https://doi.org/10.65542/djei.v2i2.42
Dasinya Journal for Engineering and Informatics. **2026**, *2*, 19.

Article

Improving Knee Osteoarthritis Classification Using Swarm-Based Optimization and Deep Learning Models

Dilan Jameel Sulaiman[1], **Baraa Wasfi Salim**[2]

[1] Information Technology Management Department, Technical College of Administration, Duhok Polytechnic University, Duhok, Iraq, dilan.sulaiman@dpu.edu.krd
[2] Information Technology Management Department, Technical College of Administration, Duhok Polytechnic University, Duhok, Iraq, baraa.salim@dpu.edu.krd
* Correspondence: dilan.sulaiman@dpu.edu.krd

Abstract

Knee Osteoarthritis (KOA) is a persistent and expensive joint condition which can be accurately identified with radiographic examination. This study suggests combining deep models with swarm intelligence to diagnose KOA through X-ray images. A systematic review was carried out to find the best strategies for detecting KOA. After the data preprocessing, features were obtained by using the pre-trained CNNs (EfficientNetB0, VGG19, ResNet50, MobileNetV2) model. Multilayer Perceptron (MLP) classifiers were trained using the features obtained. The collected features of both models were fed into classifiers like MLP. In the second phase, optimal MLP hyperparameters were calibrated via Particle Swarm Optimization (PSO) to improve the classification performance. The experiments were conducted on a balanced dataset containing 10,000 knee X-ray images across five KL grades. Among all the models we investigated, the VGG19 along with PSO and MLP achieved the accuracy of 98.8% on test set. The model also achieved 98.8% precision, 98.8% recall, and 98.8% F1-score, outperforming the corresponding non-optimized baseline models. The results indicate that the integration of swarm systems with deep learning significantly enhances the recognition performance of KOA grade. Based on such a technique, clinicians of the field of orthopedics would be able to diagnose at an earlier stage more easily and accurately.

Keywords: Knee Osteoarthritis; Particle Swarm Optimization; Deep Learning; Convolutional Neural Networks; X-ray; Image Classification.

1. Introduction

Knee Osteoarthritis (KOA) is a long-term chronic disease characterized by the degeneration of the joints which primarily affects old people, induced by aging, previous joint injury, and obesity. It is associated with symptoms to knee joints that affect inflammation, pain, stiffness and loss of movement [1]. KOA is considered one of the leading causes of disability worldwide and represents a major socioeconomic burden because of its high prevalence, long-term treatment cost, and impact on quality of life. Early diagnosis and proper classification of KOA is essential to provide adequate therapy according treatment of the same patient. Traditionally, KOA diagnosis was based on physical examination findings and radiographic methods such as X-ray and MRI. Imaging modalities have the distinct benefit of allowing for visualization of joint [degeneration] and other related pathology to allow clinicians to arrive at a more definitive diagnosis [2]. Innovations in AI and digital medical imaging have now enabled the application of Deep Learning (DL) algorithms to automate repetitive diagnostic tasks based on novel paradigms [3].

Received: February 21, 2026; Revised: May 06, 2026; Accepted: May 06, 2026; Published: May 14, 2026; Available online: May 14, 2026.

One of the celebrated Deep Learning (DL) model, particularly CNNs, is a standard approach for medical image pattern recognition as it has the capability to automatically learn abstract features from deep hierarchy [4]. without any manual extraction of features. The Kellgren-Lawrence (KL) grading system is made up of 5 grades: grade 0 (normal), grade 1 (doubtful), grade 2 (mild), grade 3 (moderate), and grade 4 (severe joint space narrowing and bone deformity) for KOA [5]. Figure 1 demonstrates various levels of knee OA according to KL grading [6].

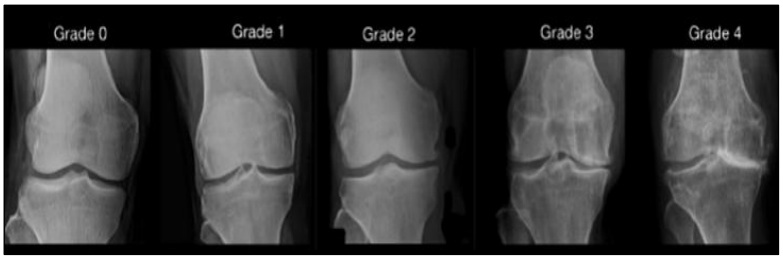

Figure 1. Knee Osteoarthritis Levels According to Kellgren-Lawrence Grades [6].

While CNNs-based models suffice, their performance is heavily reliant on tunable hyperparameter settings like the number of hidden units, dropout rates, learning rates, and epochs [7]. Setting these parameters manually can lead to poor performance. To get over this limitation, the current work presents a hybrid framework based on deep learning and Swarm Intelligence (PSO) for optimizing hyper-parameters of MLP classifier fed with image features obtained from pre-trained CNNs to enhance classification performance [8][9].

This work used several CNNs architectures, including VGG19, ResNet50, MobileNetV2, and EfficientNetB0, as feature extractors for knee X-ray images. The practical significance of this work lies in providing a reliable automated grading framework that may support clinicians in early diagnosis, reduce workload, and improve treatment planning.

The main contributions of this study are summarized as follows:

1. A hybrid framework combining deep learning feature extractors with Particle Swarm Optimization (PSO) is proposed for automatic multi-class knee osteoarthritis classification using X-ray images.
2. Four well-known pre-trained CNNs architectures, namely VGG19, ResNet50, EfficientNetB0, and MobileNetV2, were systematically evaluated as feature extraction backbones.
3. PSO was employed to automatically optimize key MLP hyperparameters, including hidden units, learning rate, dropout rate, and training epochs, reducing manual trial-and-error tuning.
4. The proposed optimization strategy significantly improved classification accuracy compared with baseline models without PSO.
5. Among all evaluated models, the VGG19 + PSO framework achieved the best overall performance, demonstrating the effectiveness of swarm-based optimization for medical image classification.
6. The proposed system provides a practical computer-aided diagnostic tool that may support clinicians in early and accurate grading of knee osteoarthritis.

The rest of the article is structured as follows: Section 2 outlines previous works on knee osteoarthritis classification using X-ray images and methods that have been employed to explain model behavior, while Section 3 presents the proposed methodology used in this study, section 4 presents the experimental results, Discusses the limitations, challenges, and future work of the proposed study in Section 5. and Section 6 concludes the study.

2. Related work

Knee osteoarthritis is gaining increasing attention from the medical imaging and machine learning communities. As KOA diagnosis relies on images and is manual, researchers have tapped into DL specifically CNNs for automating the process of classifying KOA using radiographic images including X-ray and MRIs.

