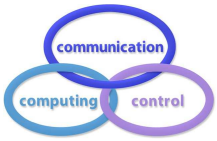
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 19, Issue: 5, Month: October, Year: 2024
Article Number: 6648, https://doi.org/10.15837/ijccc.2024.5.6648



CCC Publications

Acute Knee Injury Detection with Magnetic Resonance Imaging (MRI)

M. Mahmood, K. Alsalem, M. Elbashir, S. Abd El-Ghany, A. Abd El-Aziz

Mahmood A. Mahmood*
Department of Information Systems, College of Computer and Information Sciences
Jouf University, KSA
Skakah, KSA
*Corresponding author: mamahmood@ju.edu.sa

Khalaf Alsalem
Department of Information Systems, College of Computer and Information Sciences
Jouf University, KSA
Skakah, KSA
kosalem@ju.edu.sa

Murtada K. Elbashir
Department of Information Systems, College of Computer and Information Sciences
Jouf University, KSA
Skakah, KSA
mkelfaki@ju.edu.sa

Sameh Abd El-Ghany
Department of Information Systems, College of Computer and Information Sciences
Jouf University, KSA
Skakah, KSA
saabdelwahab@ju.edu.sa

A.A. Abd El-Aziz
Department of Information Systems, College of Computer and Information Sciences
Jouf University, KSA
Skakah, KSA
aaeldamarany@ju.edu.sa

Abstract

The anterior cruciate ligament (ACL) is a major ligament in the knee that helps to stabilize the joint and prevent excessive forward movement of the shinbone. An ACL tear is a common injury, especially among athletes who participate in sports that involve pivoting and sudden changes in direction. This paper proposes an ensemble model, which includes three deep learning models (EfficientNet-B7, ResNet-152V2, and DenseNet-201) and a genetic algorithm, to detect and classify

ACL tears using knee magnetic resonance imaging (MRI). The ensemble model was trained on the KneeMRI dataset, which comprises labeled MRI images. The deep learning models can learn to identify subtle changes in ligament structure and signal intensity that are associated with ACL tears and the genetic algorithm is used to find the optimal prediction. The proposed ensemble model was evaluated using the KneeMRI dataset. The dataset was preprocessed using data augmentation techniques. Then, the ensemble model was applied to the KneeMRI dataset, evaluated, and compared with previous models. The accuracy, recall, precision, specificity, and F1 score of our proposed ensemble model were 99.68%, 98.73%, 99.52%, 99.62%, and 98.94%, respectively. Thus, our ensemble model has an unrivaled perceptive outcome and could be used to accurately identify and classify ACL tears, improving patient outcome.

Keywords: acute knee injury, deep learning, classification, detection acute knee injury, magnetic resonance imaging.

1 Introduction

The knee is one of the largest and most complex joints in the human body. It is a pivotal joint that allows movement and stability in the lower limb. The knee joint consists of three major bones: the femur (thigh bone), tibia (shin bone), and patella (kneecap). The ends of these bones have smooth, articular surfaces covered in cartilage that allow them to move against each other with minimal friction [1]. Knee injuries are common and can result from a variety of causes, including sports activities, accidents, overuse, and degenerative conditions [2]. Meniscus tears are among the most commonly diagnosed knee injuries. The menisci are C-shaped cartilage structures located in the knee joint between the femur (thigh bone) and the tibia (shin bone). They act as shock absorbers, help distribute weight across the knee joint, and provide stability during movement. Meniscus tears can be caused by sudden twisting or pivoting movements, sports-related activities, and degenerative changes over time. The rate of meniscus tears is estimated to be 60–70 for every 100,000 people each year, with a high occurrence among individuals of all ages [3]. Additionally, in the USA, meniscus tears are the most common type of intra-articular knee injury. These tears can result in a range of symptoms, such as pain, swelling, stiffness, and a limited range of motion. In some cases, individuals may experience a locking or catching sensation in the knee. The diagnosis of a meniscus tear typically involves a combination of clinical evaluation, patient history, and imaging studies, such as magnetic resonance imaging (MRI) scans and X-rays [4, 5].

Knee osteoarthritis (OA) is a common joint condition caused by the wearing of the articular surface between the knee joints. The incidence of knee OA can vary depending on various factors, including age, genetics, lifestyle, and other health-related factors. While the exact statistics vary by region and population, it is clear that knee OA is a significant health concern. It is often considered one of the most prevalent forms of arthritis. There is a high likelihood of developing knee OA in the general population, with one in three people developing the condition [6, 7].

Knee OA is characterized by gradual degeneration of the knee joint, leading to symptoms such as pain, stiffness, and reduced mobility. It can affect people of all ages but is more common in older people, with a greater proportion of individuals aged 65 years and older having indications of knee OA. Knee OA affects more than 250,000 people overall and ranks among the 50 most common illnesses [8]. It is the most common joint illness in the USA, occurring in 13% of women and 10% of men older than 60 years. More than one-quarter of Americans will reach the age of 65 by 2030, increasing their chances of developing OA [9]. Risk factors for knee OA include aging, obesity, joint injuries, genetics, and lifestyle factors. Knee OA affects the life satisfaction of advanced-aged individuals [10].

Anterior cruciate ligament (ACL) tears are a common and significant knee injury. ACL injuries account for more than half of knee injuries, affecting 200,000 people in the USA every year. ACL injuries are a significant healthcare concern in the USA and globally. The financial cost associated with ACL injuries is more than 7 billion dollars and includes various components, such as medical expenses, surgical procedures, rehabilitation, and indirect costs related to lost productivity and workdays [11]. When athletes engage in high-impact occupational and recreational activities, especially those involving repetitive stress on the knees, ACL injuries increase [12].

The ACL is one of the four major ligaments in the knee and plays a crucial role in stabilizing

the joint by preventing excessive forward movement of the tibia (shinbone) relative to the femur (thighbone). An ACL tear can occur due to various factors, such as the following:

- **Trauma**: A sudden, forceful impact or twisting motion of the knee can result in an ACL tear. This often occurs during sports activities, such as football, basketball, soccer, skiing, or gymnastics.
- **Noncontact Injuries**: ACL tears can also occur without direct contact with another person or object. These noncontact injuries typically involve rapid changes in direction, deceleration, or pivoting.
-

Individuals who sustain an ACL tear often experience a popping sensation in the knee at the time of injury. Other common symptoms include pain, swelling, instability, and difficulty bearing weight on the affected leg. Healthcare professionals typically diagnose ACL tears through a combination of clinical evaluation, imaging studies such as MRI, and a review of the patient's medical history. ACL outcomes can be classified into three categories: Grade 0 (healthy), Grade I (partial tear), and Grade II (complete tear), as shown in Fig. 1. The treatment approach for ACL tears depends on the grade of the tear, the patient's activity level, and the goal of knee function. Treatment options include [13]:

- **Non-Surgical Options**: Some individuals with partial ACL tears or limited activity requirements may not require surgery. Rehabilitation and physical therapy are often sufficient.
- **Surgical Reconstruction**: For complete tears or individuals who have high activity demands (e.g., athletes), surgical reconstruction of the ACL is a common approach. This procedure involves replacing the torn ACL with a graft (typically from the patient's own tissues or a donor) to restore stability and function.
- **Recovery**: Recovery from an ACL tear and surgery can take several months, and a structured rehabilitation program is needed to regain strength, stability, and range of motion in the knee. The return to sports or high-impact activities should be gradual to minimize the risk of re-injury.
- **Prevention**: Many sports organizations and athletes use specific training programs and techniques to reduce the risk of ACL tear injuries, particularly in high-risk sports.

