KOA-NN: A convolutional neural network model based on spatial and channel attention mechanisms for detecting knee osteoarthritis from X-ray images

KOA-NN: Un modelo de red neuronal convolucional basado en mecanismos de atención espacial y de canal para detectar osteoartritis de rodilla a partir de imágenes de rayos X

Uriel Calderon-Uribe[1], Salvador Calderon-Uribe[2], Rocio Camarena-Martinez[1], Esau Gonzalez-Villagomez[3], Jacob Gonzalez-Villagomez[4]
[1] Laboratorio de Visión por Computadora, Tecnológico Nacional de México / ITS Sur de Guanajuato Educación Superior 2000, Uriangato, Guanajuato, México.
urielcal92@gmail.com
[2] Laboratorio de Visión Artificial y Termografía/Mecatrónica Facultad de Ingeniería, Universidad Autónoma de Querétaro, San Juan del Rio, México.
[3] Instituto Tecnológico Superior de Irapuato, Irapuato, Guanajuato, México.
[4] Departamento de Estudios Multidisciplinarios, División de Ingenierías, Campus Irapuato-Salamanca, Universidad de Guanajuato, Yuriria, Guanajuato, México.

Abstract

Knee osteoarthritis (KOA) is a deformity that causes mobility problems, pain, and wear and tear in the knee joints. Currently, one way to diagnose KOA is through the use of X-ray images. An expert analyzes the images and infers the patient's condition. However, this is a tiring and error-prone task, since the expert must analyze a large number of images to make a diagnosis. For this reason, in this work, a new convolutional neural network (CNN) model based on attention mechanisms called KOA-NN is used to identify KOA accurately. The results of this work position this new model as an alternative to the models established in the literature.

Introduction

Knee osteoarthritis (KOA) is a type of arthritis characterized by knee deformities. Various studies around the world show that KOA is a leading cause of disability in people with this condition (Pesudovs *et al.,* 2015). In addition to common treatments such as medication and physical therapy, total joint replacement surgery is the only available option to relieve the pain suffered by people with KOA. Today, expert doctors evaluate the condition of patients with KOA through X-ray images. Through these images, the expert assesses the patient's condition and infers the type of KOA they suffer from. However, experts are susceptible to errors such as fatigue or the large number of images they must analyze to reach a conclusion. For these reasons, it is important to develop new computer-assisted systems that allow experts to infer a more accurate result.

Nowadays, there are different methodologies available for the classification of knee osteoarthritis. Among the most widely used methods, convolutional neural networks (CNNs) have gained popularity due to their ability to find features automatically. Several recent studies have applied pre-trained models such as ResNet (Jahan *et al.,* 2024), VGG16 (Saini *et al.,* 2023), and DenseNet (Prakas, 2024) to classify KOA. The main advantage of using these models is their adaptability to new datasets, making them suitable for medical image analysis. However, their disadvantages include the large number of computational resources required to operate them. For this reason, it is important to develop new, low-cost models that maintain high classification rates while minimizing computational costs.

In this work, a novel system based on CNNs and attention mechanisms is used as a tool for KOA classification. First, a CNN is designed to implement attention mechanisms to infer KOAs. The attention mechanisms are used as a feature refinement tool to differentiate the KOA problem. Finally, the final model is evaluated using different metrics to validate it. The results of this study show high classification rates with a reduced number of computational resources

Materials and Methods

This section describes the techniques and methods used for KOA detection. Figure 1 shows the phases used to build KOA-NN. As can be seen in this figure, the system is divided into four phases: image preprocessing, training and test set creation, KOA-NN training, and validation. In the first phase, the image preprocessing phase, data augmentation is used to homogenize the dataset. Subsequently, in the training and test set creation phase, two datasets are created to train and evaluate the final model. Subsequently, in the training phase, the KOA model is trained using the previously created training set. Finally, in the evaluation phase, the final model is subjected to evaluation for validation.

