Current Research in MRI

Original Article

Deep Learning–Based Diagnosis of Chondromalacia Patella from Magnetic Resonance Images

Okan Demirtaş[1], Sümeyra Kuş[1], Semih Demirel[2], Habib Eser Akkaya[3], Oktay Yıldız[1], Ömer Kazcı[4]

[1]Department of Computer Engineering, Gazi University Faculty of Engineering, Ankara, Türkiye
[2]Department of Information Systems, Gazi University Faculty of Engineering, Ankara, Türkiye
[3]Department of Radiology, Ankara Training and Research Hospital, Ankara, Türkiye
[4]Department of Radiology, Presidential Health Center, Ankara, Türkiye

Cite this article as: Demirtaş O, Kuş S, Demirel S, Eser Akkaya H, Yıldız O, Kazcı Ö. Deep learning–based diagnosis of chondromalacia patella from magnetic resonance images. *Current Research in MRI*, 2025;4(1):12-21.

Corresponding author: Ömer Kazcı, e-mail: omerkazci1990@gmail.com

Received: April 23, 2025 **Accepted:** May 13, 2025 **Publication Date:** August 22, 2025

DOI:10.5152/CurrResMRI.2025.25118

Abstract

Objective: Chondromalacia patella is a degenerative disease of the articular cartilage in the patellofemoral joint, often leading to anterior knee pain and functional impairment. Accurate diagnosis is crucial to prevent disease progression and enable early intervention. Magnetic resonance imaging (MRI) is considered the gold standard for assessing cartilage integrity; however, manual interpretation is subjective and requires expert radiologists. In this study, a deep learning–based approach is proposed for the automatic classification of chondromalacia patella from MRI scans. A novel and domain-specific dataset was constructed for this study, consisting of labeled knee MRI scans obtained from routine clinical examinations.

Methods: This dataset is one of the first dedicated to chondromalacia patella classification, enhancing its originality and contribution to the field. The dataset was carefully curated and labeled by expert radiologists to ensure high diagnostic accuracy. It includes a diverse range of MRI images representing different grades of chondromalacia patella, making it a valuable resource for deep learning applications in musculoskeletal imaging.

Results: Three different models were evaluated: a custom-designed convolutional neural network (CNN) model and 2 pretrained architectures, VGG16 and ResNet50. Each model was trained and tested independently. The results indicate that both the custom CNN model and the VGG16 model achieved an accuracy of 92.66%, while the ResNet50 model performed slightly lower with an accuracy of 90.83%. Despite recent advancements in deep learning for musculoskeletal imaging, studies focusing specifically on chondromalacia patella remain limited.

Conclusion: This study addresses this gap by presenting an automated and objective diagnostic method that significantly enhances diagnostic accuracy. The results demonstrate the potential of deep learning methods in providing objective and highly accurate diagnoses from MRI scans, reducing the subjectivity associated with manual evaluation and offering a reliable decision-support tool for clinicians.

Keywords: Chondromalacia patella, convolutional neural networks, deep learning, MRI classification, transfer learning

INTRODUCTION

Chondromalacia patella is a degenerative condition affecting the articular cartilage of the patellofemoral joint, one of the most critical structures of the knee. This condition frequently manifests as anterior knee pain and functional impairment, particularly in young and physically active individuals, often occurring during or after sports activities.[1] According to the PearlDriver database, between 2007 and 2011, a total of 30 108 510 patients were recorded in the United States, among whom 2 188 753 individuals were diagnosed with patellofemoral pain (PFP). Of these cases, 437 711 were identified as chondromalacia patella, accounting for 20% of all PFP patients. Furthermore, the incidence of chondromalacia patella was 60% in females and 40% in males.[2] Chondromalacia patella, one of the potential causes of Patellofemoral Pain Syndrome, constitutes approximately 10%-25% of all visits to physical therapy clinics.[3] The structural integrity and friction-reducing properties of cartilage are essential for joint health; however, its limited self-repair capacity underscores the progressive nature of chondromalacia patella. Therefore, accurate and early diagnosis is crucial for preventing the deterioration of cartilage and determining appropriate treatment strategies.[4]

Traditional diagnostic methods include clinical examination, physical tests, and imaging techniques. Among these, magnetic resonance imaging (MRI) is considered the gold standard for evaluating chondromalacia patella due to its high-contrast and detailed visualization of cartilage structures.[5] Magnetic resonance imaging can reveal both morphological and subtle compositional changes in the cartilage. Despite these advantages, the interpretation of MRI results is a subjective process that requires substantial expertise. Radiologists' assessments can vary, and subtle early-stage cartilage lesions might be overlooked, leading to diagnostic delays.[6] These limitations highlight the need for more objective and reproducible diagnostic methods to support clinical decision-making.

A novel and domain-specific dataset was constructed for this study, consisting of labeled knee MRI scans obtained from routine clinical examinations. Given that there is no publicly available dataset specifically for chondromalacia patella, this study provides an important contribution by introducing an original dataset that can support future research. The dataset was carefully curated and labeled by expert radiologists to ensure high diagnostic accuracy. It includes a diverse range of MRI images representing different grades of chondromalacia patella, making it a valuable resource for deep learning applications in musculoskeletal imaging.