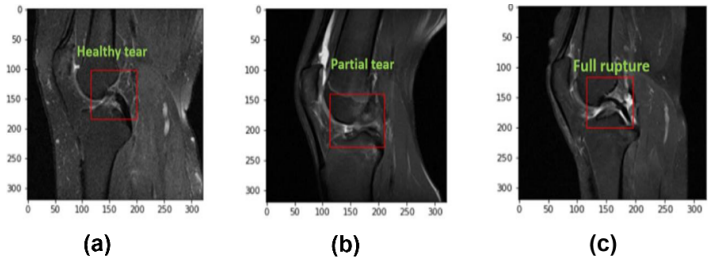


Figure 1: (a) A healthy ACL, (b) a partial tear, and (c) a full tear.

The health and proper function of the knee joint are critical for an individual's overall mobility and quality of life. An ACL tear has a significant impact on a person's daily activities and well-being. ACL tears are significant injuries with the potential for long-term consequences if not treated appropriately. Given the crucial role of the knee in mobility and daily life, timely and appropriate medical care is important when patients experience ACL tears. Proper diagnosis of ACL tears, treatment, and rehabilitation are essential for achieving the best possible outcome and returning to normal activities.

Knee MRI has a high negative predictive value for injuries such as ACL tears and meniscus tears. This means that when knee MRI shows no evidence of these abnormalities, it is highly reliable for ruling out the presence of such conditions. Hence, MRI can help prevent unnecessary surgical interventions. MRI is cost-effective for patients; however, interpreting knee MRI results requires

a high level of expertise, often by specialized radiologists or orthopedic surgeons who are trained to recognize the nuances of the images. They must differentiate between normal anatomical variations and pathological findings. Therefore, MRI interpretation is time-consuming and has inherent challenges. Moreover, when radiologists must deal with many MRI scans in a short period of time, the risk of errors in interpretation increases. Kim and Mansfield highlighted that radiologists will generally make blunders in 30% of situations when assessing X-rays, and 66% of these blunders are of an outer muscle nature [14].

Computer-aided (CAD) systems can assist radiologists in diagnosing ACL tears, thus reducing their workload and potential for error. Artificial intelligence (AI) has become an increasingly prominent and dynamic field within the realm of medical research and healthcare. AI can analyze medical images, such as X-rays, MRI, and computerized tomography (CT) scans, to assist in the early and accurate diagnosis of diseases, including ACL tears, cancer, heart conditions, and neurological disorders. AI has been particularly impactful in radiology, where it can detect and highlight abnormalities in medical images with high accuracy.

Deep learning (DL) is a subset of machine learning (ML), which, in turn, is a branch of AI [15]. DL models are designed to mimic the way the human brain processes and understands data. These models use artificial neural networks (ANNs) that consist of multiple layers (hence the term "deep") to learn and make predictions or decisions from large volumes of data. DL has made substantial contributions to the field of healthcare and medicine, revolutionizing various aspects of patient care, medical research, and healthcare management. DL models, especially convolutional neural networks (CNNs), have demonstrated exceptional performance in medical image analysis, detecting and diagnosing diseases from medical images, and identifying and delineating regions of interest in images, such as ACL tears, tumors, blood vessels, or organs [16, 17, 18].

Early detection of knee injuries, especially ACL tears, using DL is an important and promising area in healthcare. DL has the potential to revolutionize medical image analysis and improve the accuracy and efficiency of knee injury diagnosis. In this paper, we propose a fine-tuned ensemble model, which includes three DL algorithms, EfficientNet-B7, ResNet152V2, and DenseNet201, and a genetic algorithm (GA). The proposed ensemble model will help in the identification and classification of ACL tears. The GA was used to obtain the optimal epoch solution that has maximum val_accuracy and minimum loss (min_loss) from the pre-trained learning models. The ImageNet dataset was used to pre-train the three models. The three models utilized the global max pooling (GAP) 2D (GlobalMaxPooling2D) layer to reduce the spatial dimensions of the feature maps produced by convolutional layers while retaining essential information. For each feature map or input channel, GlobalMaxPooling2D calculates the maximum value across all the spatial locations. This operation results in a single value for each channel, effectively summarizing the information in that channel. Hence, we have a set of maximum values, one for each channel, which collectively form the feature vector representing the input data. Finally, we fine-tuned the resulting feature vector with a training dataset from the KneeMRI dataset, which was published online by Clinical Hospital Centre Rijeka, Croatia. The KneeMRI dataset was preprocessed by applying resizing, rescaling, flipping, rotation, zooming, and contrasting techniques.

The proposed ensemble model aims to alleviate the workload and responsibilities of primary radiologists. In a medical setting, radiologists play a crucial role in interpreting medical images, such as MRI scans. However, the increasing volume of medical images and the demand for timely diagnosis can place a significant burden on radiologists. The proposed ensemble model provides solutions that can assist radiologists in their tasks, reducing their workload and potential for error. The ensemble model achieved an accuracy, precision, recall, and F1-score of 99.68%, 99.52%, 98.73%, and 98.94%, respectively. The research contributions are summarized as follows:

- We proposed a robust ensemble model for the early diagnosis of ACL tears via MRI.
- The proposed ensemble model determines the grade of the ACL tear more accurately than the previous techniques.
- The proposed ensemble model achieves a high level of accuracy in diagnosing ACL tears. This finding implies that it can be used to effectively distinguish between patients with and without

ACL tears, reducing the likelihood of misdiagnosis.

- The proposed model offers fast diagnosis. This speed can be crucial in the healthcare setting, where rapid diagnosis enables healthcare providers to initiate treatment and care plans sooner.
- We used a GA to obtain the optimal prediction from the pre-trained learning models.
- The KneeMRI dataset was preprocessed by applying resizing, rescaling, flipping, rotation, zooming, and contrasting techniques.
- The proposed ensemble model achieved an accuracy, precision, recall, and F1-score of 99.68%, 99.52%, 98.73%, and 98.94%, respectively.

This research is structured as follows: a literature review is presented in Section 2. In Section 3, the methodology and materials of the proposed ensemble DL model are described. The ensemble model implementation and its evaluation are presented in Section 4. Section 5 discusses the results. Section 6 presents the conclusion.

2 Related Work

Bien et al. [19] proposed a DL model called MRNet for diagnosing ACL tears and meniscal tears from knee MRI scans. MRNet uses a CNN architecture designed for the specific task of diagnosing knee injuries using 3D MRI. MRNet is designed to map 3D MRI images to 2D images for the purpose of classification, enabling the identification of various knee conditions and pathologies. The knee MRI dataset was preprocessed by scaling the images to 256 × 256 pixels and converting them to PNG format. The accuracy and area under the curve (AUC) of the proposed model for diagnosing ACL tears were 86.7% and 96.5%, respectively.

