU and precision also exhibit overlapping confidence intervals with their baseline thresholds, indicating greater variability in performance. These results suggest that the proposed model improves lesion detection sensitivity and overall segmentation performance. Figure 4 illustrates the distribution of *p*-values across metrics, showing that sensitivity, 2D Dice, and 3D Dice fall below the significance threshold ($\alpha = 0.05$), while IoU and precision are not statistically significant.

To further evaluate the stability and reproducibility of the proposed model, five independent training runs were conducted using different random weight initializations under identical experimental settings. Table 7 summarizes the results of each run together with the mean and standard deviation for all evaluation metrics. The standard deviations across all metrics remained below 0.05, indicating that the proposed model produces stable and consistent performance across repeated runs and that the reported improvements are not dependent on a single training instance.

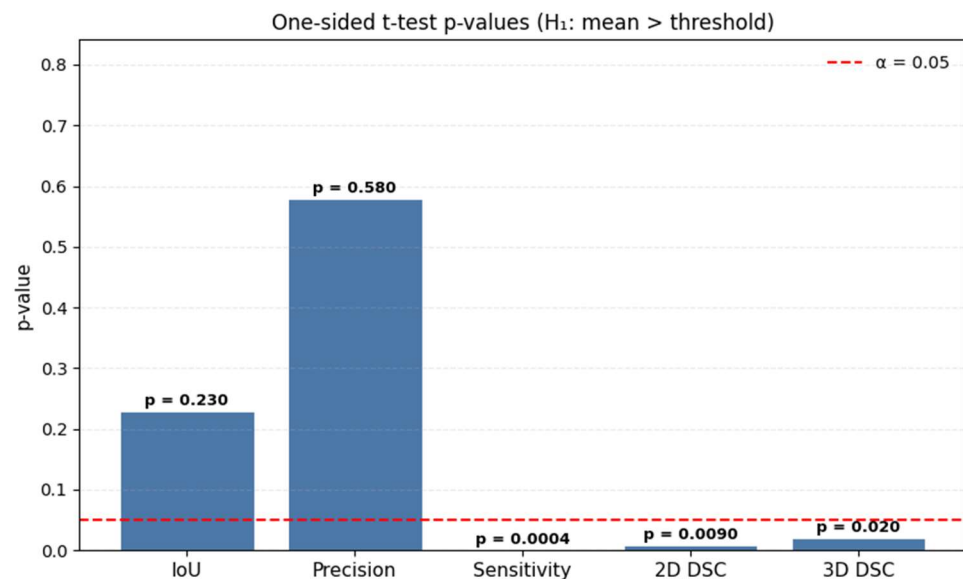


Figure 4. One-sided paired *t*-test *p*-values for each metric comparing the proposed model against U-Net + InceptionResNetV2 [24]. The dashed red line indicates the significance threshold ($\alpha = 0.05$).

Table 7. Performance stability across five repeated runs (mean $\pm$ SD).

Metric	Run1	Run2	Run3	Run4	Run5	Mean $\pm$ SD
IoU	0.45	0.43	0.51	0.44	0.53	0.472 $\pm$ 0.045
Precision	0.73	0.70	0.76	0.64	0.73	0.712 $\pm$ 0.045
Sensitivity	0.60	0.55	0.62	0.59	0.65	0.602 $\pm$ 0.037
2D DSC	0.62	0.66	0.68	0.59	0.69	0.648 $\pm$ 0.042
3D DSC	0.55	0.61	0.62	0.55	0.63	0.592 $\pm$ 0.039

4. Discussions

In this study, we propose a hybrid deep learning model that fuses the entire image and its patches for BML segmentation. The motivation behind this hybrid design stems from the limitations of using either whole-image input or image patches. While the U-Net on full images captures global context and anatomical structure, it often struggles to segment subtle or small lesions due to low contrast and texture similarity with surrounding regions. On the other hand, using only cropped patches can capture local details but lose global spatial context. To address these deficiencies, our model integrates whole-image features with patch-level features, enabling it to leverage both global context and fine-grained local information simultaneously. This fully automated approach eliminates the need to explicitly crop regions containing BML, as the hybrid model processes the entire image along with its patches for segmentation.