Recent advancements in artificial intelligence and deep learning have facilitated significant progress in the development of automated diagnostic systems for medical imaging. Deep learning models, particularly convolutional neural networks (CNNs), are capable of detecting microlevel tissue abnormalities and pathological changes that may not be easily discernible by the human eye by leveraging large-scale datasets for training.[7] In the assessment of patellar cartilage, these algorithms have demonstrated high accuracy in detecting early-stage lesions and relibly classifying disease severity.[8] Consequently, deep learning–based approaches hold substantial potential for expediting diagnostic processes and enhancing objective clinical evaluations.

Deep learning models have increasingly been integrated with MRI data for the diagnosis of various musculoskeletal conditions. Training these models on large patient cohorts has demonstrated their ability to accurately detect different pathological stages of cartilage degeneration. Additionally, their high sensitivity and specificity in identifying microscopic early-stage tissue alterations provide a significant advantage in guiding timely therapeutic interventions.[9,10] As a result, the development of automated and objective diagnostic systems offers promising advancements for both clinical practice and medical research by improving disease monitoring and treatment planning.

However, despite these technological advancements, the application of deep learning specifically to chondromalacia patella diagnosis remains underexplored. Most existing research on knee MRI has focused on broader pathologies or related cartilage conditions—such as fully automated detection of general cartilage lesions in the knee[11] or the classification of knee osteoarthritis severity from MRI.[12] Comparatively little attention has been given to developing automated methods for diagnosing patellofemoral cartilage lesions like chondromalacia patella. This gap in the literature highlights the need for a targeted investigation into whether deep learning can be effectively applied to detect chondromalacia patella from routine MRI data. Addressing this need, the present study proposes a deep learning–based approach for the automatic classification of chondromalacia patella using knee MRI scans. A custom CNN architecture was developed and trained from scratch on annotated MRI images, and in addition, 2 state-of-the-art pretrained CNN models (VGG16 and ResNet50) were fine-tuned on the same dataset for transfer learning. Each model was trained and validated independently on a large set of labeled patellofemoral cartilage images (classified as cmp for chondromalacia and normal for healthy cartilage), and their diagnostic performance in distinguishing chondromalacia patella was systematically evaluated. By implementing and comparing both a novel, domain-specific CNN and established deep learning architectures, this study provides a comprehensive assessment of deep learning's applicability to chondromalacia patella detection.

The experimental findings demonstrate the efficacy of the proposed approach. Both the custom-designed CNN and the transfer learning–based VGG16 model achieved high classification accuracy (~ 92.7% on the test set), while the ResNet50 model attained a slightly lower accuracy (~ 90.8%). This level of performance is notable, as it approaches the reliability of expert human interpretation and significantly exceeds chance, indicating that subtle cartilage abnormalities associated with chondromalacia patella can be recognized consistently by deep learning models.

These results confirm the considerable potential of deep learning in providing objective and highly accurate diagnoses from MRI scans, directly addressing the subjectivity and variability inherent in manual MRI evaluation. The ability of the models to detect chondromalacia patella with such high accuracy underscores how artificial intelligence can augment clinical practice—particularly as a decision-support tool that identifies early-stage cartilage lesions for further review, thereby facilitating earlier interventions.

In the context of the literature, this work represents a significant advancement in automated medical diagnostics for musculoskeletal imaging. To the authors' knowledge, it is one of the first studies to focus explicitly on chondromalacia patella classification using deep learning, thereby filling an important gap in the domain of knee cartilage assessment.

Overall, the study not only introduces an effective AI-driven solution for a specific diagnostic challenge but also advances the current state of knowledge by demonstrating that deep learning can be successfully harnessed for patellofemoral cartilage evaluation. This contribution lays the groundwork for future research and development of AI-assisted diagnostic tools in orthopedics, ultimately aiming to improve diagnostic accuracy and patient outcomes for knee joint disorders.

MAIN POINTS

- Understand the role of deep learning in diagnosing chondromalacia patella from magnetic resonance imaging (MRI) scans.
- Recognize the importance of creating a dedicated and expertly labeled MRI dataset for accurate model training.
- Compare the diagnostic performance of different convolutional neural network architectures, including custom and pretrained models.
- Evaluate the potential of artificial intelligence–based systems to support clinical decision-making and reduce diagnostic subjectivity.

MATERIAL AND METHODS

The knee MRI data used in this study were obtained from routine clinical examinations at Ankara Training and Research Hospital. All procedures were approved by the institutional Ehics Committee of Gazi University (Approval No: 1117365; Date: 13.12.2024). As this was a retrospective study, the imaging data were obtained from an existing database; therefore, informed consent was not required. Imaging was performed using a 1.5 Tesla MRI scanner (Siemens) with T2-weighted and Proton Density sequences at a slice thickness of 3 mm and in-plane matrix size of 512 × 512 (field of view 16 cm). Images in sagittal, coronal, and axial planes focusing on the patellofemoral joint were included to capture the cartilage from multiple views. The dataset was categorized into 2 classes: chondromalacia patella (CMP) and normal (Figure 1). In total, 1073 MRI slice images were collected, with 648 labeled as CMP and 425 as normal. For model development and evaluation, the data were stratified and split into training, validation, and

test subsets, as illustrated in Figure 2. The training set consisted of 858 images (518 CMP, 340 normal), the validation set contained 106 images (64 CMP, 42 normal), and the test set included 109 images (66 CMP, 43 normal). This stratified split ensured a balanced representation of both classes across all subsets, facilitating robust model training and evaluation. All images were initially assessed by a specialist musculoskeletal radiologist, and a second independent radiologist reviewed the labels to ensure diagnostic accuracy. Low-quality or artefactual scans were excluded prior to analysis. Before input to the deep learning models, all MR images underwent standard preprocessing. Each image was resized to 224 × 224 pixels to fit the input requirements of the CNN models. Intensity values were normalized to the range [0, 1] by linear scaling of pixel values, which facilitates network training and convergence. To prepare for training, the training images were randomly shuffled (to eliminate any unintended ordering biases) and then batched. A mini-batch size of 32 was used, balancing convergence stability with computational efficiency. This batching, along with an optimized data pipeline (including data shuffling and prefetching), helped improve training throughput and prevented the model from learning spurious sequential patterns. No additional data augmentation was applied, so as to evaluate the models on the original image distributions.