Various DL architectures show to be effective for classifying severity classes of KOA. For example, Qadir et al. [10] created a hybrid model that combines ResNet for feature extraction and BiLSTM for classifying KOA severity across five KL grades. Trained using the Mendeley V1 dataset (2,000 X-rays), with a validation set from OAI data and achieving 78.57% cross-validation, and 84.09% test accuracy. The study suffered from limited datasets, lack of optimization

approaches (for example PSO), and no probing model interpretability, which limits the clinical usability and scalability. Similarly, Thomures et al. [11] introduced an EfficientNetB5-based architecture to classify five types of KOA using transfer learning on 9 786 knee X-rays from a Kaggle dataset. I implemented a data-centric preprocessing pipeline involving outlier removal and Cleanlab-based label corrections, which improved accuracy to 82.07%, beyond previous state of the art like ResNet-101 (69%). However, this study struggles with ambiguous cases (KL Grade 1); it did not address optimization approaches or tool for model interpretability. Over-reliance on one dataset also raises issues about generalization to other clinical contexts

Researchers have also investigated deeper architectures such as ResNet-50. Wang et al. [12] A two-stage classification method was established with VGG for locating the joints center and ResNet-50 based model to rank KOA severity from the OAI dataset. The class imbalance was addressed through bandwidth filtering, histogram normalization and data rebalancing yielding 81.41% accuracy. Though effective, this is computationally costly and entirely based on a single dataset with limited scalability. It also doesn't include optimization methods and interpretability tools, making it less suitable for clinical application. Sami Mohammed et al. [13] passed six pre-trained deep neural network models to evaluate, deriving the three-class intensity level classification on OAI dataset with ResNet101 giving an accuracy of 89%. The study revealed class imbalance, no optimization techniques used, and failure to evaluate different datasets, which indicate limited generalizability Hybrid and ensemble models also show attractive results Bugday et al. [14] proposed a feature-fusion technology to concatenate the original and denoised image by DenseNet201. Then, they applied NCA for feature selection and trained a SVM classifier to achieve an accuracy of 85%. Even though the results were promising, the approach added new complexity and had some issues regarding optimization and interpretability with limited evaluation on broader datasets. YEOH et al. [15] analyzed MRI data and used 3D convolutional neural network architectures with transfer learning, achieving the highest area under curve (AUC) value of 0.945 using ResNet18. Although these results illustrate the promise of MRI-based systems, the study was focused on binary classification only and it lacked interpretation or optimization strategies. A 12-layer CNN model proposed by Rani et al. [16] using the OAI dataset reached an accuracy of 92.3% in binary classification and 78.4% in multi-class classification. Although the model could significantly outperform previous methods, it was only trained on one dataset and did not consider aspects like optimization problems, data imbalance or computational cost. Harish et al. [17] applied a simple CNN with the VGG architecture on a dataset from Kaggle with 3,836 images for 74% binary classification accuracy. It was computationally efficient, but not sophisticated, not optimized, nor generalizable as it outperformed more strategic approaches.

A summary of the most recent and notable techniques for KOA classification is shown in Table 1; along with a comparative review of models, data sets and classification regime and reported performance.

showed that accuracy of classification could be improved by using the PSO algorithm to modify parameters such as learning rate, dropout rate and number of hidden units [8]. demonstrated that PSO could greatly enhance the accuracy of classification by changing parameters like the learning rate, the dropout rate, and the number of hidden units [18]. PSO has a great potential in that it is still relatively new and underutilized because the use of PSO with different deep learning feature extractors for KOA identification remains limited, and the literature on this subject is scarce [19].

Table 1. A Summary of Recent Studies that used Deep Learning Models to Classify KOA.

Ref	Year	Model	Datasets	Accuracy	Key Finding
[1]	**2023**	**ResNet + BiLSTM**	**OAI, Mendeley**	**84.09%**	**Robust hybrid model for 5-class KOA classification**
[2]	**2025**	**EfficientNetB5**	**OAI**	**82.07%**	**Data cleaning significantly improved accuracy**
[3]	**2022**	**VGG + ResNet-50**	**OAI**	**81.41%**	**Improved preprocessing and grading accuracy**
[4]	**2023**	**ResNet101, VGG16/19, MobileNetV2**	**OAI**	**89%**	**ResNet101 best for 3-class KOA classification**

[5]	**2025**	**DenseNet201 + SVM with feature fusion**	**OAI**	**85%**	**Combined original & denoised images for better features**
[6]	**2023**	**3D CNN with transfer learning**	**OAI (MRI scans)**	**AUC = 0.945**	**MRI-based method showed strong binary performance**
[7]	**2024**	**Custom 12-layer CNN**	**OAI**	**92.3% (Binary) 78.4% (Multi)**	**Outperformed previous CNN models**
[8]	**2023**	**CNN (VGG-based)**	**OAI**	**74%**	**Simple baseline for KOA detection**

This collection of papers showcases the diversity in KOA identification models and methods, and also identifies places where further merging of optimization theory may yield improved outcomes and performance potential. Previous studies have shown promising results for KOA classification using deep learning models; however, several limitations remain. Many existing approaches relied on a single CNN architecture, limited class settings (binary or three-class classification), manually selected hyperparameters, or lacked optimization strategies and model interpretability. In addition, some studies reported performance using only one dataset, which may limit generalizability. In contrast, the proposed framework evaluates multiple pre-trained CNN architectures (VGG19, ResNet50, EfficientNetB0, and MobileNetV2) under identical settings and integrates Particle Swarm Optimization (PSO) to automatically optimize the MLP classifier hyperparameters. This combination improved classification performance and reduced manual trial-and-error tuning. Furthermore, the proposed method addresses the more challenging five-class KOA grading problem, making it a practical and robust framework for automated KOA assessment.

3. Methodology

The proposed methodology was developed to build an effective automated system for classifying the severity of Knee Osteoarthritis (KOA) using X-ray images. The framework consisted of a sequence of consecutive stages, beginning with dataset collection and preprocessing, followed by image enhancement, deep feature extraction using pre-trained CNNs models, and classifier optimization using Particle Swarm Optimization (PSO). Specifically, the CNNs models were used to extract discriminative deep features from knee X-ray images, these feature vectors were then provided to the MLP classifier, while PSO was employed to optimize the MLP hyperparameters for improved classification performance. Figure 2 illustrates the overall workflow of the proposed system for predicting KOA severity grades from raw knee X-ray images. All stages were carefully designed to improve diagnostic accuracy, model robustness, and generalization capability. The final stage involved training and testing the classification models using unseen datasets to evaluate their effectiveness and reliability. The main objective of this framework was to develop an efficient deep learning pipeline capable of accurately distinguishing KOA severity levels and supporting intelligent medical imaging systems.

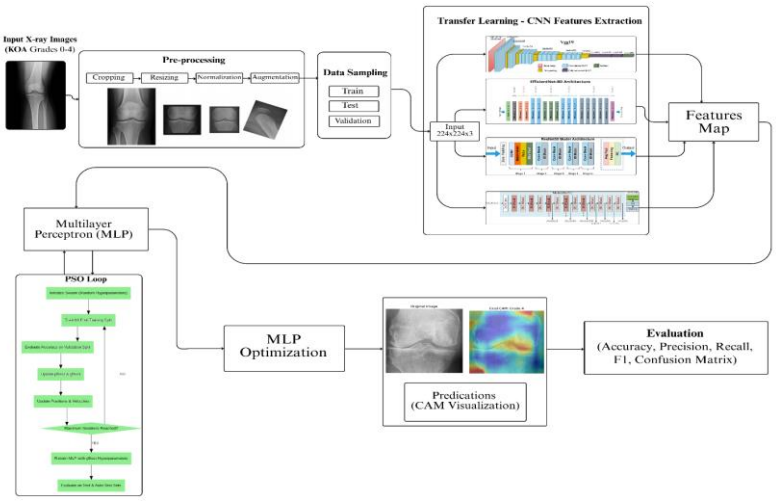

Figure 2. Workflow of the Proposed KOA Classification System.