We have compared our hybrid deep learning model with other deep learning models for image segmentation. These methods include the U-Net model [17], U-Net + EfficientNetB7 [32], Sharp U-Net [33], U-Net + ResNet-34 [34], and U-Net + InceptionResNetV2 [24]. As shown in Table 1, with a batch size of 32, the proposed hybrid model outperformed other models across almost all evaluated metrics. There are two potential reasons for this improvement. First, the patch stream explicitly encourages the network to focus on smaller subregions, making it more sensitive to fine-grained lesion details such as subtle texture and contrast variations that might be overlooked by the image branch. Second, the image stream provides broader contextual information that counterbalances the patch stream, which might otherwise focus too narrowly on small, noisy regions.

We also conducted error analysis, where we examined the 2D Dice score for each individual image in the test set, the original image, and its ground truth mask to understand the potential reasons for images with low 2D Dice score. Out of all images in the test set, about 73% of the images had a Dice score above or equal to 0.5, and about 27% of the images had a Dice score below 0.5 (9 of them produced a Dice score of 0). For the 9 images with a Dice score of 0, the lesions in the ground truth mask were extremely difficult to distinguish, even with the naked eye, which potentially explains why the model failed on those 9 images. For images with Dice scores below 0.5, the surrounding regions of lesions showed contrast and texture patterns similar to those of lesions, possibly leading to misclassification of pixels and inaccurate segmentation. To provide a more structured characterization of failure modes, we identified two primary categories of errors, with representative examples shown in Figure 5. The first category is false-positive prediction in non-lesion marrow regions. As illustrated in the first row of Figure 5 (Dice = 0.30), the model predicted lesion regions in areas where no BML was present in the ground truth. In this case, the surrounding bone marrow exhibited contrast and texture patterns similar to those of actual lesions, possibly causing the model to misclassify healthy tissue as BML. This type of error is particularly common in regions adjacent to the articular surface, where normal bone marrow signal heterogeneity can mimic lesion appearance. The second category is under-segmentation of faint or small lesions. As shown in the second row of Figure 5 (Dice = 0.64), the model captured the general location of the lesion but failed to fully delineate its extent, resulting in a predicted mask that was noticeably smaller than the ground truth. This occurs when lesion boundaries are indistinct due to low contrast with surrounding tissue, possibly causing the model to conservatively segment only the most prominent portion of the lesion while missing its peripheral regions.

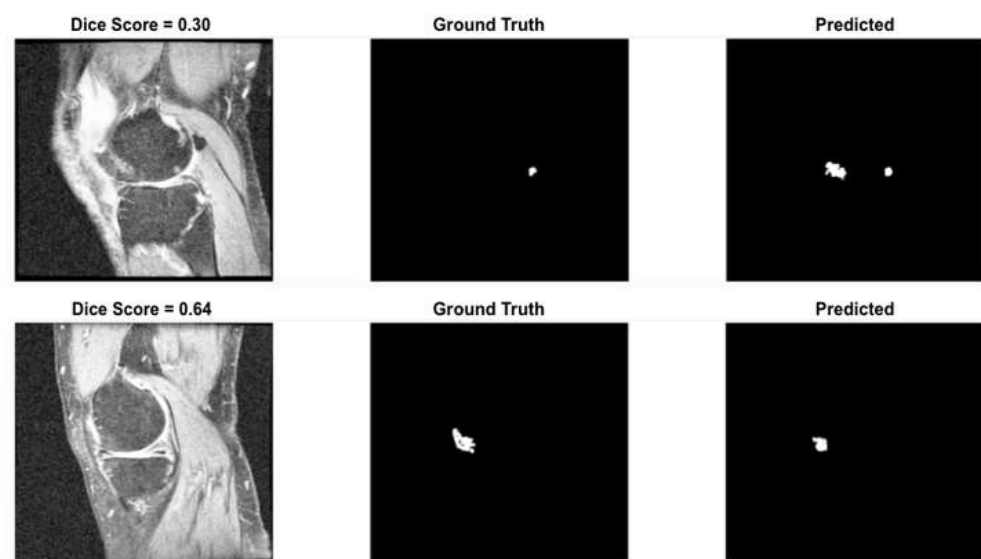


Figure 5. Sample MRI scan, their BML ground truth mask, and their BML segmentation result generated by the proposed hybrid model, showing failure cases.