Experimental Study

We developed a custom CNN architecture for the binary classification of chondromalacia patella. The network consisted of 4 convolutional blocks followed by fully connected layers (Figure 3). In the first convolutional layer, 32 filters of size 3 × 3 were applied to extract low-level visual features from the input MR images. A Rectified Linear Unit (ReLU) activation was used after each convolution to introduce nonlinearity, defined as (1):

$$f(x) = \max(0, x) \tag{1}$$

Each convolutional block also included a 2 × 2 max-pooling layer that down-sampled feature maps (reducing spatial dimensions by a factor of 2) to highlight the most salient features and reduce computational complexity. As the network depth increased, the number of filters in convolutional layers was doubled (32 → 64 → 128 → 256) in successive blocks to learn more complex and abstract feature representations.

To mitigate overfitting, dropout regularization[13] was incorporated after certain layers. For example, after the first convolutional block, a dropout layer with a rate of 0.2 was applied—meaning each neuron in that layer was retained with a probability of 0.8 and dropped with a probability of 0.2 during training. Similar dropout layers with increasing drop rates (0.3, 0.4, and 0.5 in subsequent blocks) were used as the network grew deeper, progressively randomizing neuron activations to encourage generalization. The output feature maps from the final convolutional layer were flattened into a one-dimensional vector, which was then fed into a dense fully connected layer of 256 neurons. This dense layer also used ReLU activation to learn high-level combinations of the extracted features. Another dropout layer (50% rate) followed the dense layer to further reduce overfitting.

Finally, the network's output layer was a single neuron with a sigmoid activation function, producing a probability between 0 and 1 to indicate the likelihood of the input image being classified as chondromalacia. The sigmoid function is defined as (2):

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

where x represents the input to the activation function. The sigmoid function squashes the input into a range between 0 and 1, making it particularly suitable for binary classification problems such as chondromalacia detection. The function ensures that outputs can be interpreted as probabilities, which facilitates decision-making in the classification process.

In summary, the custom CNN architecture was designed to progressively extract features through convolution and pooling while using ReLU activations and dropout regularization to ensure robust learning and prevent overfitting.

In the second phase of the study, transfer learning was employed to evaluate the performance of pretrained VGG16[14] and ResNet50[15] models. Transfer learning is based on the principle of utilizing pretrained models with frozen lower-layer weights while retraining only the upper layers for a specific target dataset, as illustrated in Figures 4 and 5.

Depending on the model selection, VGG16 or ResNet50 was loaded with pretrained weights from the ImageNet dataset, while the fully connected layers were excluded (include top=False). This approach preserved the feature extraction capabilities of the original model, ensuring that only the classification layer was retrained for the target dataset.

Initially, all layers of the base models were frozen (trainable = False). However, to enhance generalization, in the VGG16 model, the first 10

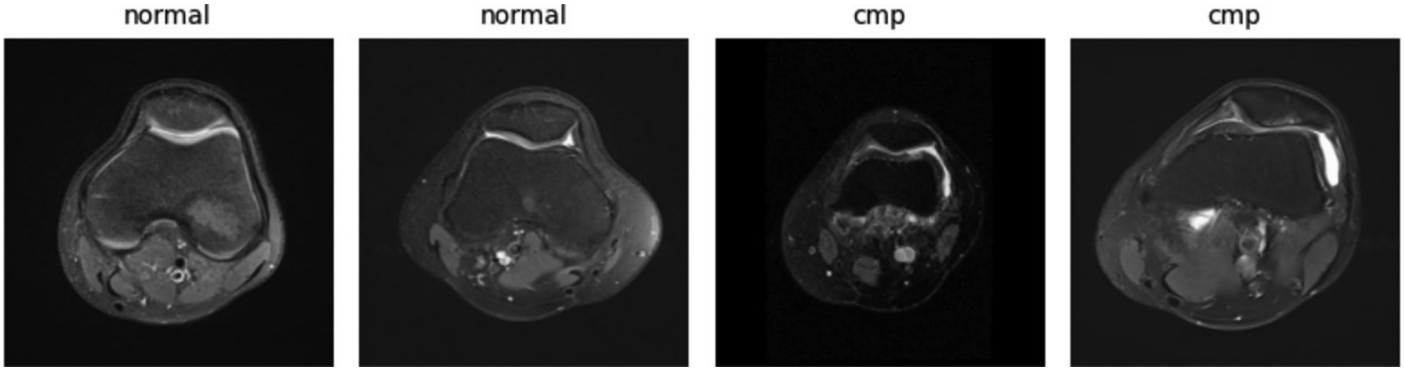

Figure 1. Examples of the dataset.