3.1. Dataset

The dataset was obtained from the Kaggle platform and is based on the publicly available Osteoarthritis Initiative (OAI) repository. The dataset was downloaded in December 2023 and consisted of 7,000 knee X-ray images annotated according to the Kellgren–Lawrence (KL) grading system, ranging from Grade 0 (Normal) to Grade 4 (Severe). The dataset was balanced across all five classes, with the following distribution: KL 0: 1,400 images; KL 1: 1,400 images; KL 2: 1,400 images; KL 3: 1,400 images; and KL 4: 1,400 images. Before applying data augmentation, the dataset was randomly divided at the image level into three mutually exclusive subsets: 80% for training, 10% for validation, and 10% for testing. The validation and test sets were preserved without augmentation to ensure an objective performance evaluation and to prevent data leakage between subsets.

3.2. Preprocessing

Several preprocessing steps were applied to improve image consistency and enhance learning performance. Each radiograph was first cropped to focus on the knee joint region of interest (ROI) and then resized to 224×224 pixels to match the input requirements of the pre-trained CNNs architectures. Pixel intensity values were normalized to the range [0, 1] to stabilize training and improve convergence. Data augmentation was applied exclusively to the training set after dataset partitioning to enhance generalization and mitigate overfitting. The augmentation pipeline included random rotations, horizontal flips, zooming, shear transformations, and brightness adjustments (see Figure. 3).

The validation and test sets underwent only deterministic preprocessing (cropping and resizing) without augmentation. This strict separation ensured that no original image or augmented variant appeared in more than one subset, thereby preventing data leakage.

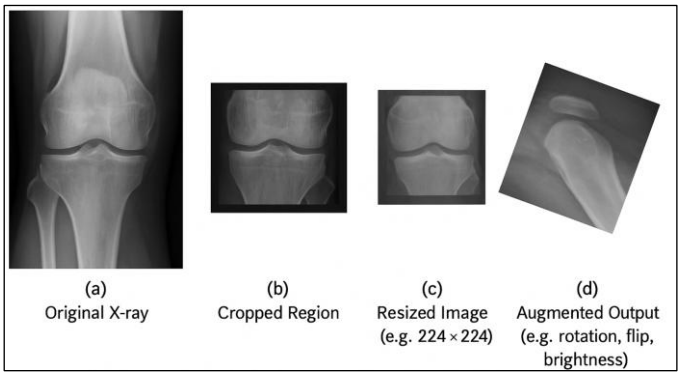

Figure 3. Example Preprocessing Pipeline

After applying the data augmentation process the dataset was balanced across all Kellgren–Lawrence (KL) grades. Specifically, a set of geometric and intensity-based augmentation techniques including rotation, horizontal flipping, width and height shifting, zooming, and brightness adjustment was applied to the original training images using the Keras ImageDataGenerator. These transformations were used to increase sample diversity and ensure an equal number of images per class. As a result, each KL grade was expanded to contain 2,000 images, leading to a balanced dataset of 10,000 images in total.

3.3. Feature Extraction Using CNNs

Feature extraction is a critical stage in deep learning-based classification of Knee Osteoarthritis (KOA). In this study, convolutional neural networks (CNNs) were employed to automatically extract discriminative features from preprocessed knee X-ray images. CNNs are highly effective for medical image analysis because they can capture spatial hierarchies and anatomical patterns associated with KOA progression. To improve generalization performance, transfer learning was adopted using pre-trained CNNs models originally trained on ImageNet. Four CNNs architectures, namely VGG19, ResNet50, EfficientNetB0, and MobileNetV2, were selected as feature extractors. These models were

selected based on their proven effectiveness in image classification tasks and their diverse architectural characteristics. VGG19 offers strong feature extraction capability, ResNet50 provides residual deep learning advantages, EfficientNetB0 balances accuracy and efficiency, while MobileNetV2 is suitable for lightweight deployment. This diversity supports a comprehensive comparison to identify the most effective feature extractor for the proposed KOA classification framework. The final classification layers of each model were removed, and the resulting outputs were used as feature vectors for the classification stage. A summary of the main architectural characteristics, parameter scale, and computational complexity of the employed CNNs models is presented in Table 2.

Table 2. Summary of the Employed Pre-Trained CNNs Architectures used for Feature Extraction

Model	Main Architecture Type	Approx. Parameters	Model Size	Input Size	Output Feature Strategy	Complexity
VGG19	**Sequential deep CNN with 3×3 convolutions**	**~143 million**	**Heavy**	**224 × 224 × 3**	**Flatten**	**High**
ResNet50	**Residual deep CNN with skip connections**	**~143 million**	**Medium**	**224 × 224 × 3**	**Global Average Pooling**	**Medium**
Efficient NetB0	**Compound-scaled CNN**	**~5.3 million**	**Light-weight**	**224 × 224 × 3**	**Global Average Pooling**	**Low**
MobileNetV2	**Depth wise separable CNN with inverted residuals**	**~3.5 million**	**Light-weight**	**224 × 224 × 3**	**Global Average Pooling**	**Very Low**

VGG19 is a Deep Neural Network (DNN) with 19 layers that is known for being easy to use and always building the same way. This program makes use of tiny 3x3 filters, and is applicable to medical images that are well organized [20]. Figure 4 illustrates the internal structure of the VGG19 network to be used in this study to explain how the features are extracted. The network is initiated by convolutional layers that identify the important features of the knee X-ray images and then the pooling layers that reduce the dimensionality and the complexity of the computation. It removes the fully connected layers at the final end of network to vary the model to act as a feature extractor and not a classifier [21].

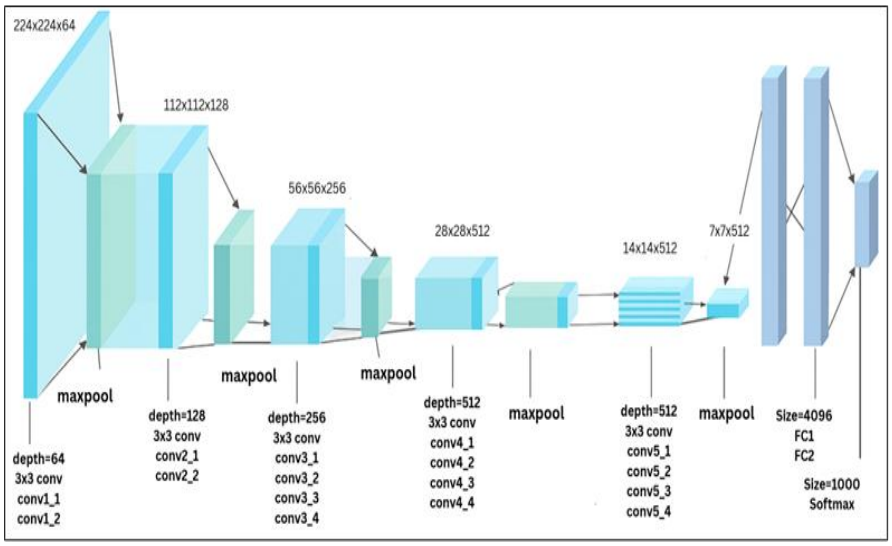

Figure 4. VGG-19 architecture with convolutional, pooling, fully connected layers, and SoftMax classifier [9].