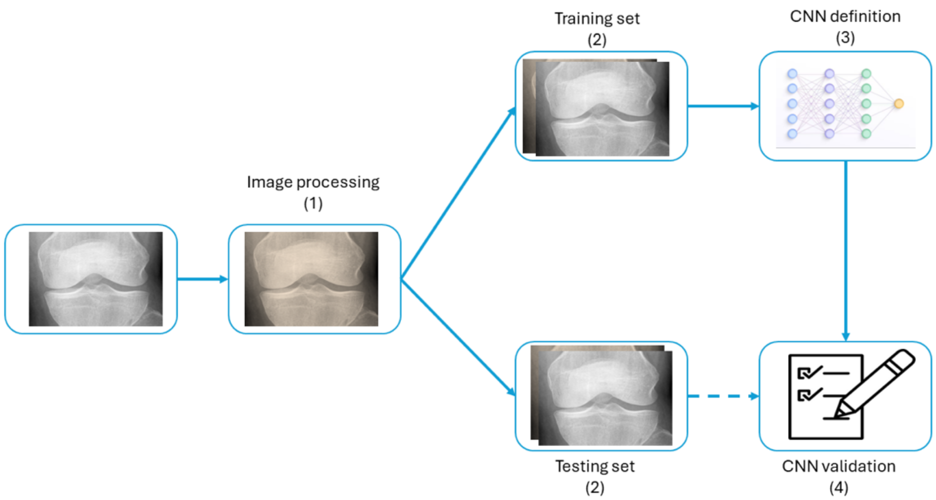

Figure 1. Overall description of the KOA-NN model.

Dataset Description

To carry out the development of the KOA-NN, the dataset established in (Chen 2018) was used. This dataset is composed of X-ray images divided into two classes: KOA and healthy. Figure 2 shows the distribution of images present in the dataset.

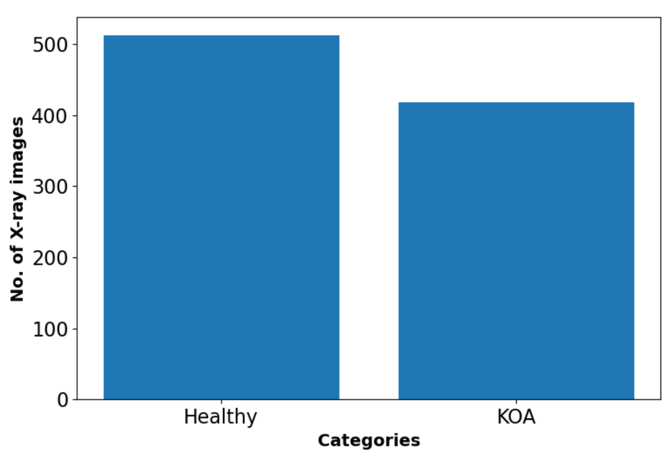

Figure 2. Images present in the dataset.

As can be seen in this figure, there are more X-ray images for the healthy class than for the KOA class. To address this problem, different data augmentation techniques were implemented. In this case, the data augmentation techniques used were brightness adjustment and random horizontal and vertical flipping. Equations 1 and 2 describe the transformations mentioned above.

$$g(i,j) = \alpha f(i,j) + \beta \qquad (1)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \qquad (2)$$

In equation 1, α represents the gain factor used to adjust the contrast in the resulting image $g(i,j)$. The β factor is used to control the brightness in the resulting image, and $f(i,j)$ is the original image. In equation 2, x and y represent the position of the new pixel in the image, θ represents the angle to rotate, and u and v represent the position of the pixel to rotate. In this work, the α factor is adjusted to 1, the β factor is a random value between [0, -0.5], and θ is a scattering value between [0, 45] degrees. Figure 3 shows an example from the transformations applied to the healthy class.

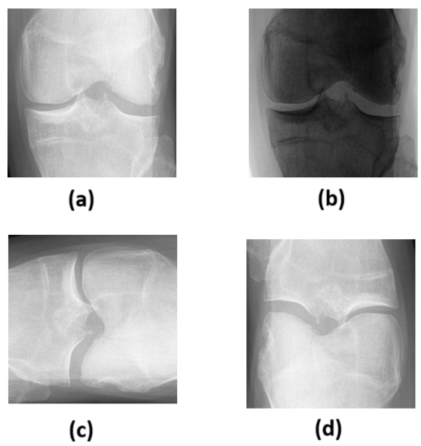

Figure 3. Transformations applied to healthy class, a) Original image, b) Brightness adjustment, c) Rotation 90º, and d) Image inversion.