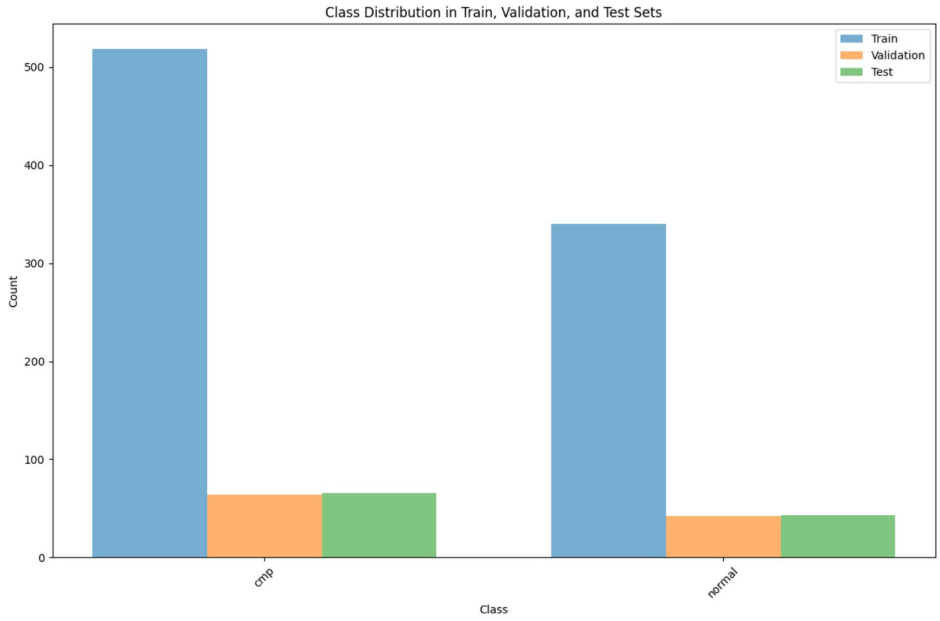

Figure 2. Distribution of the dataset across training, validation, and test sets for chondromalacia patella and normal classes.

layers were frozen while the remaining layers were made trainable, whereas in the ResNet50 model, the last 20 layers were fine-tuned to adapt the model to the target dataset.

The feature maps extracted from the base model were processed through a Global Average Pooling (GAP) layer. The GAP layer computes the mean activation of each feature map, reducing the number of trainable parameters and mitigating the risk of overfitting.

Following the GAP layer, a fully connected (dense) layer with 128 neurons and a ReLU activation function was incorporated. To further prevent overfitting, a 50% dropout layer was applied. Finally, a single neuron with a sigmoid activation function was used in the output layer for binary classification.

All models (the custom CNN and the 2 transfer learning models) were trained using a consistent procedure to ensure a fair comparison. The network weights were optimized using the Adam optimization algorithm, which is an adaptive stochastic gradient descent method. Adam maintains moving averages of the gradient and its square to adapt the learning rate for each parameter. At each training step t, the first moment m_t (mean of gradients) and second moment v_t (uncentered variance) are updated as (3-4):

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{3}$$

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{4}$$

where g_t is the gradient of the loss with respect to the model parameters at step t, and β_1 and β_2 are the decay rates for the moving averages (typically $\beta_1 = 0.9$, $\beta_2 = 0.999$). The learning rate was initially set to 1×10^{-4} for all experiments, as this small rate provided stable learning for complex models. Model parameters were updated using backpropagation to minimize the classification loss.

The binary cross-entropy loss function is employed, which is suitable for binary classification problems. This loss function measures the discrepancy between predicted probabilities and true labels, and is defined as (5):

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(y_i\right) + \left(1-y_i\right)\log\left(1-y_i\right)\right] \tag{5}$$

where:

- N is the total number of training samples,
- $y_i \in \{0,1\}$ is the true label of sample i (with 1 representing chondromalacia patella and 0 representing normal cartilage),
- $\hat{y}_i$ is the predicted probability that sample i belongs to the positive class (CMP).

This loss function penalizes confident misclassifications heavily, guiding the optimizer towards improved predictions.

During training, dropout layers were activated as described in the model architecture. This means that in each iteration, different subsets of neurons were randomly dropped, helping the Adam optimizer generalize the model. Training was carried out in mini-batches of 32 images, striking a balance between computational efficiency and gradient estimation stability. The models were trained for a maximum of 100 epochs or until convergence.

To prevent overfitting, an early stopping strategy was implemented based on validation performance. Specifically, if the validation loss did not improve for a patience period of 10 consecutive epochs, training was halted. The model parameters corresponding to the epoch with the lowest validation loss were then restored as the final model.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 222, 222, 32)      896

 max_pooling2d (MaxPooling2  (None, 111, 111, 32)      0
 D)

 dropout (Dropout)           (None, 111, 111, 32)      0

 conv2d_1 (Conv2D)           (None, 109, 109, 64)      18496

 max_pooling2d_1 (MaxPoolin  (None, 54, 54, 64)        0
 g2D)

 dropout_1 (Dropout)         (None, 54, 54, 64)        0

 conv2d_2 (Conv2D)           (None, 52, 52, 128)       73856

 max_pooling2d_2 (MaxPoolin  (None, 26, 26, 128)       0
 g2D)

 dropout_2 (Dropout)         (None, 26, 26, 128)       0

 conv2d_3 (Conv2D)           (None, 24, 24, 256)       295168

 max_pooling2d_3 (MaxPoolin  (None, 12, 12, 256)       0
 g2D)