- The ResNet50 is applicable in fine-grained feature extraction and training deep networks because it has residual connections to minimize the impact of vanishing gradients [22].
- EfficientNetB0 applies a compound scaling technique to balance the depth, width, and resolution, which is an option that is lightweight in implementation yet high-performance [23].
- MobileNetV2 that applies depth wise separable convolutions, reversed residuals to reduce computations but still achieve accuracy is optimized to run in mobile devices and edge devices [24].

The images were analyzed by each CNNs to generate complex, high-dimensional feature vectors which acted as inputs in the step of classifier.

All models received input images of size 224 × 224 × 3 pixels. VGG19 produced a feature map of size 7 × 7 × 512, which was flattened into a feature vector of length 25,088. The original VGG19 architecture consists of repeated convolutional and max-pooling layers followed by fully connected and SoftMax classification layers. In this study, the final classification layers were removed so that the model could be used solely for feature extraction. ResNet50 generated a 2,048-dimensional feature vector using Global Average Pooling (GAP), while both EfficientNetB0 and MobileNetV2 produced compact feature vectors of 1,280 dimensions. The variation in feature dimensions reflects differences in network depth, architecture, and pooling strategy. VGG19 preserves more spatial details through flattening, whereas the remaining models use GAP to reduce dimensionality and computational complexity. The extracted feature vectors were subsequently fed into the MLP classifier, which was further optimized using Particle Swarm Optimization (PSO) to improve classification accuracy and generalization performance.

3.4. Multilayer Perceptron (MLP)

The MLP served as a classification head for the features extracted from four pre-trained CNNs architectures, which are VGG19, EfficientNetB0, ResNet50 and MobileNetV2. The CNNs were used as fixed feature extractors and their output feature vectors were fed into the MLP. The MLP Models architecture, number of hidden layers, number of neurons in each layer, used activation functions, dropout rates, and learning rate were tailored based on each model to allow for a tradeoff between complex models and generalization performance. The hidden layers employed the ReLU activation function to improve nonlinear learning capability.

The output layer in all cases consisted of five neurons, corresponding to the five Kellgren–Lawrence (KL) grades of knee osteoarthritis severity (Normal, Grade 1, Grade 2, Grade 3, and Grade 4). A SoftMax activation function was applied in the output layer to perform multi-class classification The MLP classifier was subsequently optimized using Particle Swarm Optimization (PSO) to identify the most effective hyperparameter configuration and further improve classification accuracy.

3.5. Particle Swarm Optimization (PSO)

In the proposed pipeline, the CNNs models were first used to extract deep features from knee X-ray images. These feature vectors were then used as inputs to the MLP classifier. PSO was subsequently employed to optimize the MLP hyperparameters using validation accuracy as the fitness function before final testing. It is important to distinguish between the roles of Adam and PSO in the proposed framework. Adam was used as the gradient-based optimizer to update network weights during MLP training, whereas PSO was employed to optimize the MLP hyperparameters, including hidden units, dropout rate, learning rate, and training epochs. Therefore, Adam optimized model parameters, while PSO optimized the training configuration.

As a metaheuristic optimization approach, PSO was integrated to optimize the classifier design. Inspired by the collective actions of flocks of birds, PSO simulated a population-based search, with each particle exploring a hyperparameter space. The goals were to minimize classification loss and improve validation accuracy. Key parameters tuned included the number of training epochs, learning rate, dropout rate, and hidden units. The optimization strategy

significantly improved the classifier's performance by automatically identifying efficient configurations without the need for any human modification.

In PSO, candidate solution is considered as a particle which roams around the search space. There are two basic attributes of a particle:

- Position (xi): representing the current candidate solution.
- Velocity (vi): controlling the direction and step size of the particle's movement.

Each particle keeps track of:

- Its personal best position (pbest) – the best solution it has achieved so far.
- The global best position (gbest) – the best solution found by the entire swarm.

During optimization, the velocity and location of particles are adjusted at every iteration using [18], the following equations:

$$\left(X_{i,d}(t) - dgBest\right) \cdot 2^r \cdot 2^C + (x_i, d(t) - i, dPBest) \cdot 1^r \cdot 1^C + v_{i,d}(t) = (1 + v_{i,d}(t)) \quad (1)$$

$$1 + v_{i,d}(t) = 1 + x_{i,d}(t) \quad (2)$$

Where:

- w = inertia weight (balance between exploitation and exploration).
- 1C, 2C= Factors of cognitive and social acceleration.
- 1r, 2r = Random numbers in [0,1].

These updates allow the swarm to balance:

- Exploration: Exploring newer areas of the search space.
- Exploitation: Refining search around the best-known solutions.

This procedure keeps on till the convergence requirements are reached (e.g., reaching maximum iterations or minimal error) [18].

It is an optimization that follows a systematic process. The second step was to define the MLP with a 5-class SoftMax output layer as the model under optimization. The fitness function was configured to maximize validation accuracy, while training was performed for the Adam optimizer and sparse categorical cross-entropy loss (with a batch size of 32). A validation set was used for scoring and the training set was used for fitting. PSO parameters were set to swarm size of 10 particles and a maximum of 5 iterations. For VGG19 and EfficientNetB0, as both models were exhibiting signs of overfitting, a smaller range was applied to the dropouts that were incremented by 0.01 until reaching [0.10-0.60]. In the case of ResNet50 and MobileNetV2, a wider search range was applied without dropout adjustment. In PSO we represent a candidate solution as four key hyperparameters: Hidden Units, Dropout Rate, Learning Rate and Number of Epochs. In Table 3, we summarize the search ranges used for each CNNs feature extractor.

Table 3. PSO hyperparameter Search Spaces Per Feature Extractor

CNNs features → MLP	Hidden units	Dropout	Learning rate	Epochs
VGG19 → MLP	**64 - 384**	**0.20 - 0.40**	**0.0001 - 0.005**	**0 - 50**
EfficientNetB0 → MLP	**64 - 512**	**0.10 - 0.40**	**0.0001 - 0.005**	**0 - 50**
ResNet50 → MLP	**64 - 512**	**0.10 - 0.50**	**0.0001 - 0.005**	**0 - 50**
MobileNetV2 → MLP	**64 - 512**	**0.10 - 0.50**	**0.0001 - 0.005**	**0 - 50**

Through repetitive assessment of their performance, each model converged to the best-performing set of hyperparameters, yielding significant improvements in performance over settings chosen manually. Table 4 summarizes the final optimized hyperparameters identified by PSO for all four CNNs feature extractors.