Once the dataset is homogenized, two different datasets, training and testing, are generated. On the one hand, the training dataset is used to fine-tune the KOA-NN weights, and on the other hand, the test set is used to validate the final model. The ratio of the test and training sets is 10% and 90%, respectively. That is, the number of images required to test the model and the number of images required to train the model.

KOA-NN architecture

In deep learning, convolutional neural networks are a type of neural network used for image processing. The learning process consists of extracting image features using kernels. The objective of feature extraction is to find patterns that allow differentiation between two or more classes. In this work, a neural network based on attention mechanisms is used to classify KOA and healthy individuals. Figure 4 shows the CNN structure used in the KOA identification.

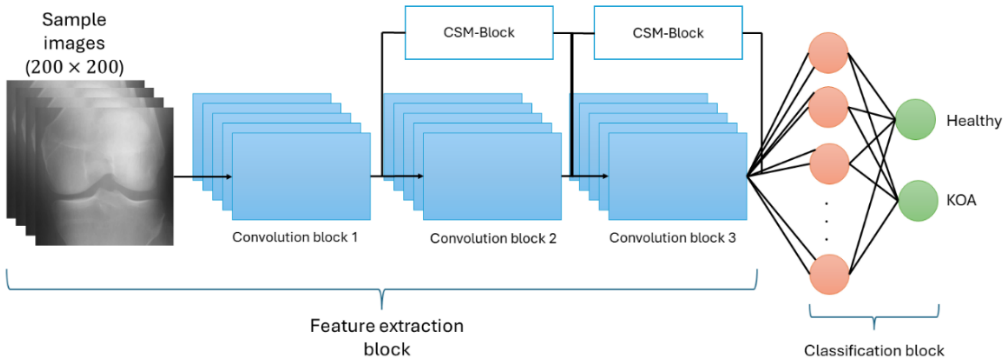

Figure 4. Proposed CNN architecture for KOA classification (CSM refers to channel and spatial attention mechanism).

As can be seen in this figure, the model is composed of three convolutional layers and a feedforward neural network. The first convolutional block (convolutional block 1) is composed by a convolutional layer with a kernel size of 3×3, padding the "same", and ReLU (Rectified Linear Unit) as an activation function. The convolutional block 2 is formed by a kernel with size of 3×3, padding the "same", and ReLU as an activation function. Finally, the convolutional block 3 is composed by a kernel with size of 3×3, padding the "same", and ReLU as an activation function. Between the convolutional block 2 and 3, there are an attention block, which allows for refining each of the features obtained by each convolutional block (Woo *et al.*, 2018). Figure 5 shows the structure of each attention block used in the final CNN.