 dropout_3 (Dropout)         (None, 12, 12, 256)       0

 flatten (Flatten)           (None, 36864)             0

 dense (Dense)               (None, 256)               9437440

 dropout_4 (Dropout)         (None, 256)               0

 dense_1 (Dense)             (None, 1)                 257

=================================================================
Total params: 9826113 (37.48 MB)
Trainable params: 9826113 (37.48 MB)
Non-trainable params: 0 (0.00 Byte)
```

Figure 3. The structure of the custom neural network.

This ensured that the best-performing model on validation data was retained rather than one that might have started overfitting with continued training.

The entire training process was executed using the TensorFlow/Keras deep learning framework. Model training and validation were performed on an NVIDIA GPU, which accelerated the computation of convolution operations and gradient updates. By the end of training, 3 trained models were obtained (the custom CNN, VGG16-based, and ResNet50-based), each with learned parameters optimized to distinguish chondromalacia patella from normal cartilage in MRI slices.

Results

After training, each model's performance was evaluated on the held-out test set using standard classification metrics. The focus was on metrics that reflect both overall accuracy and the model's ability to correctly identify the positive class (chondromalacia patella).

Accuracy was calculated as the proportion of correctly classified images among all test images (6):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

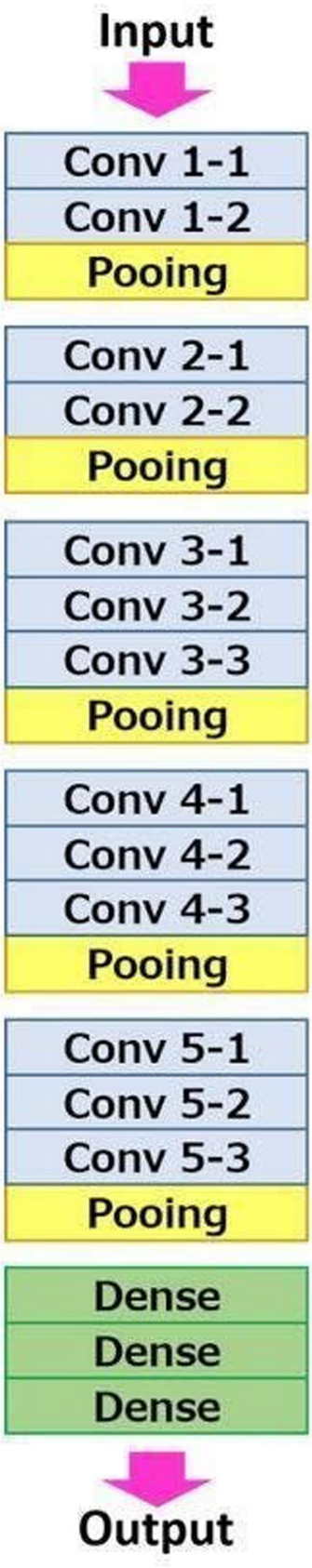

Figure 4. VGG16 model arhitecture.

where:

- TP (True Positives) - Correctly predicted chondromalacia cases.
- TN (True Negatives) - Correctly predicted normal cases.
- FP (False Positives) - Normal cases incorrectly predicted as chondromalacia.
- FN (False Negatives) - Chondromalacia cases incorrectly predicted as normal.

While accuracy provides an overall measure of performance, it can be less informative when dealing with imbalanced datasets. Therefore, additional evaluation metrics were used to assess the model's performance more comprehensively.

Precision, also known as positive predictive value, measures how reliable the model's chondromalacia predictions are. It is defined as (7):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

A higher precision indicates that when the model predicts chondromalacia, it is often correct. High precision minimizes false alarms, which is particularly important in clinical applications where unnecessary interventions should be avoided.

Recall, also known as sensitivity, measures how well the model identifies true chondromalacia cases. It is given by (8):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

A high recall means that the model is effective at detecting chondromalacia and misses few actual cases, which is crucial for ensuring that patients with the condition are correctly identified.

To balance precision and recall, the F1-score is computed, which is the harmonic mean of precision and recall (9):

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

The F1-score is particularly useful when the positive class (chondromalacia) is of primary importance. It provides a single metric that reflects both the reliability of positive predictions and the ability to detect true cases.

To further analyze the model's classification performance, the confusion matrix is computed, which compares the model's predictions with the ground-truth labels on the test set. This structured representation details the number of correctly and incorrectly classified samples across different categories.

In reporting results, the emphasis was on both accuracy and F1-score as indicators of overall effectiveness, while precision and recall provide further insight into the model's ability to correctly identify chondromalacia cases. These metrics help assess the model's capability to detect chondromalacia while minimizing false positives and false negatives, ensuring both high sensitivity and specificity in clinical applications.

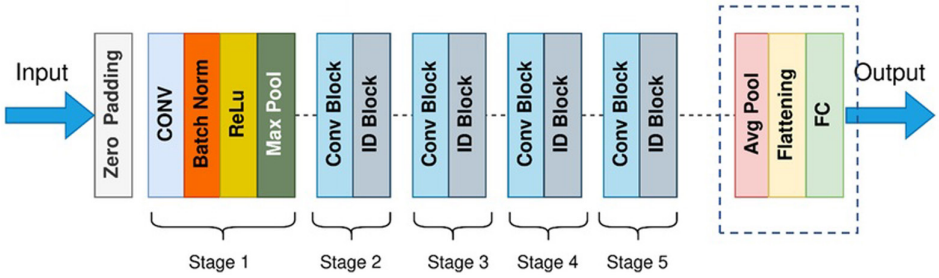

Figure 5. ResNet50 model architecture.

Table 1 the accuracy, precision, recall, and F1-score metrics obtained for each model across the training, validation, and test phases. The ResNet50 model was pretrained on the ImageNet dataset and fine-tuned by making the last 20 layers trainable. The model achieved a test accuracy of 90.83%, with precision, recall, and F1-score values of 86.67%, 90.70%, and 88.64%, respectively. The validation accuracy of ResNet50 was 90.57%, indicating that overfitting was effectively minimized and the model maintained a strong generalization capability.