Table 4. Best Hyperparameter Configurations Identified by PSO Across All Models

Model	Best Hidden units	Best Dropout	Best Learning rate	Best Epochs
VGG19 → MLP	**384**	**0.30**	**0.0001**	**28**
EfficientNetB0 → MLP	**512**	**0.18**	**0.0002**	**21**
ResNet50 → MLP	**455**	**0.17**	**0.0001**	**27**
MobileNetV2 → MLP	**240**	**0.16**	**0.0001**	**23**

3.6. Model Training, Testing and Evaluation

The dataset was divided into training, validation, and testing subsets. During the training phase, CNNs-extracted feature vectors were used as inputs to the MLP classifier. PSO was applied to identify the optimal hyperparameter configuration based on validation accuracy. After optimization, the best model was retrained and finally evaluated on an independent unseen test set to assess classification performance and generalization capability. In order to improve learning efficacy and reduce the chance of overfitting, the DL models were trained with a combination of training and validation sets. Each of the models were optimized using PSO and then tested on the given test set. Performance measures were defined to assess the model using the model accuracy of categorizing disability and specific grades of KOA in the knee as a whole.

The subsequent static rate measures were implemented in order to further evaluate the models:

- **Accuracy**: It is the ratio of number of correctly classified instances (TP + TN) to the total instances.
- **Precision:** The ratio of true positive predictions to the total number of positive predictions.
- **Recall:** This is the proportion of genuine positives that were successfully detected out of all real positives.
- **F1 Score:** A correct evaluation based on the harmonic mean of recall and accuracy.

Equations (1) through (4) describe the performance metrics. "TP" represents for true positive; "TN" stands for true negative; "FP" stands for false positive; and "FN" refers for false negative.

$$\text{Accuracy}=(TP+TN)/(TP+FN+TN+FP) \quad (1)$$

$$\text{Precision}=TP/(TP+FP) \quad (2)$$

$$\text{Recall}=TP/(TP+FN) \quad (3)$$

$$\text{F1 Score}=2\times(\text{Precision}\times\text{Recall})/(\text{Precision}+\text{Recall}) \quad (4)$$

This procedure produced an effort that each model's capacity to use the features of X-ray images to categorize the severity of KOA was accurately measured. After PSO optimization, Table 5 shows additional evaluation metrics, including F1 score, recall, and precision and accuracy.

Table 5. Performance Metrics of the Proposed Models Including F1 Score, Recall, and Accuracy

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)

VGG19 + PSO	**98.8%**	**98.8%**	**98.8%**	**98.8%**
ResNet50 + PSO	**89.2%**	**90.4%**	**89.2%**	**88.5%**
EfficientNetB0 + PSO	**88.5%**	**88.9%**	**88.5%**	**88.4%**
MobileNetV2 + PSO	**92.2%**	**92.2%**	**92.2%**	**92.1%**

Based on these criteria, the VGG19+PSO model did better than the other designs, getting the highest accuracy and F1 score on all KOA scores.

4. Results and discussions

In order to determine the effectiveness of swarm-based optimization to classify knee osteoarthritis (KOA), numerous deep learning models have been created and tested with and without the presence of PSO. The VGG19 + PSO model was the most accurate with 98.8%. Confusion matrices confirmed this fantastic result and ensured the accurate prediction of all the levels of KOA severity.

Conversely, models that omitted PSO were quite poor. As an example, MobileNetV2, without PSO optimization, was able to achieve an accuracy of 68.8% as opposed to ResNet50, without PSO, which achieved 56.2%. These findings indicate that PSO is an effective way of enhancing the accuracy of classification through automated hyperparameter optimization.

4.1. Accuracy and Loss Trends

Figures 5 (accuracy) and (loss) show the VGG19 + PSO model's training and validation performance. While the training accuracy gradually increases to around 88% by the last epoch, the validation accuracy shows a robust and steady improvement, eventually approaching 95%. At the same time, there is a steady lower trend in both training and validation losses, suggesting efficient learning and generalization.

The model benefited from the adjusted hyperparameters that PSO brought as well as the data augmentation techniques that were used, as seen by the notable increase in validation performance. Furthermore, the model's resilience and capacity to preserve generalization across unknown data are validated by the lack of overfitting.

To further reduce overfitting, several regularization strategies were employed, including dropout layers, data augmentation, and EarlyStopping based on validation performance. Transfer learning using pre-trained CNNs models also improved feature robustness and generalization capability. In addition, final model evaluation was performed on an independent unseen test set to assess performance reliability.

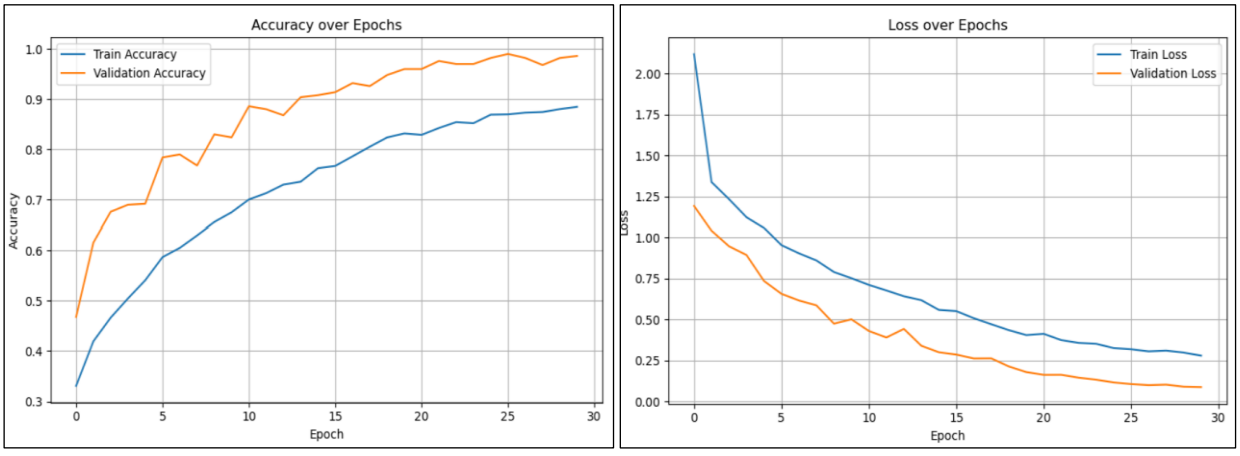

Figure 5. Training, Validation Accuracy and Loss for VGG19 with PSO.

4.2. Model Comparison: With and Without PSO

A comparison between models trained using PSO and their original counterparts without PSO was carried out in order to assess the effect of PSO on model performance. PSO was applied to improve model learning and classification performance during the hyperparameter tuning phase. The findings, which are shown in Table 6, show that using PSO significantly increased accuracy across all models.

The VGG19 model showed the greatest improvement, improving by 33.6% after being optimized using PSO. ResNet50 also saw an increase of 33.0%, demonstrating the significant impact of PSO on deeper architectures. Significant gains of 18.7% and 23.4% were also shown by EfficientNetB0 and MobileNetV2, respectively.

Table 6. The Accuracy Results

Model	Without PSO (Accuracy)	With PSO (Accuracy)	Gain
VGG19 + PSO	**65.2%**	**98.8%**	**+33.6%**
ResNet50 + PSO	**56.2%**	**89.2%**	**+33.0%**
EfficientNetB0 + PSO	**69.8%**	**88.5%**	**+18.7%**
MobileNetV2 + PSO	**68.8%**	**92.2%**	**+23.4%**

The PSO algorithm played a key role in fine-tuning hyperparameters such as the number of hidden units, learning rate, dropout, and epochs improving the classification performance across all models.

