An Attention-Based Deep Learning Approach to Knee Injury Classification from MRI Images

Kowshik Deb Nath*, A. F. M. Minhazur Rahman[†], and Md. Ali Hossain[‡]

Department of Computer Science & Engineering

Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh

Email- *kowshikcseruet1998@gmail.com [†]m.r.saurov@gmail.com, [‡]ali.ruet@gmail.com

Abstract—Knee injuries, prevalent in athletic and aging populations, pose significant challenges to healthcare professionals due to their complex nature and the critical function of the knee joint. Early and accurate diagnosis is paramount to ensure effective treatment and minimize long-term complications. Traditional diagnostic methods, including physical examinations and imaging techniques like MRI, require expert interpretation and can sometimes be inconclusive. This study introduces an approach to knee injury classification using deep learning techniques by leveraging convolutional neural networks (CNNs) with Attention Mechanism. This research work integrates powerful feature extraction capabilities of CNN and feature refinement of attention mechanism for the binary and multi-class classification of knee MRI images, with the aim of accurately identifying specific knee injury types. Based on our experiment on two comprehensive knee MRI datasets, our custom CNN model achieved 88% testing accuracy on Dataset-1 (Binary classification) and 77% accuracy on Dataset-2 (Multi-class classification). Meanwhile, the Attention-based CNN model achieved 100% accuracy on Dataset-1 (Binary Classification) and 91% accuracy on Dataset-2 (Multi-Class Classification). This approach not only holds promise for enhancing diagnostic accuracy but also for reducing the time to diagnosis.

Index Terms—Knee Injuries, MRI Knee Images, Deep Learning Algorithms, Attention Mechanism, Convolutional Neural Networks(CNNs), Binary Classification, Multi-class classification, Data augmentation, Radiologists.

I. INTRODUCTION

Knee injuries are a significant concern, especially in sports, accounting for a major portion of severe injuries that lead to prolonged absences from sports participation [1]-[4]. Specifically, ruptures in the anterior cruciate ligament (ACL) make up over half of these cases, impacting around 200,000 Americans annually [1], [5]–[7]. Additionally, knee cartilage issues affect nearly 900,000 Americans each year, leading to over 200,000 surgeries [5]-[8]. Meniscal injuries rank as the second most frequent knee-related ailment, with a 12-14% incidence rate [9], and a prevalence of 60-70 cases per 100,000 in the UK. The financial burden of ACL injuries alone exceeds \$7 billion in the U.S. [10]. Knee injuries are linked to both immediate and long-term pain, disability, and a decline in overall health quality [12], [13], [24]. Particularly, active young individuals are at a higher risk of knee injuries due to their involvement in intense activities, increasing their chances of developing osteoarthritis (OA) [14]. Statistically, about half of those with ACL or meniscal injuries will show signs of knee OA within 10 to 20 years post-injury [8], [15]. Other potential outcomes

of knee injuries include structural muscle injuries in the lower limb [16] and tendinopathies [17]. These statistics highlight the societal and economic implications of knee injuries, emphasizing the need for efficient and cost-effective diagnostic methods.

A. Types of Knee Injuries

The anatomical structure of the knee joint is categorized as depicted in Figure 1, according to the taxonomy presented in [18].



Fig. 1. Taxonomy of knee joint anatomy.

In the provided illustrations, various knee injuries are depicted: Figure 2 showcases an ACL ligament injury, Figure 3 represents a PCL injury, and Figure 4 demonstrates an MCL injury.



Fig. 2. Anterior cruciate ligament anatomy [31].

B. Traditional and Deep Learning based Approaches

Arthroscopy, traditionally used for diagnosing knee issues, is limited due to its invasive nature, leading to MRI as a non-invasive alternative. MRI, however, is challenging to interpret and prone to human error, causing inconsistencies in knee injury diagnoses [19]–[24].



Fig. 3. Posterior cruciate ligament anatomy [32].



Fig. 4. Medial collateral ligament anatomy [33].