Similarly, the VGG16 model was fine-tuned by unfreezing its last 10 layers, followed by training on the target dataset. The model achieved a test accuracy of 92.66 %, with precision, recall, and F1-score values of 90.70% across all metrics. The validation accuracy was 89.62%, demonstrating superior test accuracy compared to ResNet50.

The custom CNN model, designed from scratch, consisted of three convolutional blocks followed by fully connected layers. The model was trained for 42 epochs, achieving a test accuracy of 92.66%, with precision, recall, and F1-score values of 88.89%, 93.02%, and 90.91%, respectively. Notably, the validation accuracy of 96.23% suggests that the custom CNN exhibited a higher generalization performance compared to the transfer learning–based models.

Overall, the experimental results indicate that both VGG16 and the custom CNN model demonstrated superior classification performance, while ResNet50 exhibited slightly lower accuracy. The findings highlight the effectiveness of transfer learning for medical image classification, while also emphasizing the potential of custom-built architectures tailored to specific tasks.

Figure 6 presents the training and validation performance of three models: (a) Custom CNN, (b) VGG16, and (c) ResNet50. Each subfigure displays accuracy (left) and loss (right) over training epochs. The custom CNN model shows a consistent increase in training and validation accuracy, with minor fluctuations in validation accuracy. The validation loss decreases smoothly, indicating good generalization with minimal overfitting. In contrast, the VGG16 model exhibits noticeable fluctuations in validation accuracy, suggesting instability in learning. The ResNet50 model demonstrates stable performance, with training and validation accuracy curves closely aligned. The validation loss decreases steadily and follows the training loss trend, suggesting strong generalization and minimal overfitting.

Table 2 presents the confusion matrices for Custom CNN, VGG16, and ResNet50. ResNet50 achieves the highest number of correctly classified CMP samples (62), followed by Custom CNN (61) and VGG16 (60). In the normal class, Custom CNN correctly classifies 40 instances, while ResNet50 and VGG16 classify 39. Misclassification rates are slightly higher in VGG16, with 6 CMP samples incorrectly labeled as normal and 4 normal samples predicted as CMP. ResNet50 exhibits the lowest overall misclassification, indicating superior generalization performance.

DISCUSSION

This study demonstrates the effectiveness of deep learning models in classifying chondromalacia patella from MRI images. The comparative analysis of ResNet50, VGG16, and a custom CNN model highlights the significance of architectural selection in medical image classification. The superior performance of the custom CNN model, which achieved 92.66% test accuracy and an F1-score of 90.91%, suggests that task-specific architectures can exhibit better generalization capabilities for specific medical imaging tasks.

Deep learning–based studies on MRI analysis of musculoskeletal disorders have been previously investigated. Prior research has demonstrated the suitability of CNNs for classifying knee pathologies.[16] A recent study using 3D CNNs for knee osteoarthritis classification achieved 86.5% accuracy, demonstrating the effectiveness of deep learning in musculoskeletal imaging.[17] Similarly, a VGG16-based classification approach reported 86.7% accuracy, supporting the feasibility of transfer learning for knee pathology diagnosis.[18]

In comparison, this study showed that ResNet50 achieved 90.83% accuracy, surpassing previous models based on 3D CNNs and VGG16. However, the custom CNN model outperformed all other approaches, highlighting the importance of architecture optimization in improving classification performance. The ability of the custom CNN to achieve superior generalization suggests that designing models tailored to specific medical imaging tasks may be more effective than relying solely on transfer learning.

Furthermore, the low false positive and false negative rates observed in the confusion matrix suggest that the custom CNN model effectively distinguishes between CMP and normal cases, reducing the risk of misclassification. This improvement can be attributed to the use of dropout regularization, fine-tuned convolutional layers, and an optimized training strategy.