4.3. Confusion Matrix Analysis

The confusion matrices in Figure 6 (with PSO) and F (Test Set without PSO), VGG19 +PSO demonstrate that the model achieved perfect predictions across all five KOA classes (Normal, Grade 1 to Grade 4).

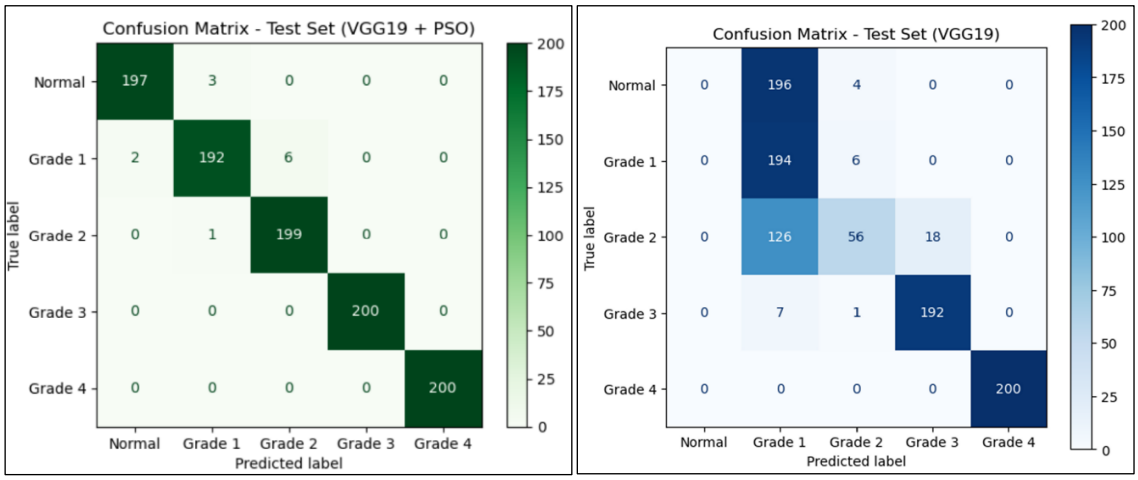

Figure 6. Confusion matrices on tested datasets for VGG19 with and without PSO.

Such high consistency validates the model's durability and generalization ability when applied to unknown data, in addition to its ability to properly identify the severity of KOA. This result indicates that PSO may enhance MLP-based classification better.

4.4. Swarm Optimization Progress

To improve the MLP classifier's performance, PSO employed to modify many important hyperparameters, such as the number of hidden units, learning rate, dropout rate, and epochs. During 60 iterations, the PSO successfully struck a balance between exploration and exploitation to find high-performing combinations with little human labor. The validation accuracy rapidly increased, confirming the algorithm's capacity to converge on optimum solutions. The ability of swarm intelligence to automate the adjustment of hyperparameters for deep learning models is shown here, and Figure 7 depicts the development of the PSO.

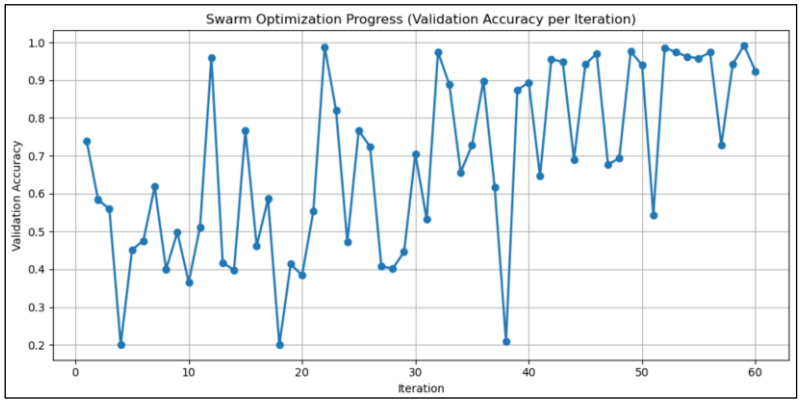

Figure 7. Enhancement in the Precision of Validation during PSO Iterations.

4.5. Grad-CAM Visualization for Model Interpretability

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to enhance comprehension of the VGG19 + PSO model's decision-making process. The areas in the input knee X-ray pictures that go into the categorization output are highlighted by this technique. A greater understanding of the regions impacting each forecast was made possible by superimposing the generated heatmaps on top of the original images. The model concentrated on small structural alterations and subtle changes in the joint space, which are frequently hard to see with the naked eye, as seen in Figure 8 for early-stage osteoarthritis (Grade 1). The model focused on specific characteristics such marginal osteophytes and joint space shortening in situations of severe osteoarthritis (Grade 4). The model's focus during prediction is revealed by these visuals, which also validate that the model is in line with clinically significant aspects.

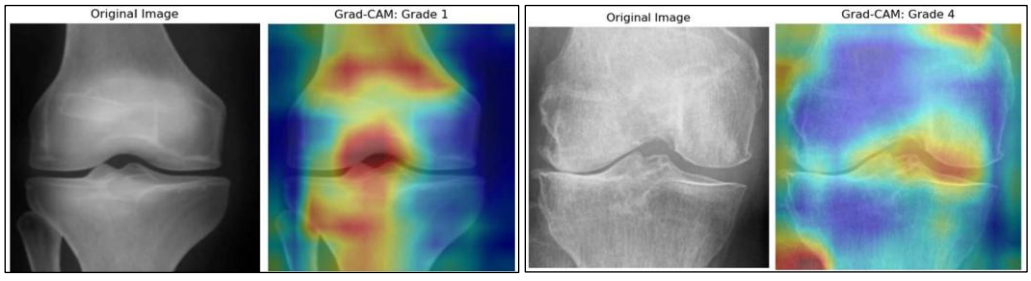

Figure 8. Grade 4-Classified Active Regions Are Displayed on the Knee X-Ray using the Grad-CAM Overlay.

4.6. Comparison With Previous Works on KOA Classification

In this study, four deep learning architectures (VGG19, ResNet50, EfficientNetB0, and MobileNetV2) were evaluated under two configurations: baseline (without PSO) and optimized (with PSO). The integration of PSO for hyperparameter tuning significantly improved the performance of all models, with VGG19 + PSO achieving the highest accuracy of 98.8% as show in Table 7.