In the proposed hybrid model, using non-overlapping patches with a stride equal to the patch size (112 pixels) was preferred over overlapping patches for two main reasons. First, overlapping patches would substantially increase the number of patches per image, thereby increasing the computational cost of the patch encoder branch, which already accounts for most of the additional overhead in the hybrid model, as shown in Table 5. Second, the fusion of patch-level features with the global image stream in the proposed architecture helps mitigate any boundary discontinuity issue, as the image branch provides

continuous spatial context across the entire image, compensating for any local information loss at patch boundaries.

With a batch size of 32, we evaluated 2D and 3D DSC, as well as IoU, Pearson correlation coefficient, precision, and sensitivity. Like higher 2D and 3D DSC in our model, the IoU of the proposed hybrid model outperformed those of other models compared in this study, further demonstrating that the proposed model was more effective at capturing object boundaries. Our model also achieved the highest Pearson correlation coefficient among all the models compared, indicating a strong volume-level correlation between the segmentation result and the ground truth mask. As for precision, our model outperformed all the other models but U-Net + InceptionResNetV2 in [24]. One potential reason is that the proposed hybrid model might have captured local noise patterns from the patch branch, leading to false-positive predictions. Moreover, our model achieved the highest sensitivity score, indicating that it is less likely to miss parts of a lesion and more able to correctly identify pixels within the lesion.

A result of DSC of 0.68 and Pearson correlation coefficient of 0.98 between BML segmentation volume and ground truth volume on the same testing dataset, reported in [24] using the U-Net + InceptionResNetV2 model, was achieved using a batch size of 64. With an effective batch size of 64, as shown in Table 2, our hybrid model has a slightly lower 2D Dice score, indicating it is slightly less accurate on a slice-wise overlap basis. However, it has a notable improvement in 3D Dice, compared with the model in [24]. This indicates that, although slice-level agreement is not uniformly improved, the proposed model more effectively captures the overall lesion burden across the full 3D volume. Since 3D Dice is computed by stacking predicted 2D masks into a full 3D volume per patient and evaluating overlaps across the entire volume, it better reflects volumetric coherence rather than local per-slice overlap. In addition, the identical $r = 0.98$ for both methods further suggests that our model also preserves volume trends well, whereas its higher 3D DSC indicates improved volumetric agreement with the ground truth. Moreover, the proposed hybrid model achieved a slightly higher IoU and greater sensitivity than the model in [24], indicating superior spatial overlap and enhanced detection of lesion regions. However, precision is marginally lower, which reflects a trade-off in which the model favors higher sensitivity (reduced false negatives) at the cost of lower precision, leading to a slight increase in false-positive predictions. Overall, these results suggest that the proposed hybrid model prioritizes more consistent volumetric lesion characterization, even when slice-wise overlap metrics are not uniformly superior.

The improved 3D Dice scores indicate better voxel-level agreement with manual annotations, while the strong Pearson correlation between predicted and manual BML volumes suggests that the proposed model can produce volumetric measurements that are highly consistent with expert annotations. This is clinically important because BML volume is considered a meaningful imaging biomarker in osteoarthritis and has been associated with pain severity, disease progression, and cartilage degeneration [14,35]. More accurate volumetric segmentation may therefore facilitate longitudinal monitoring of osteoarthritis progression, assessment of treatment response, and biomarker-driven patient stratification in clinical research and therapeutic trials [14,35].