Additionally, the choice of the Adam optimization algorithm with a learning rate of 1e-4 contributed to stable training performance. Studies by Kingma and Ba[12] (2014) demonstrate that Adam outperforms traditional stochastic gradient descent in medical image analysis.

The batch size of 32 was chosen based on recent findings that smaller batch sizes can improve model generalization and latent feature

Table 1. Comparison of Training, Validation, and Test Metrics

Model	Training Accuracy	Training Precision	Training Recall	Training F1 Score	Validation Accuracy	Validation Precision	Validation Recall	Validation F1 Score	Test Accuracy	Test Precision	Test Recall	Test F1 Score
ResNet50	0.9720	0.9566	0.9735	0.9650	0.9057	0.8810	0.8810	0.8810	0.9083	0.8667	0.9070	0.8864
VGG16	0.9277	0.9041	0.9147	0.9094	0.8962	0.8605	0.8810	0.8706	0.9266	0.9070	0.9070	0.9070
Custom CNN	0.9720	0.9731	0.9559	0.9644	0.9623	0.9750	0.9286	0.9512	0.9266	0.8889	0.9302	0.9091

CNN, convolutional neural network.

Table 2. Confusion Matrices for Different Models

	True: CMP		True: NORMAL	
Model	CMP	NORMAL	CMP	NORMAL
Custom CNN	61	5	3	40
VGG16	60	6	4	39
ResNet50	62	4	4	39

CMP, chondromalacia patella.

extraction in medical imaging tasks,[19] while also balancing computational efficiency as demonstrated in studies with MobileNet V2 and ShuffleNet.[20]

The superior performance of the custom CNN model can be attributed to several key factors. First, its architecture was specifically designed for chondromalacia patella detection. While standard architectures such as ResNet50 and VGG16 are pretrained on large-scale datasets like ImageNet, the custom model was optimized to focus on the distinct characteristics of musculoskeletal MRI scans. Second, the low false positive and false negative rates observed in the confusion matrix suggest that the model effectively differentiates between varying severity levels of chondromalacia patella.

From a clinical perspective, the implementation of deep learning models for MRI-based chondromalacia patella diagnosis offers several advantages. One of the most notable benefits is the reduction of inter-observer variability, which is a common challenge in radiological assessments. Studies have reported inconsistencies in manual grading and interpretation of MRI scans, highlighting the need for automated decision-support systems to enhance diagnostic reliability and standardization.[6]

The high classification accuracy and F1-score achieved by the custom CNN model suggest that deep learning can provide objective and reproducible evaluations, assisting radiologists and medical specialists in clinical decision-making.[7,8] Furthermore, deep learning–based classification can significantly reduce interpretation time, allowing for more efficient and timely diagnoses, ultimately improving patient management and treatment planning.

This study highlights the effectiveness of deep learning models, particularly a custom-designed CNN, in diagnosing chondromalacia patella from MRI scans. The experimental results demonstrate that the custom CNN and VGG16 models achieve superior performance, while ResNet50 exhibits slightly lower accuracy. The findings emphasize the potential of custom-built architectures and transfer learning in medical image classification, while also underscoring the importance of further research into domain-specific optimizations. Despite the promising results, this study has several limitations. A primary concern is the limited dataset size, which may affect the model's generalizability across diverse patient populations. Utilizing larger and more varied MRI datasets could enhance model robustness and clinical applicability. Studies have shown that deep learning models trained on extensive and heterogeneous data perform better across different clinical settings. Another significant issue is the domain gap between natural images used in pretrained models (e.g., ImageNet) and medical images like MRI scans. This disparity can limit the effectiveness of transfer learning in medical imaging applications. Research indicates that models pretrained on medical-specific datasets, such as RadImageNet, outperform those trained on natural image datasets, underscoring the importance of domain-specific pretraining. Future

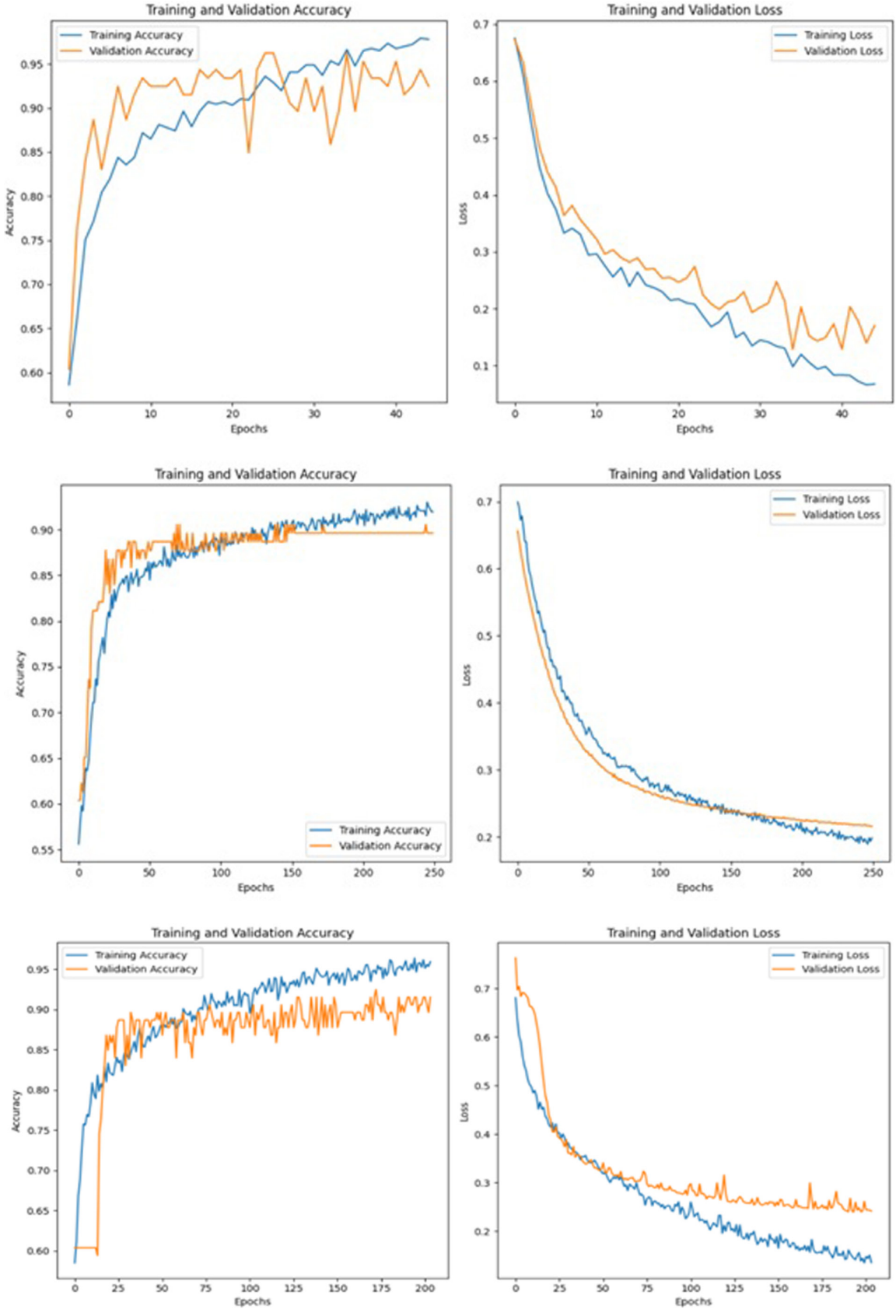