Recent advancements in deep learning have seen its application in identifying ACL tears from knee MRI scans. In their study, Sridhar et al. [29] applied a Deep Convolution Neural Network using the Inception-v3 framework, achieving a high training accuracy of 99.04% but a lower 95.42% in testing. This gap suggested a notable overfitting in the model. Meanwhile, Wahid et al. [30] adopted a multi-layer Convolutional Sparse Coding technique, attaining 95% accuracy with a dataset of 623 MRI images, which included various tear types. However, the limitation in their study was the small dataset size, which might have restricted the model's potential for higher accuracy.

These studies shed light on the potential of machine and deep learning in medical imaging. Yet, they highlight the necessity for ongoing research and expanded datasets to improve the precision and dependability of these models.

C. Our Contribution

Our study presents a method that merges Convolutional Neural Networks (CNNs) with an attention mechanism to highlight specific features. This model excels in multi-class classification, identifying different tear types like "ACL", "FCL", "MCL", "Normal", and "PCL". To counter overfitting and boost dataset variety, we employed data augmentation strategies. Utilizing a vast dataset, our approach demonstrates the effectiveness of the attention mechanism in enhancing classification accuracy.

II. OVERVIEW OF ATTENTION MECHANISM

Attention mechanisms have become a pivotal concept in deep learning, particularly in tasks that require the model to focus on specific parts of the input data. The idea of attention in deep learning is inspired by the human cognitive process of selectively concentrating on a particular aspect of information while ignoring others [25].

A. Concept of Attention

The basic idea of attention is to allow the model to weigh different parts of the input differently. In a sense, the model



Fig. 5. Base structure of attention module.

attends to certain parts of the input when making predictions or representations, as illustrated in Figure 5. This is analogous to how humans pay attention to specific details when observing a scene or listening to a speech [26]. In deep learning models, attention mechanisms can be implemented in various ways. A common approach is to assign a weight to each part of the input. These weights determine how much focus the model should give to different parts of the input when making predictions. The weights are often computed based on the compatibility between the input and some context, such as a query or another part of the input [27].

B. Squeeze-and-Excitation Network

The Squeeze-and-Excitation Network (SENet) [28] introduces an attention mechanism to adaptively recalibrate channel-wise feature responses in a convolutional neural network. This recalibration is achieved by explicitly modeling the interdependencies between channels of the convolutional features. The core idea is to use global information to selectively emphasize informative features and suppress less useful ones in a channel-wise manner. The SE block can be divided into two steps: the squeeze operation and the excitation operation as illustrated in Figure 6.

Squeeze: This operation captures the global information of the feature map. It involves global average pooling which compresses the spatial dimensions $H \times W$ of the feature map, resulting in a 1D vector of size C (number of channels).

$$z_{c} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_{c}(i,j)$$
(1)

where u_c is the input feature map and z_c is the output of the squeeze operation for channel c.

Excitation: This operation captures channel-wise dependencies. It involves passing the squeezed 1D vector through two fully connected (FC) layers, the first acting as a bottleneck and the second producing the scaling factors. Between these FC layers, a ReLU activation function is applied, and after the second FC layer, a sigmoid activation is applied to obtain the scaling factors in the range between 0 to 1.

$$s_c = \sigma \left(W_2 \delta \left(W_1 z_c \right) \right) \tag{2}$$

where σ is the sigmoid function, δ is the ReLU function, and W_1 and W_2 are the weights of the two FC layers.



Fig. 6. A detailed diagram of the squeeze and excitation network with proper dimensions and the different operations.

Finally, the original feature map u_c is scaled by the output s_c of the excitation operation:

$$\tilde{u}_c = s_c \cdot u_c \tag{3}$$

where \tilde{u}_c is the recalibrated feature map.

III. PROPOSED METHODOLOGY

Our methodology for ACL tear detection in knee MRI images combines Convolutional Neural Networks (CNNs) with attention mechanisms. The attention layer, placed right after the initial convolutional layers, was chosen through a mix of theoretical analysis and empirical testing. This positioning allows the layer to effectively refine early-stage features, enhancing the model's precision in detecting subtle variations in knee injuries. We opted for the SENet attention layer due to its proficiency in medical imaging, particularly for its channelwise feature recalibration capabilities. Crucially, during training, the attention mechanism learns to focus on the most pertinent features for each knee injury class, such as ACL, FCL, MCL, Normal, and PCL. This adaptive learning, based on the correlation between input features and class labels, sharpens the model's ability to distinguish between various injury types, thereby improving its accuracy and efficiency.