Azcona et al. [20] compared DL models based on transfer learning and DL models trained from scratch to detect meniscus tears and ACL tears from MRI. The authors used ResNet-18 and the KneeMRI dataset. The fine-tuned DL model based on transfer learning with data augmentation achieved the best performance. The fine-tuned DL model based on transfer learning for detecting ACL tears achieved an AUC of 96%.

Chang et al. [21] compared three CNN models with different dimensionalities and views for diagnosing ACL tears. The input views included dynamic patch-based sampling, cropped slices, and full slices. The dimensions were single, three, or five slices. The authors used the KneeMRI dataset. The images in the dataset were rescaled to 256 × 256 voxels and normalized using the z score algorithm. The model with dynamic patch-based sampling achieved the best performance, with an AUC of 97% and an accuracy of 96.7%.

Liu et al. [22] proposed a DL system for detecting ACL tears. The proposed system used LeNet-5 and YOLO, which are two CNNs, to isolate ACL tears from MRI. The DenseNet CNN was used to classify abnormalities in isolated ACL tears. The KneeMRI dataset was used. The specificity and sensitivity of the proposed system were 96%. The sensitivity of the radiologists ranged from 96% to 98%, and the specificity ranged from 90% to 98%. There was no statistically significant difference in analytic execution between the proposed system and clinical radiologists. The AUC of the receiver operating characteristic (ROC) curve for the proposed system was 98%, which means that the proposed system achieved good performance.

Namiri et al. [23] used two CNN models to detect ACL tears. The first CNN model had 3D kernels, and the second had 2D kernels. The authors used 1243 MRI images from 224 patients. The 3D-CNN model achieved an accuracy of 89% and a weighted Cohen k of 0.83. The 2D-CNN model achieved an accuracy of 92% and a weighted Cohen k of 0.83 for intact ACL tears, the 3D model achieved a sensitivity of 89% and a specificity of 88%, and the 2D model achieved a sensitivity of 93% and a specificity of 90%. For full ACL tears, the 3D model achieved a sensitivity of 76% and a specificity of 100%, and the 2D model achieved a sensitivity of 82% and a specificity of 94%. All the reconstructed ACL tears were correctly classified by the 2D CNN model.

Table 1: Class distribution of the KneeMRI dataset

Class	Count
Healthy	690
Partial tear	172
Complete tear	55

Zhang et al. [24] proposed a DL model based on the 3D DenseNet model to classify ACL tears. The authors used the KneeMRI dataset. The dataset was split into a training set (70%), a test set (10%), and a validation set (20%). A comparison among 3D DenseNet, VGG16 and ResNet revealed that 3D DenseNet achieved the highest accuracy (95.7%).

Germann et al. [25] used a deep convolutional neural network (DCNN) for the diagnosis of ACL tears. The DCNN was trained on knee MRI. The MRI images were preprocessed by rescaling and cropping sagittal fluid-sensitive and coronal views. The sensitivity for clinical use was 97.5–97.9%, where all p values were ≥ 0.118, and the sensitivity was 96.1% for DCNNs. The sensitivity of the DCNN was 93.1%, and it was 99.6–100% for clinical use, where all p values were < 0.001.

Awan et al. [26] proposed a DL model based on the ResNet-14 model for the detection of ACL tears. The ResNet CNN model had 14 layers and was combined with six different directions by utilizing data augmentation and class balancing. A total of 917 sagittal MRI images of knees were used. The MRI images were in grayscale. The MRI dataset was normalized, the regions of interest were labeled, and a hybrid approach with undersampling and oversampling was applied. By using 5-fold cross-validation, the proposed model achieved a sensitivity, accuracy, precision, specificity, and F1-score of 91.6%, 92%, 91.6%, 94.6%, and 92.3%, respectively. Moreover, the proposed model achieved AUCs of 99.9%, 97%, and 98% for full ACL tears, partial tears, and healthy patients, respectively.

Jeon et al. [27] proposed a 3D DL model for classifying ACL tears. The KneeMRI dataset was used. The authors assumed that the features required for detecting ACL tears are highly homogeneous and locally confined. The homogenous features were presented using fewer convolutional filters and squeeze modules. The search for local features was implemented using Gaussian positional encoding and attention modules. The proposed model achieved ROC curves of 98% and 98.3% on the KneeMRI and Chiba datasets, respectively.

Astuto et al. [28] proposed 3D CNNs for diagnosing regions of interest within MRI, including ACL tears, bone marrow, menisci, and cartilage grade abnormalities. The KneeMRI dataset was used. The sensitivity of the binary lesion for all tissues was 88%, the specificity was 89%, and the AUC was 90%.

3 Materials and Methods

3.1 Material Description

The KneeMRI dataset was assembled reflectively from test records made on a Siemens Avanto 1.5T MR scanner and obtained by the proton thickness weighted fat concealment method at the Clinical Medical Clinic Community Rijeka, Croatia, from 2006 to 2014. The KneeMRI dataset comprises 917 images of left or right knees, where each image is in a 12-bit grayscale. The KneeMRI dataset has three classes: healthy (intact ACL), partial ACL tear, and complete ACL tear as shown in table 1. The regions of interest were determined by rectangles, extracted from the original images manually, and annotated [29]. Ensemble models are beneficial because they leverage the diversity of different models, combining their strengths to improve overall performance and generalizability. They are particularly effective when individual models have complementary strengths and weaknesses. This approach can make the ensemble more robust to variations in input data and more accurate in its predictions. Therefore, to diagnose ACL tears early using MRI, we propose a fine-tuned ensemble CNN model utilizing EfficientNet-B7, ResNet152V2, and DenseNet201, which are DL models. In the proposed ensemble, we used a GA to obtain the optimal prediction solution that has maximum val_accuracy and minimum loss (min_loss) from ensemble pre-trained learning models. The three models utilized the global max pooling (GAP) 2D (GlobalMaxPooling2D) layer to reduce the spatial dimensions of the feature maps produced by convolutional layers while retaining essential information.

For each feature map or input channel, GlobalMaxPooling2D calculates the maximum value across all the spatial locations. This operation results in a single value for each channel, effectively summarizing the information in that channel. Hence, we have a set of maximum values, one for each channel, which collectively form the feature vector representing the input data. Finally, we fine-tuned the resulting feature vector over the training set of the KneeMRI dataset. Figure 2 depicts the architecture of the proposed fine-tuned ensemble model. The steps of the proposed ensemble model are as follows:

Phase 1 : KneeMRI dataset preprocessing: The KneeMRI dataset [29] was preprocessed by applying resizing, rescaling, flipping, rotation, zooming, and contrasting techniques.

Phase 2 : KneeMRI dataset splitting: The preprocessed KneeMRI dataset was split into a training set (70%), a test set (15%), and a validation set (15%).

Phase 3 : Pre-training for the EfficientNet-B7, ResNet152V2, and DenseNet201 models: Through the supervised pre-training phase of the transfer learning process, the models used were pre-trained on ImageNet. Moreover, the three pre-trained models utilized the GlobalMaxPooling2D layer to reduce the spatial dimensions of the feature maps produced by convolutional layers while retaining essential information.

Phase 4 : Integrating the pre-trained individual feature vectors: For each feature map or input channel, GlobalMaxPooling2D calculates the maximum value across all the spatial locations. This operation results in a single value for each channel.