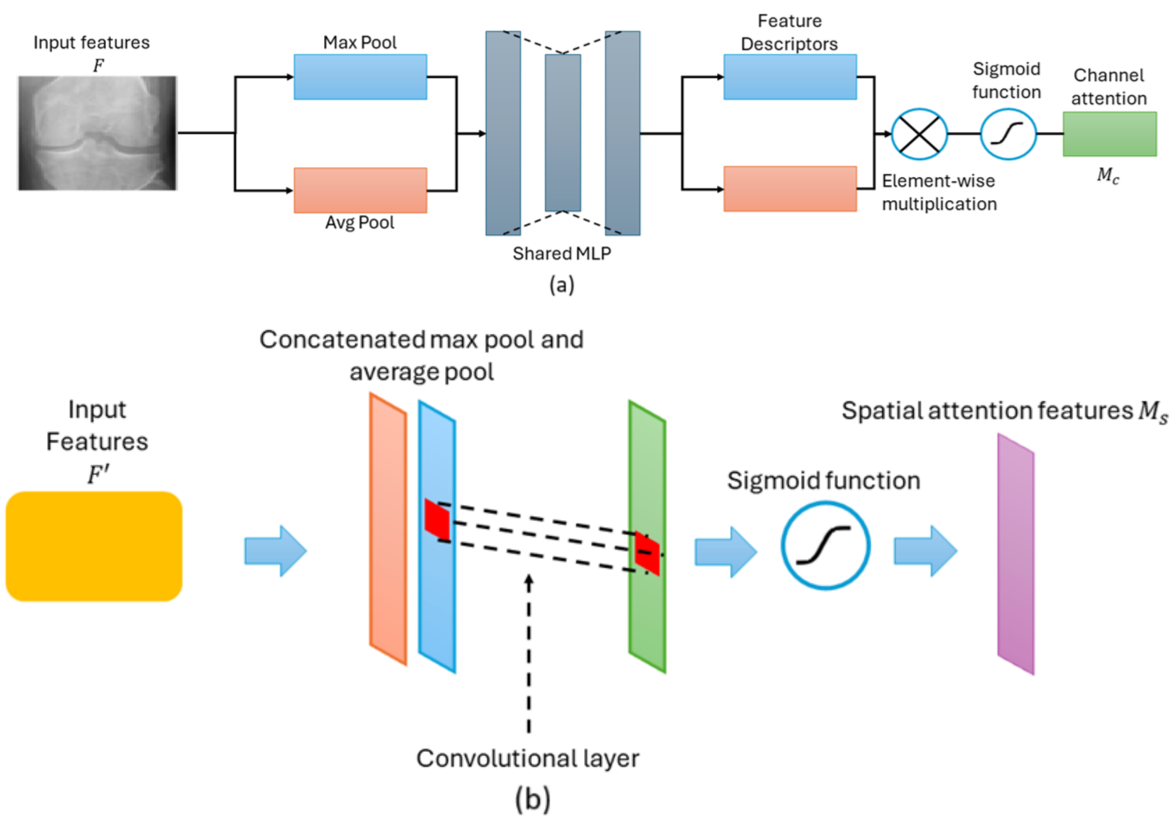

Figure 5. Representation of channel and spatial attention mechanism, a) Channel attention layer, b) Spatial attention layer.

According to the figure shown above, the attention block is composed of two mechanisms: the channel-based attention mechanism and the spatial attention mechanism. The objective of the channel-based attention mechanism is to summarize the characteristics of each channel and use them as a weighting factor in the final result. On the other hand, the spatial attention mechanism summarizes the characteristics in each batch and uses them to highlight them in the final batch. Finally, the feed-forward neural network is formed by two dense layers. The first layer contains 512 neurons and ReLU as an activation function. The final dense layer contains two neurons and sigmoid as an activation function. Adam, learning and binary crossentropy were used to set up the optimizer and the loss function respectively.

Model evaluation

In order to evaluate the results of the final model, various metrics were used to measure its performance. These metrics are:

Accuracy: which is a measure used to measure the correct predictions of the model.

$$Accuracy = \frac{TP+TN}{TP+TN+ FP+FN} \qquad (3)$$

Where TP represents the positive instances assigned to the positive class. TN are the negative instances classified as negative. FP are positive instances assigned as negative instances. Finally, FN represents the negative instances classified as positive instances.

Precision: the number of true positives divided by the total number of positive predictions.

$$Precision = \frac{TP}{TP+FP} \qquad (4)$$

Recall: is a measure of how often a model correctly identifies positive instances (true positives) from all the actual positive samples in the dataset.

$$recall = \frac{TP}{TP +FN} \qquad (5)$$

Results

To evaluate the results of the final model, a comparison was performed between a model without attention mechanisms and one with attention mechanisms. The objective of this comparison is to observe which features are learned by both models. Figure 6 shows the results obtained from both models using the GradCAM technique (Xiao *et al.,* 2021).