Table 7. Comparative Results of the Proposed Models (With and Without PSO) With Previous Studies on KOA Classification

Ref	Year	Dataset	No. of Class	No. of Dataset	Method	Use of PSO	Accuracy
[10]	**2023**	**OAI**	**3**	**2,874**	**EfficientNet B5**	**No**	**97%**
[11]	**2023**	**OAI**	**5**	**5,778**	**EfficientNet-B0**	**No**	**69.74%**
[12]	**2023**	**OAI**	**5**	**2,578**	**EfficientNet B3**	**No**	**92%**
[13]	**2023**	**OAI**	**5**	**9,786**	**ResNet-34,** **VGG-19** **DenseNet 121** **DenseNet 161**	**No**	**95%** **96%** **96%** **97%**
[14]	**2023**	**OAI**	**2** **3** **4** **11**	**4,255**	**VGG16-GAP-KNN**	**No**	**2-class: 95.74%** **3-class: 92.60%** **4-class: 92.12%** **11-class: 89.75%**
[15]	**2023**	**OAI**	**2**	**1,752**	**VGG-16, RCNN (transfer learning)**	**No**	**98.6%**
[16]	**2024**	**OAI**	**5**	**9,786**	**ResNet50** **Xception** **Inception** **ResNetV2** **Ensemble**	**No**	**62%** **59%** **61%** **68%**
[17]	**2024**	**OAI**	**5**	**8,260**	**ResNet-101,** **Evidential Deep Learning (EDL)**	**No**	**72%**
This Study	**2025**	**OAI**	**5**	**9,786**	**VGG19** **ResNet50** **EfficientNetB0** **MobileNetV2**	**No**	**64.2%** **56.2%** **69.8%** **68.8%**
This Study	**2025**	**OAI**	**5**	**9,786**	**VGG19** **ResNet50** **EfficientNetB0** **MobileNetV2**	**Yes**	**98.8%** **89.2%** **88.5%** **92.2%**

The PSO optimization method was employed in this study, which greatly improved the model performance. VGG19 accuracy jumping from 64.2% to 98.8%, an increase of 34.6%

ResNet50, for example, had only a baseline of 56.2%, but an improved score of 89.2% on the dataset, indicating how much of a difference using PSO can help boost even weaker models. MobileNetV2 + PSO (92.2%) performed on par with much deeper networks, including EfficientNet B3 (92%), establishing that the lightweight MobileNet models can be made competitive with more complex architectures when combined with PSO. Previous works including [26] report the best results of 5-class KOA classification at accuracy of 69.74% indicating that the task while possible is challenging without advanced methods to optimize performance. The VGG19 + PSO (98.8%) optimized with ensemble techniques such as DenseNet161 (97%) performing slightly better which indicates that for achieving state of the art performance is not a significant need to use heavy models in multiple numbers, since the PSO can give better results. Finally, the PSO algorithm was shown to be effective in obtaining a better categorization accuracy considering all five

classes of KOA by balancing the results for each of them specifically and being capable of strength from its parallelized optimization.

5. Limitations, Challenges and Future Work

Despite the promising results achieved by the proposed framework, several limitations should be acknowledged. The experiments were conducted using a single publicly available dataset, which may limit generalizability to other clinical settings. In addition, external validation using independent datasets is still required. Some deep architectures, particularly VGG19, may require higher computational resources compared with lightweight models.

The future research work will concentrate on improving the framework in a number of ways. First, it is better to increase the data base with pictures of various hospitals and a wide range of patients to enhance the generalizability of the model and minimize the possible bias. Second, the incorporation of lightweight architectures and model compression methods will reduce computational expenses and the system will be more affordable to clinics that do not have ample hardware. Third, even more advanced architectures, including Vision Transformers and hybrid CNNs-transformer models, could be explored in order to enhance performance and interpretability. Lastly, there will be real world clinical validation studies performed in order to test the usability, reliability, and efficacy of the model in real healthcare settings.

6. Conclusions

This paper employed a hybrid of the swarm intelligence-based optimization and the deep learning feature selection to develop an effective framework to enhance the detection of KAOs. The study studied four popular CNNs models, namely, VGG19, ResNet50, MobileNetV2, and EfficientNetB0 as feature extractors. PSO was also applied to optimizing the hyperparameters of MLP classifiers. The PSO enabled automatic tuning of important parameters like the number of neurons, dropout rates, learning rates and number of epochs at the push of a button as opposed to manually testing whatever combination was working with your data. The VGG19 + PSO model achieved outstanding performance with 98.8% test accuracy across all evaluated configurations. This strong performance can be attributed to the rich feature extraction capability of VGG19, effective PSO-based hyperparameter optimization, balanced training data, and robust preprocessing procedures. Confusion matrices, loss trends and visualizations that were used to determine training accuracy confirmed such reliability and consistency of the improved model. Moreover, the training and validation curves showed stable convergence without clear signs of overfitting, while final evaluation on an independent unseen test set supported the robustness of the model. The conclusions were drawn based on consistent experimental evidence obtained from multiple deep learning architectures under identical training and evaluation settings. This paper has shown that there are significant advantages in combining deep feature derivation with smart optimization in medical image classification projects. The suggested system boosts prediction accuracy and has a scalable and consistent method of automated KOA grading. These innovations can aid in early detection and help in clinical decision-making in the end improving patient outcomes.

Author Contributions: Dilan Jameel Sulaiman contributed to the conceptualization, methodology, software development, data analysis, visualization, and manuscript writing. Baraa Salim contributed to supervision, validation, review, and editing of the manuscript. All authors have read and approved the final version of the manuscript.

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank Duhok Polytechnic University for providing academic support and research facilities for this study.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	**Artificial Intelligence**
CNNs	**Convolutional Neural Networks**
DL	**Deep Learning**
Grad-CAM	**Gradient-weighted Class Activation Mapping**

KL	**Kellgren-Lawrence**
KOA	**Knee Osteoarthritis**
MLP	**Multilayer Perceptron**
MRI	**Magnetic Resonance Imaging**
OAI	**Osteoarthritis Initiative**
PSO	**Particle Swarm Optimization**
ROI	**Region of Interest**
XAI	**Explainable Artificial Intelligence**