Phase 5 : Fine-tuning the resulting feature vector: We fine-tuned the resulting feature vector and trained it on the training set and validation set of the KneeMRI dataset.

Phase 6 : Measuring the performance: The fine-tuned proposed ensemble model was evaluated on the test set of the KneeMRI dataset using evaluation metrics, such as recall, precision, accuracy, specificity, and F1-score.

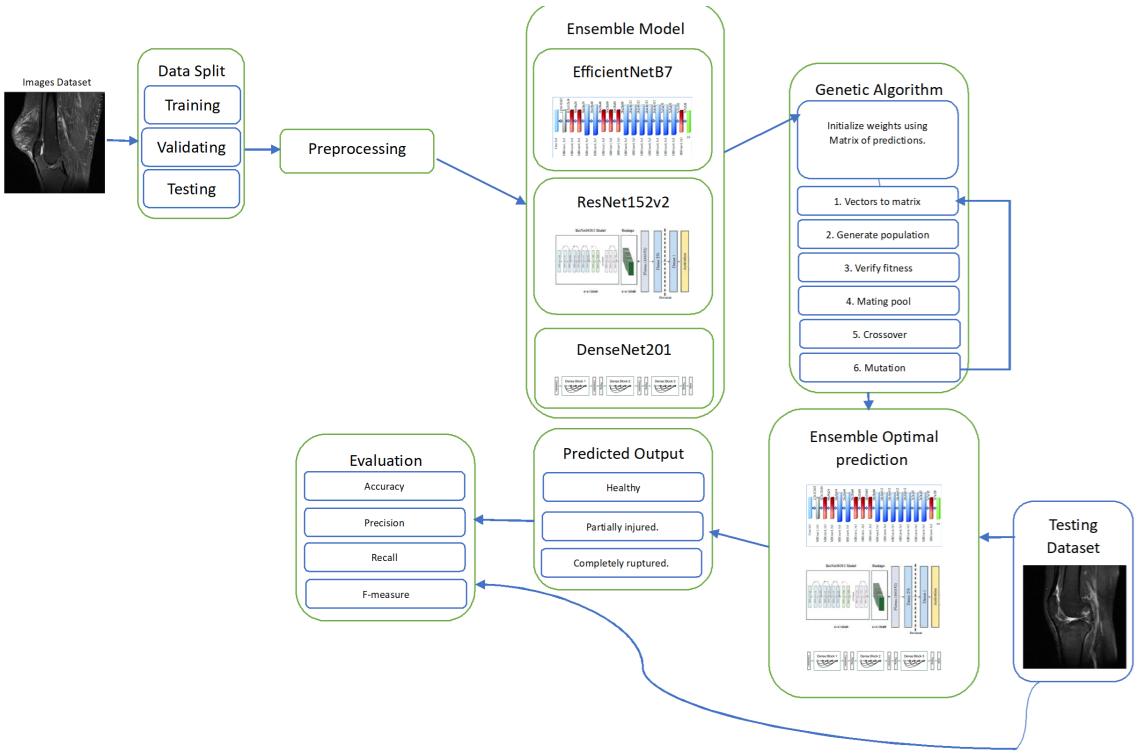


Figure 2: Architecture of the proposed ensemble model.

3.2 Dataset preprocessing

Image preprocessing is a critical step in image analysis and computer vision applications. It involves a series of techniques and operations applied to raw or captured images to prepare them for further analysis or interpretation. The importance of image preprocessing lies in its ability to enhance the quality of images, improve feature extraction, and make subsequent tasks more accurate and efficient. In our proposed ensemble, we applied a resizing method to resize the images to a size of 224×224, a rescaling method to rescale the values of the images to intervals of [0,1], a flipping method to flip the images horizontally, a rotation of 0.1, a zooming in on the images of 0.1, and a contrast of 0.1.

3.3 EfficientNet-B7

EfficientNet-B7 is a deep neural network (DNN) with many layers, making it capable of capturing complex patterns in images. EfficientNet-B7 is a specific variant of the EfficientNet family of CNNs known for its efficiency and performance in image recognition tasks. EfficientNet was introduced in [30]. The "B7" in the name indicates the scaling factor used in designing this model. EfficientNet models are scaled in terms of depth, width, and resolution to balance performance and computational efficiency. A higher scale number (e.g., B7) typically indicates a larger and more powerful model. EfficientNet-B7 is suitable for a wide range of computer vision tasks, including image classification, object detection, and image segmentation. It can be used in applications such as autonomous driving and medical imaging [30].

4 ResNet-152V2

ResNet-152V2, or residual network-152 version 2, is a deep convolutional neural network with a 152-layer architecture that is part of the ResNet family. The depth of the network allows it to capture intricate features in images, making it suitable for a wide range of computer vision tasks. ResNet is known for its ability to train DNNs effectively using skip connections or residual connections, which help mitigate the vanishing gradient problem. Skip connections allow information to flow directly through the network without significant degradation. This enables the training of much deeper networks and helps prevent the vanishing gradient problem. ResNet-152V2 is an improved version of the original ResNet-152, offering better performance and accuracy [31].

4.1 DenseNet-201

DenseNet-201, short for Densely Connected Convolutional Network 201, is a popular CNN architecture used in computer vision tasks, particularly for image classification. It is part of the DenseNet family and is known for its densely connected layers and efficient use of parameters. DenseNet-201 is one of the variants within this family with 201 layers in the network. DenseNet-201 introduces densely connected blocks, where each layer is connected to every subsequent layer within a block. This dense connectivity allows for feature reuse, better gradient flow, and more efficient training. DenseNet-201 includes transition blocks that control the growth of feature maps as they progress through the network. Transition blocks typically consist of 1x1 convolutional layers and pooling layers to reduce the spatial dimensions.

4.2 Genetic Algorithm

A genetic algorithm (GA) is a type of evolutionary algorithm that is inspired by the process of natural selection. GAs are used to solve optimization problems by iteratively evolving a population of candidate solutions. Each candidate solution is represented as a chromosome, which is a string of bits. The chromosomes are evaluated using a fitness function, which assigns a score to each solution based on how well it satisfies the problem constraints. At each iteration, the GA selects the fittest solutions to be parents. The parents are then used to create offspring through crossover and mutation. Crossover is the process of combining the chromosomes of two parents to create a new offspring. Mutation is the process of randomly changing bits on a chromosome. The offspring are then evaluated

using the fitness function, and the fittest offspring are selected to be the parents of the next generation. This process is repeated until a satisfactory solution is found or until a maximum number of iterations is reached.

Main steps of a genetic algorithm
1. Initialize the weights using the prediction matrix.
2. Generate population.
3. Evaluate the fitness of everyone in the population.
4. Select the parents for the next generation. The parents have the highest fitness value.
5. Create offspring. This is done by combining the chromosomes of two parents using crossover and mutation.
6. Replace the current population with the offspring.
7. Repeat steps 3-6 until a satisfactory solution is found or until a maximum number of iterations is reached.

The GA prevents learning models' epochs from being selected as the optimization step in the proposed ensemble. The GA commences with a population of potential solutions, which, in this case, are the ensemble epochs transformed into a set of possible parameters. These solutions are then evaluated using a fitness function that aligns with the goal of achieving high validation accuracy (val_accuracy) and low loss (min_loss). The elite solutions are then used to generate a new population, replacing the previous one through crossover and mutation operators, and run for a sufficient number of generations, usually limited by a predefined maximum or fixed value if the fitness improvement rate is slow.