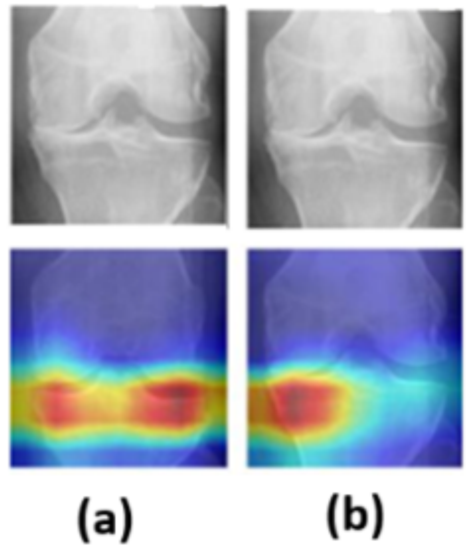

Figure 6. *Results obtained to apply the GradCAM method, a) Proposed model without channel and spatial attention mechanism, b) Proposed model with channel and spatial attention mechanism.*

As can be seen in this figure, the features learned by the model using attention mechanisms stand out more than those learned by a conventional model. Note that, despite the improvement in the features learned by the KOA-NN, this entails a higher computational cost and a longer training time. Finally, Table 1 shows a comparison of different state-of-the-art models.

Table 1: Comparison between different models established in the state of the art (MLP refers to multi-layer perceptron, DNN refers to Dense neural network. MDL refers to Mixture deep learning, CSAM refers to channel and spatial attention mechanism).

Author	Method	Accuracy %	Precision %	Recall %	F1-score %
Brahim et al., 2019	MLP	82.98	80.65	82.15	-
Moustakidis et al., 2020	DNN	79.6	-	-	-
Ahmed 2022	MDL	90	91	90.8	-
Proposed	CNN + CSAM	100	100	100	100

As can be seen in this table, the KOA-NN model stands out from the other models. This is due to the way in which features are acquired. Attention mechanisms allow for refining and extracting the essential features of each class.

Conclusions

In this work, we propose a new system for osteoarthritis detection. The system is based on a CNN with attention mechanisms that allow inference (based on X-ray images) between two types of diseases: osteoarthritis and a healthy individual. According to different validation metrics, the KOA-NN achieves an accuracy of 90%. This makes it ideal for implementation in medical settings as a healthcare tool. Future work will focus on improving this network to allow for the classification of more deformities.

References

Ahmed, S. M., & Mstafa, R. J. (2022). Identifying Severity Grading of Knee Osteoarthritis from X-ray Images Using an Efficient Mixture of Deep Learning and Machine Learning Models. *Diagnostics,* 12(12), 2939. https://doi.org/10.3390/diagnostics12122939

Brahim, A., Jennane, R., Riad, R., Janvier, T., Khedher, L., Toumi, H., & Lespessailles, E. (2019). A decision support tool for early detection of knee OsteoArthritis using X-ray imaging and machine learning: Data from the OsteoArthritis Initiative. *Computerized Medical Imaging and Graphics,* 73, 11-18.

Chen, P. (2018). “Knee Osteoarthritis Severity Grading Dataset”, Mendeley Data, V1, doi: 10.17632/56rmx5bjcr.1

Jahan, M., Hasan, M. Z., Samia, I. J., Fatema, K., Rony, M. A. H., Arefin, M. S., & Moustafa, A. (2024). KOA-CCTNet: An enhanced knee osteoarthritis grade assessment framework using modified compact convolutional transformer model. IEEE Access.

Moustakidis, S., Papandrianos, N. I. Christodolou, E. Papageorgiou, E. & Tsaopoulos, D. Dense Neural Networks in Knee Osteoarthritis Classification: A Study on Accuracy and Fairness*. Neural Comput. Appl.,* 5, 1–13.

Pesudovs, K., Melaku, Y. A., & Global Burden of Disease Study 2013 Collaborators. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013 diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet,* 386(9995), 743-800.

Prakash, P. (2024). Severity Classification of Knee Osteoarthritis from XRay Images using Deep Learning. [Doctoral dissertation] Dublin, National College of Ireland.

Saini, D., Khosla, A., Chand, T., Chouhan, D. K., & Prakash, M. (2023). Automated knee osteoarthritis severity classification using three-stage preprocessing method and VGG16 architecture. *International Journal of Imaging Systems and Technology,* 33(3), 1028-1047.

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) pp. 3-19.

Xiao, M., Zhang, L., Shi, W., Liu, J., He, W., & Jiang, Z. (2021, September). A visualization method based on the Grad-CAM for medical image segmentation model. In 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS) (pp. 242-247). IEEE.