Figure 6. Comparison of training and validation performance across different models. (A) Training and validation accuracy/loss of custom CNN. (B) Training and validation accuracy/loss of VGG16. (C) Training and validation accuracy/loss of ResNet50.

work should focus on improving scalability, expanding the dataset, and developing fully automated deep learning pipelines to enhance clinical applicability. Given the scarcity of labeled medical imaging data, exploring semi-supervised learning techniques presents a viable alternative. Methods like self-supervised learning have shown promise in leveraging unlabeled data to improve model performance in medical image analysis. Additionally, incorporating training models on medical-specific datasets may further improve generalization. By addressing these aspects, deep learning approaches can contribute more effectively to the automated diagnosis of musculoskeletal conditions, ultimately improving clinical decision-making and patient outcomes.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Ethics Committee Approval: This study was approvad by the institutional ethics committee of Gazi University (Approval No: 1117365; Date: 13.12.2024).

Informed Consent: As this was a retrospective study, the imaging data were obtained from an existing database; therefore, informed consent was not required.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – O.O.; Design – O.S.; Supervision – O.O.; Resources – S.S.; Materials – O.H.; Data Collection and/or Processing – O.H.; Analysis and/or Interpretation – S.S.; Literature Search – O.O.; Writing Manuscript – O.O.; Critical Review – O.O.; Other – S.S.

Declaration of Interests: The authors have no conflict of interest to declare.

Funding: The authors declared that this study has received no financial support.

REFERENCES

1. McConnell J. The physical therapist's approach to patellofemoral disorders. *Clin Sports Med*. 2002;21(3):363-387. [CrossRef]
2. Glaviano NR, Kew M, Hart JM, Saliba S. Demographic and epidemiological trends in patellofemoral pain. *Int J Sports Phys Ther*. 2015;10(3):281-290.
3. Pak J, Lee JH, Lee SH. A novel biological approach to treat chondromalacia patellae. *PLoS One*. 2013;8(5):e64569. [CrossRef]
4. Powers CM. The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: a theoretical perspective. *J Orthop Sports Phys Ther*. 2003;33(11):639-646. [CrossRef]
5. Cucchiarini M, De Girolamo L, Filardo G, et al. Basic science of osteoarthritis. *J Exp Orthop*. 2016;3(1):22. [CrossRef]
6. Recht M, Bobic V, Burstein D, et al. Magnetic resonance imaging of articular cartilage. Clinical orthopaedics and related Research®. 2001;391:379-396.
7. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. [CrossRef]
8. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. 2017;19(1):221-248. [CrossRef]
9. Gold GE, Han E, Stainsby J, Wright G, Brittain J, Beaulieu C. Musculoskeletal mri at 3.0 t: relaxation times and image contrast. *AJR Am J Roentgenol*. 2004;183(2):343-351. [CrossRef]
10. Schiphof D. *Identifying Knee Osteoarthritis*. Dutch Arthritis Foundation. Amsterdam. Rotterdam: Optima Grafische Communicatie; 2011.
11. Liu F, Zhou Z, Samsonov A, et al. Deep learning approach for evaluating knee mr images: achieving high diagnostic performance for cartilage lesion detection. *Radiology*. 2018;289(1):160-169. [CrossRef]
12. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv Preprint ArXiv:1412.6980 2014.
13. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. arXiv Preprint ArXiv:1207.0580 2012.
14. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv Preprint ArXiv:1409.1556 2014.
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE; 2016:770-778. [CrossRef]
16. Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet. *PLoS Med*. 2018;15(11):e1002699. [CrossRef]
17. Guida C, Zhang M, Shan J. Knee osteoarthritis classification using 3d CNN and mri. *Appl Sci*. 2021;11(11):5196. [CrossRef]
18. Solak FZ. Classification of knee osteoarthritis severity by transfer learning from x-ray images. *Karaelmas Fen Ve Mu¨hendislik Dergisi*. 2024;14(2):119-133.
19. Kerley CI, Cai LY, Tang Y, et al. Batch size go big or go home: counter-intuitive improvement in medical autoencoders with smaller batch size. In: *Medical Imaging*: Image Processing. 2023;12464:106-115. SPIE
20. Biswas K, Pal R, Patel S, et al. A novel momentum-based deep learning techniques for medical image classification and segmentation. In: Berlin: International Workshop on Machine Learning in Medical Imaging; 2024:1-11. Springer