The stopping criteria for the DNN learning process are different from the stopping criteria based on the output of a four-dimensional vector as derived by GA but are embedded in the metrics evaluated by GA. In DNN training, validation accuracy and loss values are measured for each epoch to identify overfitting. Subsequently, the GA adapts the epoch selection depending on these values after the training phase is accomplished. The performance of the GA process depends more on the convergence of the epoch combinations; hence, the GA process halts with a combination that produces the best performance where there are no drastic improvements in the fitness scores or where a fixed number of generations have been reached.

5 Experimental Results and Discussion

5.1 Performance Metrics

The performance of the proposed model was assessed using the following metrics: specificity, accuracy, precision, recall, and F1-measure, which are defined in Equations (1)-(5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

$$Specificity = \frac{TN}{FP + TN} \tag{5}$$

False positives (FPs), false negatives (FNs), true negatives (TNs), and true positives (TPs) are all measures of how well a model performs. Accuracy is the proportion of predictions that are correct. Precision is the proportion of positive predictions that are actually correct. Recall is the proportion of actual positives that are correctly predicted. The F-measure is a measure of how well a model predicts both positive and negative cases.

5.2 Model Implementation

The KneeMRI dataset was divided: 70% for training, 15% for validation, and 15% for testing. The computer used for the experiment had a 64-bit operating system, an x64-based processor, an Intel(R) Core(TM) i7-3612QM CPU @ 2.10 GHz and 2.1 GHz, and 10 GB of RAM. The Python programming language was used to implement the model.

Figure 3 and Table 2 show the results for the EfficientNet-B7 deep learning model applied to the KneeMRI dataset. We distinguished between healthy, partially torn, and completely ruptured ACLs. Table 2 shows that the EfficientNet-B7 model had an accuracy, precision, recall, specificity, and F1-measure average of 70.77%, 69.94%, 74.73%, 66.67%, and 72.25%, respectively.

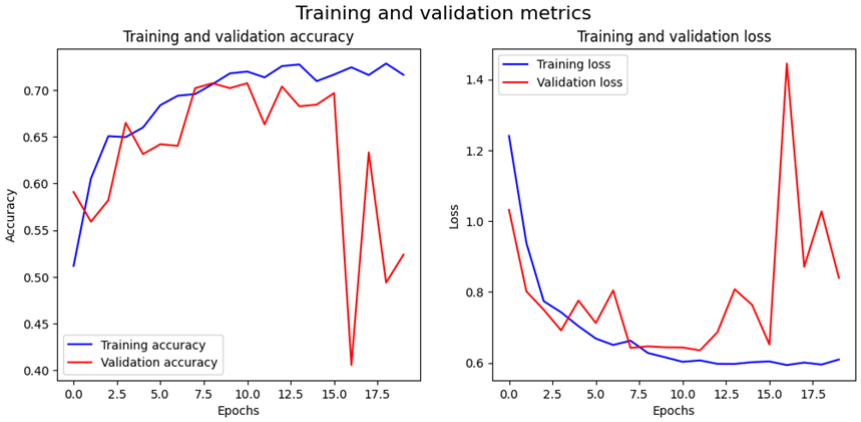


Figure 3: Accuracy and loss of the training and validation metrics for EffeicientNet-B7

Table 2: Performance of the EfficientNet-B7 model

	Class	Precision	Recall	F1-Measure	Specificity	Accuracy
EfficientNet- B7	Healthy	0.7007	0.7558	0.7287	0.6634	**0.7076**
	Partially torn	0.6946	0.7557	0.7219	0.6674	**0.7069**
	Completely ruptured	0.7028	0.7303	0.7173	0.6692	**0.7084**
Average		0.6994	0.7473	0.7226	0.6667	**0.7077**

Figure 4 and Table 3 show the results for the ResNet152V2 deep learning model applied to the KneeMRI dataset. We distinguished between healthy, partially torn, and completely ruptured ACLs. Table 3 shows that ResNet152V2 had an accuracy, precision, recall, specificity, and F1-measure average of 71.43%, 70.30%, 76.02%, 66.67%, and 73.05% respectively.

Figure 5 and Table 4 show the results for the DenseNet201 deep learning model applied to the KneeMRI dataset. We distinguished between healthy, partially torn, and completely ruptured ACLs.

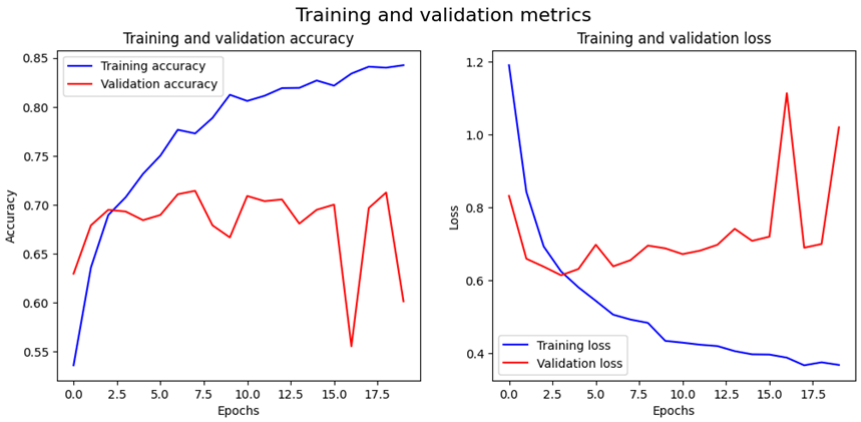


Figure 4: Accuracy and loss of training and validation metrics for ResNet152V2

Table 3: Performance of the ResNet152V2 model

	Class	Precision	Recall	F1-Measure	Specificity	Accuracy
ResNet152V2	Healthy	0.7043	0.7688	0.7367	0.6634	**0.7142**
	Partially torn	0.6982	0.7687	0.7298	0.6674	**0.7135**
	Completely ruptured	0.7064	0.7430	0.7251	0.6692	**0.7150**
Average		0.7030	0.7602	0.7305	0.6667	**0.7143**

Table 4: Performance of the DenseNet201 model

	Class	Precision	Recall	F1-Measure	Specificity	Accuracy
DenseNet201	Healthy	0.7186	0.7873	0.7529	0.6928	**0.7371**
	Partially torn	0.7123	0.7873	0.7458	0.6971	**0.7364**
	Completely ruptured	0.72070.7	0.7608	0.7411	0.6990	**0.7379**
Average		0.7172	0.7785	0.7466	0.6963	**0.7372**

Table 4 shows that DenseNet201 had an accuracy, precision, recall, specificity, and F1-measure average of 73.72%, 71.72%, 74.66%, 69..63%, and 74.66%, respectively.