References

1. R. Ahmed and A. Shariq Imran, "Knee Osteoarthritis Analysis Using Deep Learning and XAI on X-rays," IEEE Access, 2024, doi: 10.1109/ACCESS.2017.DOI.
2. X. Wang, S. Liu, and C. Zhou, "Classification of Knee Osteoarthritis Based on Transfer Learning Model and Magnetic Resonance Images," in Proceedings - 2022 International Conference on Machine Learning, Control, and Robotics, MLCR 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 67–71. doi: 10.1109/MLCR57210.2022.00021.
3. D. Shen, G. Wu, and H. Il Suk, "Deep Learning in Medical Image Analysis," Annu Rev Biomed Eng, vol. 19, pp. 221–248, Jun. 2017, doi: 10.1146/annurev-bioeng-071516-044442.
4. J. Qadir, "Enhancing Skin Disease Diagnosis: A Hybrid Approach Combining Vision Transformer and Feature Selection Techniques.," Zanin Journal of Science and Engineering, vol. 1, no. 1, pp. 54–71, Mar. 2025, doi: 10.64362/zjse.37.
5. M. Jahan et al., "KOA-CCTNet: An Enhanced Knee Osteoarthritis Grade Assessment Framework Using Modified Compact Convolutional Transformer Model," IEEE Access, vol. 12, pp. 107719–107741, 2024, doi: 10.1109/ACCESS.2024.3435572.
6. M. D. Fall, "Quantifying Uncertainty in Knee Osteoarthritis Diagnosis," in Proceedings - International Symposium on Biomedical Imaging, IEEE Computer Society, 2024. doi: 10.1109/ISBI56570.2024.10635586.
7. A. Khalid, E. M. Senan, K. Al-Wagih, M. M. Ali Al-Azzam, and Z. M. Alkhraisha, "Hybrid Techniques of X-ray Analysis to Predict Knee Osteoarthritis Grades Based on Fusion Features of CNN and Handcrafted," Diagnostics, vol. 13, no. 9, May 2023, doi: 10.3390/diagnostics13091609.
8. D. Sarkar, T. Khan, and F. Ahmed Talukdar, "Hyperparameters optimization of neural network using improved particle swarm optimization for modeling of electromagnetic inverse problems," Int J Microw Wirel Technol, vol. 14, no. 10, pp. 1326–1337, Dec. 2022, doi: 10.1017/S1759078721001690.
9. J. Wang, Z. Lei, X. Chang, and D. Huang, "IPSO-CNN: Malicious Code Classification with Improved PSO Optimized CNN," in 2023 5th International Conference on Frontiers Technology of Information and Computer, ICFTIC 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 906–910. doi: 10.1109/ICFTIC59930.2023.10455878.
10. A. Qadir, R. Mahum, and S. Aladhadh, "A Robust Approach for Detection and Classification of KOA Based on BILSTM Network," Computer Systems Science and Engineering, vol. 47, no. 2, pp. 1365–1384, 2023, doi: 10.32604/csse.2023.037033.
11. T. Momenpour and A. Abu Mallouh, "Optimizing CNN-Based Diagnosis of Knee Osteoarthritis: Enhancing Model Accuracy with CleanLab Relabeling," Diagnostics, vol. 15, no. 11, p. 1332, May 2025, doi: 10.3390/diagnostics15111332.
12. Y. Wang, S. Li, B. Zhao, J. Zhang, Y. Yang, and B. Li, "A ResNet-based approach for accurate radiographic diagnosis of knee osteoarthritis," CAAI Trans Intell Technol, vol. 7, no. 3, pp. 512–521, Sep. 2022, doi: 10.1049/cit2.12079.
13. A. S. Mohammed, A. A. Hasanaath, G. Latif, and A. Bashar, "Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images," Diagnostics, vol. 13, no. 8, Apr. 2023, doi: 10.3390/diagnostics13081380.
14. B. Bugday, H. Bingol, M. Yildirim, and B. Alatas, "Enhancing knee osteoarthritis detection with AI, image denoising, and optimized classification methods and the importance of physical therapy methods," PeerJ Comput Sci, Feb. 2025, doi: 10.7717/peerj.
15. P. S. Q. Yeoh, K. W. Lai, S. L. Goh, K. Hasikin, X. Wu, and P. Li, "Transfer learning-assisted 3D deep learning models for knee osteoarthritis detection: Data from the osteoarthritis initiative," Front Bioeng Biotechnol, vol. 11, 2023, doi: 10.3389/fbioe.2023.1164655.
16. S. Rani et al., "Deep learning to combat knee osteoarthritis and severity assessment by using CNN-based classification," BMC Musculoskelet Disord, vol. 25, no. 1, Dec. 2024, doi: 10.1186/s12891-024-07942-9.

17. H. Harish, A. Patrot, S. Bhavan, S. Gousiya, and A. Livitha, "Knee Osteoarthritis Prediction Using Deep Learning," in 2023 International Conference on Recent Advances in Information Technology for Sustainable Development, ICRAIS 2023 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 14–19. doi: 10.1109/ICRAIS59684.2023.10367065.
18. T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, "Particle Swarm Optimization: A Comprehensive Survey," IEEE Access, vol. 10, pp. 10031–10061, 2022, doi: 10.1109/ACCESS.2022.3142859.
19. T. Li, H. Luo, and C. Wu, "A PSO-based fine-tuning algorithm for CNN," in Proceedings of 2021 5th Asian Conference on Artificial Intelligence Technology, ACAIT 2021, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 704–709. doi: 10.1109/ACAIT53529.2021.9731225.
20. S. U. Rehman and V. Gruhn, "A Sequential VGG16+CNN-Based Automated Approach With Adaptive Input for Efficient Detection of Knee Osteoarthritis Stages," IEEE Access, vol. 12, pp. 62407–62415, 2024, doi: 10.1109/ACCESS.2024.3395062.
21. G. Harish Kumar and K. Jaisharma, "Enhancing the Handwritten Digit Recognition by Employing Novel Progressive VGG19 Model and Compare with SVM Performance," in 2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICCCNT61001.2024.10724278.
22. M. I. F. Jamil, R. Samad, D. Pebrianti, M. Mustafa, N. R. H. Abdullah, and N. H. Noordin, "A Comparative Study of Deep Learning Models for the Classification of Knee Osteoarthritis in X-Ray Images," Institute of Electrical and Electronics Engineers (IEEE), Sep. 2024, pp. 228–233. doi: 10.1109/icom61675.2024.10652557.
23. A. Haseeb et al., "Knee Osteoarthritis Classification Using X-Ray Images Based on Optimal Deep Neural Network," Computer Systems Science and Engineering, vol. 47, no. 2, pp. 2397–2415, 2023, doi: 10.32604/csse.2023.040529.
24. A. Asnidar et al., "Application of MobileNetV2 Architecture to Classification of Knee Osteoarthritis Based on X-ray Images," in 2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation, ICAMIMIA 2023 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 375–380. doi: 10.1109/ICAMIMIA60881.2023.10427581.
25. R. Singh, N. Sharma, D. Upadhyay, S. Devliyal, and R. Gupta, "A Fine-Tuned EfficientNet B5 Transfer Learning Model for the Classification of Knee Osteoarthritis," in 2023 3rd International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/SMARTGENCON60755.2023.10442712.
26. A. Pandey and V. Kumar, "Enhancing Knee Osteoarthritis Severity Classification using Improved Efficientnet," in 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1351–1356. doi: 10.1109/UPCON59197.2023.10434740.
27. R. Singh, N. Sharma, R. Chauhan, D. Rawat, and R. Gupta, "Knee Osteoarthritis Classification Using EfficientNet B3 Transfer Learning Model," in 2023 2nd International Conference on Futuristic Technologies, INCOFT 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/INCOFT60753.2023.10425390.
28. T. Tariq, Z. Suhail, and Z. Nawaz, "Knee Osteoarthritis Detection and Classification Using X-Rays," IEEE Access, vol. 11, pp. 48292–48303, 2023, doi: 10.1109/ACCESS.2023.3276810.
29. Y. X. Teoh, A. Othmani, S. L. Goh, J. Usman, and K. W. Lai, "Predicting Knee Osteoarthritis Pain Severity through A Deep Hybrid Learning Model: Data from the Osteoarthritis Initiative," in Proceedings - 2023 2023 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 4148–4153. doi: 10.1109/BIBM58861.2023.10385415.
30. A. Marimuthu, A. R. Kavitha, and S. S. Abdullah, "Minimal Knee Joint Space Width Detection in Digital X-Ray Images using Deep Learning," in 2023 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICDSAAI59313.2023.10452517.
31. V. Bhateja et al., "Ensemble CNN Model for Computer-Aided Knee Osteoarthritis Diagnosis," International Journal of Service Science, Management, Engineering, and Technology, vol. 15, no. 1, 2024, doi: 10.4018/IJSSMET.349913.
32. M. D. Fall, "Quantifying Uncertainty in Knee Osteoarthritis Diagnosis," in Proceedings - International Symposium on Biomedical Imaging, IEEE Computer Society, 2024. doi: 10.1109/ISBI56570.2024.10635586.