This study has several limitations. First, all experiments were conducted exclusively on the OAI dataset without external validation on independent cohorts. Although the OAI dataset is a large, high-quality, and widely used resource, models trained and evaluated on a single dataset may learn dataset-specific imaging characteristics, scanner-dependent patterns, or annotation conventions that may not generalize to other clinical settings. Therefore, the current results primarily demonstrate the effectiveness of the proposed framework within the OAI cohort, and further validation on external multi-center datasets is needed

to confirm its generalizability. Second, because the proposed framework combines both patch-based and whole-image representations, it requires greater memory consumption than architectures that rely solely on either image-based or patch-based inputs. Third, it should be noted that while CLAHE and Gaussian filtering improved visual contrast and training stability in our experiments, we did not perform a dedicated ablation study to quantify the individual contribution of each preprocessing component. Therefore, their exact impact on model performance cannot be fully isolated in the current study. Another limitation of this study is that the model was trained and evaluated only on slices containing visible BMLs, rather than on full volumetric scans that include lesion-free slices. The last potential limitation is the inherent class imbalance between lesion and non-lesion pixels. Although the use of 3D volumetric context helps mitigate this issue, it may still influence performance in cases with very small or low-contrast lesions.

5. Conclusions

In this study, we introduce a hybrid deep learning model that combines images and patches of each image for simultaneous image and patch feature extraction using U-Net as the fundamental architecture and InceptionResNetV2 as the encoder. The proposed approach, with a physical batch size of 32, achieved a 2D DSC of 0.68, a 3D DSC of 0.62, an IoU of 0.51, a precision of 0.76, and a sensitivity of 0.62, with an r of 0.85 between manual and automatic volume measurements. Using an effective batch size of 64 via gradient accumulation, the model achieved a 2D DSC of 0.63, a 3D DSC of 0.65, an IoU of 0.48, a precision of 0.75, and a sensitivity of 0.6, with an r of 0.98 between manual and automatic volume measurements. This approach is fully automated and requires no human involvement. Compared with existing deep learning models for image segmentation, the proposed model achieved better overall performance across almost all evaluated metrics at a batch size of 32. Using an effective batch size of 64, the proposed model demonstrated improved 3D DSC, IoU, and sensitivity compared with the best-performing baseline model proposed in [24], while r remained high at 0.98.

Future work will focus on improving computational efficiency and generalizability by optimizing the model to reduce memory usage using cropped regions where BMLs are most likely to occur, as well as by training and validating the framework on larger multi-center datasets that include additional bone types and diverse imaging acquisition settings. In addition, future studies will evaluate the model's performance across different BML sizes, particularly for smaller lesions. Another future direction is to replace the ImageNet pre-trained InceptionResNetV2 with a domain-specific pretrained encoder, which may further improve the model's overall segmentation accuracy. Finally, exploring overlapping patch strategies with reduced computational overhead, such as using shared-weight lightweight encoders, remains a direction for future investigation.

Author Contributions: Conceptualization, L.L., J.S. and M.Z.; methodology, L.L. and D.S.P.; software, D.S.P. and H.T.; validation, D.S.P. and H.T.; formal analysis, D.S.P. and L.L.; data curation, J.B.D. and T.M.; writing—original draft preparation, D.S.P. and L.L.; writing—review and editing, L.L., J.S., M.Z. and J.B.D.; supervision, L.L., J.S. and M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Authors Jeffrey B. Driban and Timothy McAlindon's contributions were supported in part by NIH Grant R24 AR085006.