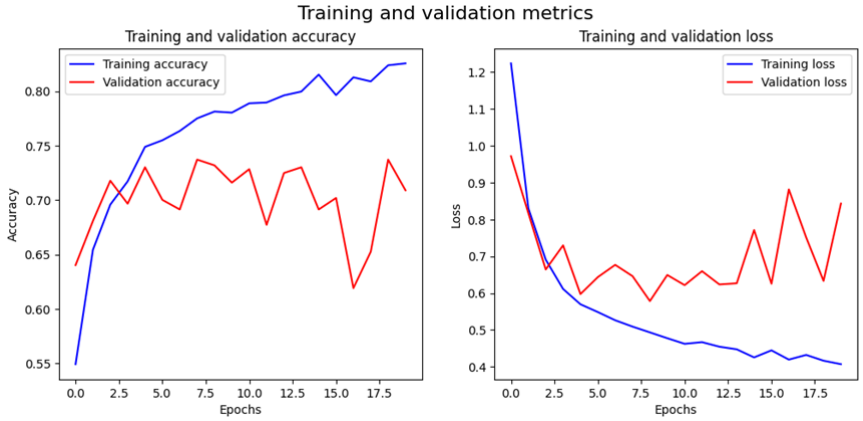


Figure 5: Accuracy and loss of training and validation metrics for DenseNet201

Figure 6 and Table 5 show the results for the ensemble deep learning model (EfficientNet-B7, ResNet-152V2, and DenseNet-201) applied to the KneeMRI dataset. We distinguished between healthy, partially torn, and completely ruptured ACLs. Table 5 shows that the ensemble had an accuracy, precision, recall, specificity, and F1-measure average of 99.68%, 99.52%, 98.73%, 99.62%, and 98.94%, respectively.

For the healthy class, the precision, recall, and F1-measure were 99.71%, 99.85%, and 99.78%, respectively. For the partially torn class, the precision, recall, and F1-measure were 98.84%, 99.84%, and 99.84%, respectively. For the completely ruptured class, the precision, recall, and F1-measure were 100%, 96.49%, and 98.21%, respectively.

Table 6 shows the confusion matrices for the ensemble model, EfficientNet-B7, ResNet152V2, and DenseNet201. The test dataset had three classes: healthy, partially torn, and completely ruptured. Table 6 shows that the ensemble model correctly predicted 688 samples out of 690 samples for the healthy class, 171 out of 172 samples for the partially torn class, and 54 samples out of 55 samples for the completely ruptured class. EfficientNet-B7 correctly predicted 488 samples for the healthy class, 122 samples for the partially torn class, and 39 samples for the completely ruptured class. ResNet152V2 correctly predicted 493 samples for the healthy class, 123 samples for the partially torn class and 39 samples for the completely ruptured class. DenseNet201 correctly predicted 509 samples for the healthy class, 127 samples for the partially torn class, and 41 samples for the completely ruptured class.

Hyperparameters are the parameters that have an impact on the final performance of the optimizer or any other method, including the one proposed herein, and their selection and tuning is

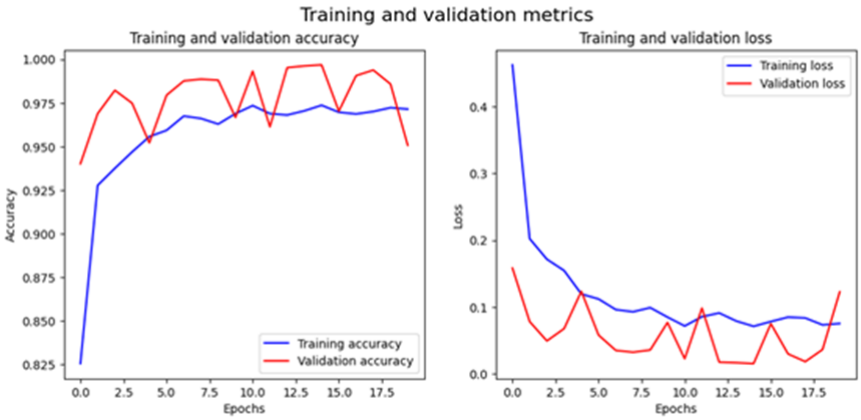


Figure 6: Accuracy and loss of training and validation metrics for the ensemble model

Table 5: Performance of the ensemble model.

	Class	**Precision**	**Recall**	**F1-Measure**	**Specificity**	**Accuracy**
Ensemble model	Healthy	0.9971	0.9985	0.9978	0.9912	**0.9967**
	Partially torn	0.9884	0.9984	0.9884	0.9973	**0.9957**
	Completely ruptured	1	0.9649	0.9821	1	**0.9978**
Average		0.9952	0.9873	0.9894	0.9962	**0.9968**

vital. Hyperparameters were identified using a systematic methodology where initial values were set according to industry standards and previous research. We used a grid search to compare many different values before implementing a random search to sample hyperparameter values. Moreover, the hyperparameters are batch size, learning rate , and number of epochs.

Figure 7 shows the effect of the number of batches on the validation accuracy and loss of the ensemble model. When the batch size goes from 16 to 128, the number of members of a group for training also increases, which in turn increases the validation accuracy and decreases the validation loss. The optimization took 0.9968 seconds when we ran it at a batch size of 128. Likewise, the mean validation loss decreased slightly from 0.6313 to 0.6230 over the same range. This trend seems to imply that a larger batch size is beneficial for model performance, presumably due to the stabilizing effect that it brings to gradient estimates during the training process. Therefore, a batch size of 128 was identified as ideal, with the highest validation accuracy and the least validation loss.

Figure 8 shows the effect of the learning rate on the validation accuracy and validation loss of the ensemble model. The adjustment process for the learning rate has been suspended. The specified learning rates under review are 0.1, 0.01, 0.001, 0.0001, and 0.00001. The maximum validation accu-

Table 6: Confusion matrices

		Healthy	**Partially torn**	**Completely ruptured**
Ensemble model	Healthy	688	1	1
	Partially torn	0	171	1
	Completely ruptured	0	1	54
EffiecientNet-B7	Healthy	488	90	112
	Partially torn	20	122	30
	Completely ruptured	6	10	39
ResNet152V2	Healthy	493	93	104
	Partially torn	25	123	24
	Completely ruptured	6	10	39
DenseNet201	Healthy	509	80	101
	Partially torn	15	127	30
	Completely ruptured	6	8	41

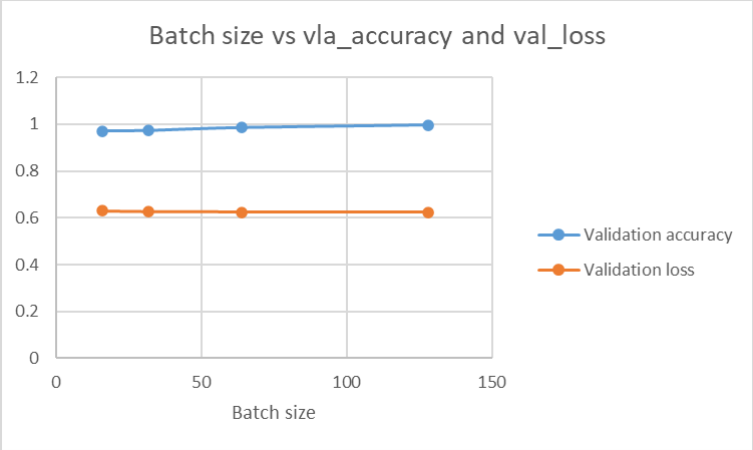


Figure 7: Batch size vs val_Accuracy and val_loss of ensemble model

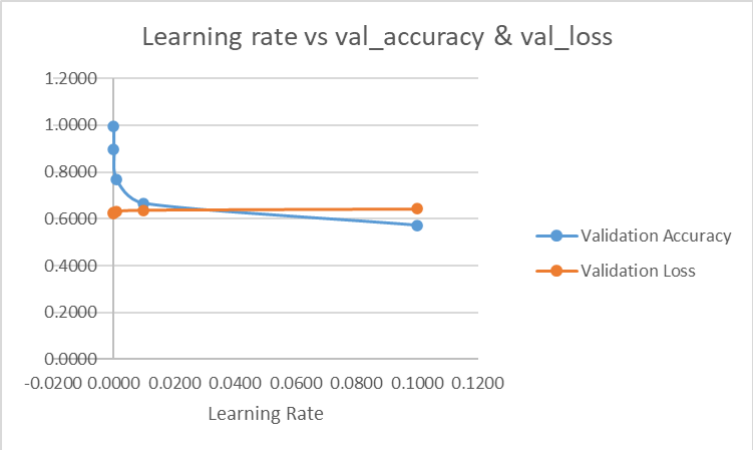


Figure 8: Learning rate vs val_Accuracy and val_loss of ensemble model

racy of the model was 0.9968 and the smallest validation loss was 0.6228. However, the learning rates may be too small for the model to converge at the highest accuracy and the lowest loss. Therefore, a learning rate between zero and one seems to offer the best fit for this model, producing the best accuracy and least loss.

5.3 Comparison with state-of-the-art methods

Table 7 demonstrates that our ensemble model achieves state-of-the-art accuracy in detecting and classifying ACL tears, outperforming previous methods by 1.68%. Our model's accuracy for detecting and classifying ACL tears is 99.68%, compared to the previous state-of-the-art accuracy of 98%. The improved accuracy of our model could lead to more accurate diagnoses of ACL tears, which could improve patient outcomes.

6 Conclusion

The proposed ensemble model provides solutions that can assist radiologists in their tasks, reducing their workload and the potential for errors in ACL tear detection. Early detection allows for timely medical intervention. In many cases, ACL tears may require surgical repair. Early tear detection can lead to a quicker surgical schedule and better outcomes. This paper proposes a robust ensemble model which incorporates DL models with a GA to differentiate between healthy, partially torn, and completely ruptured ACL states. The KneeMRI dataset was used to create and measure the performance of the ensemble model. The dataset was preprocessed by resizing, rescaling, zooming,

Table 7: Comparison between the proposed ensemble deep learning model and previous state-of-the-art methods

Reference	**Methodology**	**performance**	**Datasets**
Bien et al. [19] and AUC: 96.5%	MRNet	Accuracy: 86.7%	KneeMRI dataset
Azcona et al. [20]	ResNet-18 and transfer learning	AUC: 96%.	KneeMRI dataset
Chang et al. [21]	CNN with different inputs of dimensionalities and view	Accuracy: 96.7% and AUC of 97%.	KneeMRI dataset
Liu et al. [22]	LeNet-5 and YOLO	AUC: 98%	KneeMRI dataset
Namiri et al. [23]	3D and 2D CNNs	2D's accuracy: 92%	KneeMRI dataset
Zhang et al. [24]	3D DenseNet	Accuracy: 95.7%	KneeMRI dataset
Germann et al. [25]	DCNN	Sensitivity: 96.1%	KneeMRI dataset
Awan et al. [26]	ResNet-14	Accuracy: 92% %	KneeMRI dataset
Jeon et al. [27]	3D CNN	AUC: 98%	KneeMRI dataset
Astuto et al. [28]	3D CNN	AUC: 90%	KneeMRI dataset
Proposed approach	Ensemble deep learning	Accuracy: 99.68%	KneeMRI dataset

flipping, contrasting, and rotating images. The proposed model was tested on the KneeMRI dataset and compared to competitive methods using the same dataset. The proposed ensemble model provided a maximum precision, recall, F1 measure and accuracy of 99.52%, 98.73%, 98.94%, and 99.68%, respectively, demonstrating state-of-the-art results.

However, the model has some restrictions. The main drawback of the proposed model is that it is not effective when working with large datasets. One of the primary limitations therein is that its overall evaluation focuses on a single type of imaging, which may not reflect all the details of ACL injury. Furthermore, it also depends upon the quality of the data used as well as the variations available in the data set. Despite the comprehensiveness of the KneeMRI dataset, there might be some clinical scenarios not included in the identified and labeled cases, which may affect the generalization of the developed model. Additionally, the implementation of the ensemble model for many attributes/predictors may run into a number of difficulties, especially when applied in real-time consultation in hospitals and clinics.

Future work should focus on improving the performance and generality of the developed model. Another avenue is to explore the incorporation of other types of images intermediate to traditional X-Ray and MRI, such as ultrasound or computer tomography. Furthermore, it might be possible to improve the model's stability and performance by using a range of more complex forms of data augmentation and experimenting with the parameters of the GA. Lastly, the use of transfer learning: using models with similar medical imaging tasks so that they do not have to go through the same tedious labeling of datasets again.

Funding

This work was funded by the Deanship of Graduate Studies and Scientific Research at Jouf University under grant no. DGSSR-2023-02-02114.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Mangone, M.; Diko, A.; Giuliani, L.; Agostini, F.; Paoloni, M.; Bernetti, A.; Santilli, G.; Conti, M.; Savina, A.; Iudicelli, G.; et al.A Machine Learning Approach for Knee Injury Detection from Magnetic Resonance Imaging. *Int. J. Environ. Res. Public Health*,**20(6059)** (2023).

[2] Musahl, V.; Karlsson, J. Anterior cruciate ligament tear. *N. Engl. J. Med.*, **380** (2019), 2341–2348.

[3] Ahmed, I.; Radhakrishnan, A.; Khatri, C.; Staniszewska, S.; Hutchinson, C.; Parsons, N.; Price, A.; Metcalfe, A. Meniscus tears are more common than previously identified, however, less than a quarter of people with a tear undergo arthroscopy. *Knee Surg. Sport. Traumatol. Arthrosc*, **29** (2021), 3892–3898

[4] Lohmander, L.S.; Englund, P.M.; Dahl, L.L.; Roos, E.M.The long-term consequence of anterior cruciate ligament and meniscus injuries. *Am. J. Sport. Med.*, **35** (2007), 1756–1769.

[5] Englund, M.; Roemer, F.W.; Hayashi, D.; Crema, M.D.; Guermazi, A. Meniscus pathology, osteoarthritis and the treatment controversy. *Nat. Rev. Rheumatol.*, **8** (2012), 412–419.

[6] Conaghan, P.G.; Porcheret, M.; Kingsbury, S.R.; Gammon, A.; Soni, A.; Hurley, M.; Rayman, M.P.; Barlow, J.; Hull, R.G.; Cumming, J.; et al. Impact and therapy of osteoarthritis: The Arthritis Care OA, *Nation 2012 survey. Clin. Rheumatol*, **34** (2015), 1581–1588.