Data Availability Statement: No new data was created. Data were pulled from public database OAI (<https://nda.nih.gov/oai>, accessed on 30 January 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Osteoarthritis. Available online: <https://www.cdc.gov/arthritis/osteoarthritis/> (accessed on 30 January 2026).
2. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **2020**, *396*, 1204–1222. [\[PubMed\]](#)
3. Martig, S.; Boisclair, J.; Konar, M.; Spreng, D.; Lang, J. MRI characteristics and histology of bone marrow lesions in dogs with experimentally induced osteoarthritis. *Vet. Radiol. Ultrasound* **2007**, *48*, 105–112. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Leydet-Quilici, H.; Le Corroller, T.; Bouvier, C.; Giorgi, R.; Argenson, J.N.; Champsaur, P.; Pham, T.; de Paula, A.M.; Lafforgue, P. Advanced hip osteoarthritis: Magnetic resonance imaging aspects and histopathology correlations. *Osteoarthr. Cartil.* **2010**, *11*, 1429–1435. [\[CrossRef\]](#)
5. Saadat, E.; Jobke, B.; Chu, B.; Lu, Y.; Cheng, J.; Li, X.; Ries, M.D.; Majumdar, S.; Link, T.M. Diagnostic performance of in vivo 3-T MRI for articular cartilage abnormalities in human osteoarthritic knees using histology as standard of reference. *Eur. Radiol.* **2008**, *18*, 2292–2302. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Zanetti, M.; Bruder, E.; Romero, J.; Hodler, J. Bone marrow edema pattern in osteoarthritic knees: Correlation between MR imaging and histologic findings. *Radiology* **2000**, *215*, 835–840. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Gabrielle Bergman, A.; Willén, H.K.; Lindstrand, A.L.; Pettersson, H.T. Osteoarthritis of the knee: Correlation of subchondral MR signal abnormalities with histopathologic and radiographic features. *Skelet. Radiol.* **1994**, *23*, 445–448. [\[CrossRef\]](#)
8. Hunter, D.J.; Gerstenfeld, L.; Bishop, G.; Davis, A.D.; Mason, Z.D.; Einhorn, T.A.; Maciewicz, R.A.; Newham, P.; Foster, M.; Jackson, S.; et al. Bone marrow lesions from osteoarthritis knees are characterized by sclerotic bone that is less well mineralized. *Arthritis Res. Ther.* **2009**, *11*, R11. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Driban, J.B.; Tassinari, A.; Lo, G.H.; Price, L.L.; Schneider, E.; Lynch, J.A.; Eaton, C.B.; McAlindon, T.E.; OAI Investigators Group. Bone marrow lesions are associated with altered trabecular morphometry. *Osteoarthr. Cartil.* **2012**, *20*, 1519–1526. [\[CrossRef\]](#)
10. Felson, D.T.; Chaisson, C.E.; Hill, C.L.; Totterman, S.M.; Gale, M.E.; Skinner, K.M.; Kazis, L.; Gale, D.R. The association of bone marrow lesions with pain in knee osteoarthritis. *Ann. Intern. Med.* **2001**, *134*, 541–549. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Felson, D.T.; McLaughlin, S.; Goggins, J.; LaValley, M.P.; Gale, M.E.; Totterman, S.; Li, W.; Hill, C.; Gale, D. Bone marrow edema and its relation to progression of knee osteoarthritis. *Ann. Intern. Med.* **2003**, *139*, 330–336. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Hunter, D.J.; Zhang, Y.; Niu, J.; Goggins, J.; Amin, S.; LaValley, M.P.; Guermazi, A.; Genant, H.; Gale, D.; Felson, D.T. Increase in bone marrow lesions associated with cartilage loss: A longitudinal magnetic resonance imaging study of knee osteoarthritis. *Arthritis Rheum.* **2006**, *54*, 1529–1535. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Hunter, D.J.; Guermazi, A.; Lo, G.H.; Grainger, A.J.; Conaghan, P.G.; Boudreau, R.M.; Roemer, F.W. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthr. Cartil.* **2011**, *19*, 990–1002. Erratum in *Osteoarthr. Cartil.* **2011**, *19*, 1168. [\[CrossRef\]](#)
14. Driban, J.B.; Price, L.L.; Lo, G.H.; Pang, J.; Hunter, D.J.; Miller, E.; Ward, R.J.; Eaton, C.B.; Lynch, J.A.; McAlindon, T.E. Evaluation of bone marrow lesion volume as a knee osteoarthritis biomarker-longitudinal relationships with pain and structural changes: Data from the osteoarthritis initiative. *Arthritis Res. Ther.* **2013**, *15*, R112. [\[PubMed\]](#)
15. Dijkstra, A.J.; Anbeek, P.; Yang, K.A.; Vincken, K.L.; Viergever, M.A.; Castelein, R.M.; Saris, D.B.F. Validation of a novel semiautomated segmentation method for MRI detection of cartilage-related bone marrow lesions. *Cartilage* **2010**, *1*, 328–334. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Pang, J.; Driban, J.B.; Destenaves, G.; Miller, E.; Lo, G.H.; Ward, R.J.; Price, L.L.; Lynch, J.A.; Eaton, C.B.; Eckstein, F.; et al. Quantification of bone marrow lesion volume and volume change using semi-automated segmentation: Data from the osteoarthritis initiative. *BMC Musculoskelet. Disord.* **2013**, *14*, 3. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Preiswerk, F.; Sury, M.S.; Wortman, J.R.; Neumann, G.; Wells, W.; Duryea, J. Fast quantitative bone marrow lesion measurement on knee MRI for the assessment of osteoarthritis. *Osteoarthr. Cartil. Open* **2022**, *4*, 100234. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Dodin, P.; Abram, F.; Pelletier, J.P.; Martel-Pelletier, J. A fully automated system for quantification of knee bone marrow lesions using MRI and the osteoarthritis initiative cohort. *J. Biomed. Graph. Comput.* **2013**, *3*, 51–65.
19. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yüille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
20. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
21. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv* **2023**, arXiv:2312.00752.
22. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
23. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer International Publishing: Cham, Switzerland, 2015.
24. Ponnusamy, R.; Zhang, M.; Wang, Y.; Sun, X.; Chowdhury, M.; Driban, J.B.; McAlindon, T.; Shan, J. Automatic segmentation of bone marrow lesions on MRI using a deep learning method. *Bioengineering* **2024**, *11*, 374. [\[CrossRef\]](#) [\[PubMed\]](#)