[7] Vriezekolk, J.; Peters, Y.A.S.; Steegers, M.A.H.; Davidson, E.N.B.; Ende, C.H.M.V.D. Pain descriptors and determinants of pain sensitivity in knee osteoarthritis: A community-based cross-sectional study. *Rheumatol. Adv. Pract.*, **6** (2022).

[8] Mohammed, A.S.; Hasanaath, A.A.; Latif, G.; Bashar, A. Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images. *Diagnostics*, **13** (2023), 1380.

[9] US EPA. *An Aging Nation: The Older Population in the United States|Health & Environmental Research Online (HERO).* (accessed on 22 July 2022) Available online: https://hero.epa.gov/hero/index.cfm/reference/details/reference_id/2990744

[10] M, G.K.; Goswami, A.D. Automatic Classification of the Severity of Knee Osteoarthritis Using Enhanced Image Sharpening and CNN. *Appl. Sci.*, **13(1658)** (2023).

[11] Mather, R.C., III; Koenig, L.; Kocher, M.S.; Dall, T.M.; Gallo, P.; Scott, D.J.; Bach Jr, B.R.; Spindler, K.P.; Group, M.K. Societal and economic impact of anterior cruciate ligament tears. *J. Bone Jt. Surg. Am.*, **95(1751)** (2013).

[12] Pouly, M.; Koller, T.; Gottfrois, P.; Lionetti, S. Artificial intelligence in image analysis-fundamentals and new developments. *Der Hautarzt Z. Fur Dermatol. Venerol. Und Verwandte Geb.*, **71** (2020), 660–668.

[13] Siouras, A.; Moustakidis, S.; Giannakidis, A.; Chalatsis, G.; Liampas, I.; Vlychou, M.; Hantes, M.; Tasoulis, S.; Tsaopoulos, D. Knee Injury Detection Using Deep Learning on MRI Studies: A Systematic Review. *Diagnostics*, **12(547)** (2022).

[14] Kim, Y.W.; Mansfield, L.T. Fool me twice: Delayed diagnoses in radiology with emphasis on perpetuated errors. *AJR Am. J. Roentgenol.*, **202** (2014), 465–470.

[15] European Society of Radiology (ESR); Codari, M.; Melazzini, L.; Morozov, S.P.; van Kuijk, C.C.; Sconfienza, L.M.; Sardanelli, F. Impact of artificial intelligence on radiology: A EuroAIM survey among members of the European Society of Radiology. *Insights Into Imaging*, **10(105)** (2019).

[16] Avola, D.; Cannistraci, I.; Cascio, M.; Cinque, L.; Diko, A.; Fagioli, A.; Foresti, G.L.; Lanzino, R.; Mancini, M.; Mecca, A.; et al. A Novel GAN-Based Anomaly Detection and Localization Method for Aerial Video Surveillance at Low Altitude. *Remote Sens.*, **14(4110)** (2022).

[17] Otter, D.W.; Medina, J.R.; Kalita, J.K. A survey of the usages of deep learning for natural language processing. *IEEE Trans. Neural Netw. Learn. Syst.*, **32** (2020), 604–624.

[18] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2016), pp. 770–778.

[19] Bien, N.; Rajpurkar, P.; Ball, R.L.; Irvin, J.; Park, A.; Jones, E.; Bereket, M.; Patel, B.N.; Yeom, K.W.; Shpanskaya, K. Deep-learning assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Med.*, **15** (2018).

[20] Azcona, D.; McGuinness, K.; Smeaton, A.F. A Comparative Study of Existing and New Deep Learning Methods for Detecting Knee Injuries using the MRNet Dataset. *In Proceedings of the 2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA), Kuala Lumpur, Malaysia,*, (18–20 September 2020), pp. 149–155.

[21] Chang, P.D.; Wong, T.T.; Rasiej, M.J. Deep Learning for Detection of Complete Anterior Cruciate Ligament Tear. *J. Digit. Imaging*, **32** (2019), 980–986.

[22] Liu, F.; Guan, B.; Zhou, Z.; Samsonov, A.; Rosas, H.; Lian, K.; Sharma, R.; Kanarek, A.; Kim, J.; Guermazi, A. Fully automated diagnosis of anterior cruciate ligament tears on knee MR images by using deep learning. Radiol. *Artif. Intell.*, **1(180091)** (2019).

[23] Namiri, N.K.; Flament, I.; Astuto, B.; Shah, R.; Tibrewala, R.; Caliva, F.; Link, T.M.; Pedoia, V.; Majumdar, S. Deep Learning for Hierarchical Severity Staging of Anterior Cruciate Ligament Injuries from MRI. *Radiol. Artif. Intell.*, **2** (2020).

[24] Zhang, L.; Li, M.; Zhou, Y.; Lu, G.; Zhou, Q. Deep Learning Approach for Anterior Cruciate Ligament Lesion Detection: Evaluation of Diagnostic Performance Using Arthroscopy as the Reference Standard. *J. Magn. Reson. Imaging*, **52** (2020), 1745–1752.

[25] Germann, C.; Marbach, G.; Civardi, F.; Fucentese, S.F.; Fritz, J.; Sutter, R.; Pfirrmann, C.W.; Fritz, B. Deep Convolutional Neural Network–Based Diagnosis of Anterior Cruciate Ligament Tears: Performance Comparison of Homogenous Versus Heterogeneous KneeMRI CohortsWith Different Pulse Sequence Protocols and 1.5-T and 3-T Magnetic Field Strengths. *Investig. Radiol.*, **55(499)** (2020).

[26] Awan, M.J.; Rahim, M.S.M.; Salim, N.; Mohammed, M.A.; Garcia-Zapirain, B.; Abdulkareem, K.H. Efficient detection of knee anterior cruciate ligament from magnetic resonance imaging using deep learning approach. *Diagnostics*, **11(105)** (2021).

[27] Jeon, Y.S.; Yoshino, K.; Hagiwara, S.; Watanabe, A.; Quek, S.T.; Yoshioka, H.; Feng, M.J.I.J.O.B.; Informatics, H. Interpretable and lightweight 3-D deep learning model for automated ACL diagnosis. *IEEE J. Biomed. Health Inform.*, **25** (2021), 2388–2397.

[28] Astuto, B.; Flament, I.K.; Namiri, N.; Shah, R.; Bharadwaj, U.; M. Link, T.; D. Bucknor, M.; Pedoia, V.; Majumdar, S.J.R.A.I. Automatic Deep Learning–assisted Detection and Grading of Abnormalities in Knee MRI Studies. Radiol. *Artif. Intell.*, **3** (2021).

[29] tajduhar, I.; Mamula, M.; Miletić, D., Unal,G.; Semi-automated detection of anterior cruciate ligament injury from MRI, *Computer Methods and Programs in Biomedicine,* **140** (2017), p.p. 151–164.

[30] Tan, M. and Quoc V. Le; Tan and Le EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *CoRR*, 2019.

[31] Bilal, M., Maqsood, M., Yasmin, S. et al. A transfer learning-based efficient spatiotemporal human action recognition framework for long and overlapping action classes. *J Supercomput*, **78** (2022), 2873–2908.

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Mahmood et al (2024). Acute Knee Injury Detection with Magnetic Resonance Imaging (MRI), *International Journal of Computers Communications & Control*, 19(5), 6648, 2024.
https://doi.org/10.15837/ijccc.2024.5.6648