25. Stirling, C.E.; Neeteson, N.J.; Walker, R.E.; Boyd, S.K. Deep learning-based automated detection and segmentation of bone and traumatic bone marrow lesions from MRI following an acute ACL tear. *Comput. Biol. Med.* **2024**, *178*, 108791. [[CrossRef](#)] [[PubMed](#)]
26. Xie, X.; Zhang, W.; Pan, X.; Xie, L.; Shao, F.; Zhao, W.; An, J. Canet: Context aware network with dual-stream pyramid for medical image segmentation. *Biomed. Signal Process. Control* **2023**, *81*, 104437.
27. Qiu, Z.; Hu, Y.; Chen, X.; Zeng, D.; Hu, Q.; Liu, J. Rethinking dual-stream super-resolution semantic learning in medical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *46*, 451–464. [[PubMed](#)]
28. OAI. The Osteoarthritis Initiative. Available online: <https://nda.nih.gov/oai> (accessed on 30 January 2026).
29. Driban, J.B.; Price, L.L.; LaValley, M.P.; Lo, G.H.; Zhang, M.; Harkey, M.S.; Canavatchel, A.; McAlindon, T.E. Novel framework for measuring whole knee osteoarthritis progression using magnetic resonance imaging. *Arthritis Care Res.* **2022**, *74*, 799–808. [[CrossRef](#)]
30. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; Romeny, B.H.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* **1987**, *39*, 355–368. [[CrossRef](#)]
31. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
32. Lin, S.Y.; Lin, C.L. Brain tumor segmentation using U-Net in conjunction with EfficientNet. *PeerJ Comput. Sci.* **2024**, *10*, E1754. [[CrossRef](#)] [[PubMed](#)]
33. Zunair, H.; Hamza, A.B. Sharp U-Net: Depthwise convolutional network for biomedical image segmentation. *Comput. Biol. Med.* **2021**, *136*, 104699. [[CrossRef](#)] [[PubMed](#)]
34. Abedalla, A.; Abdullah, M.; Al-Ayyoub, M.; Benkhelifa, E. 2st-unet: 2-stage training model using u-net for pneumothorax segmentation in chest x-rays. In *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020*; IEEE: Piscataway, NJ, USA, 2020.
35. Driban, J.B.; Lo, G.H.; Lee, J.Y.; Ward, R.J.; Miller, E.; Pang, J.; Price, L.L.; McAlindon, T.E. Quantitative bone marrow lesion size in osteoarthritic knees correlates with cartilage damage and predicts longitudinal cartilage loss. *BMC Musculoskelet. Disord.* **2011**, *12*, 217. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.