In this section, Tables I and II detail the configurations of our CNN 7 and Attention-based CNN architecture 8 for both binary and multiclass classification, respectively. These tables highlight key aspects such as batch sizes, trainable parameters, optimizer choice, and loss functions, offering a clear view of our model's architecture vital for ACL tear detection.

A. Attention Module

Given an input feature, the attention block illustrated in Figure 8 can be mathematically represented as follows:

Let the input feature be represented as I.

1. The gating signal
$$g$$
 is computed as:

$$g = \text{ReLU}(W_g * I + b_g)$$
$$g = \text{ReLU}(W'_g * g + b'_g)$$

TABLE I CNN ARCHITECTURE CONFIGURATION

Parameter	Value		
Total Parameters	7,757,509		
Trainable Parameters	7,757,509		
Non-trainable Parameters	0		
Batch Size	32		
Optimizer	RMSprop		
Loss Function	Categorical Crossentropy		

TABLE II ATTENTION-BASED CNN ARCHITECTURE CONFIGURATION

Parameter	Value			
Total Parameters	21,736,229			
Trainable Parameters	21,730,725			
Non-trainable Parameters	5,504			
Batch Size	32			
Optimizer	RMSprop			
Loss Function	Categorical Crossentropy			
SENet Reduction Ratio	8			

where W_g and W'_g are the weights for the two convolutional layers applied to the gating signal, and b_g and b'_g are the corresponding biases.

2. The transformation x of the input tensor is computed as:

$$x = \operatorname{ReLU}(W_x * I + b_x)$$
$$x = \operatorname{ReLU}(W'_x * x + b'_x)$$

where W_x and W'_x are the weights for the two convolutional layers applied to the input tensor, and b_x and b'_x are the corresponding biases.

3. The attention coefficients ψ are computed as:

$$\psi = \sigma(x \odot g)$$

where σ is the sigmoid activation function and \odot denotes element-wise multiplication.

4. The output of the attention block is:

$$O = I \odot \psi$$

B. Squeeze-and-Excitation Block

In our model architecture, as visualized in Figure 8, we've adeptly incorporated the principles of attention, as elucidated in Subsection II-B. The attention mechanism, which allows models to assign varying importance to different parts of the input, has been embedded at strategic points in our design.

Specifically, we've employed spatial attention in our second block, where the model refines its focus on specific spatial hierarchies within the input. This is akin to how humans selectively concentrate on certain aspects of a visual scene.

Furthermore, the Squeeze and Excitation (SE) channel attention has been integrated into the third, fourth, and fifth blocks of our architecture. This mechanism refines the features by allowing the model to weigh the importance of different channels, ensuring that the model captures global contextual information from the input. The amalgamation of spatial and SE channel attention in our architecture ensures that the model is not only sensitive to where specific features are but also to the interplay between different features across the channels. This dual attention strategy, rooted in the concepts described in Subsection II-B, enhances the model's ability to extract and refine features, leading to improved performance.

C. Framework of Proposed Method

In our research, we employed a variety of methods to ensure the robustness and accuracy of our findings:

We designed a custom Convolutional Neural Network tailored to the specific features and nuances of knee MRI images in Figure 7. This network allowed us to capture intricate patterns from input knee MRI images.



Fig. 7. CNN architecture for binary and multi-class classification.

In Figure 8, we present the attention-based CNN architecture tailored for both binary and multi-class classification of knee injury.

Our model employs convolutional layers to discern spatial patterns. The attention block refines the model's focus on pivotal input regions, enhancing prediction accuracy. Meanwhile, the Squeeze-and-Excitation block recalibrates channelwise responses, amplifying crucial features. Together, these blocks ensure the model's attentiveness and adaptability to diverse data nuances.

IV. EXPERIMENTAL ANALYSIS

Our model utilizes two key attention mechanisms to identify different knee injuries from MRI images. The Squeeze-and-Excitation (SE) block in Figure 8 enhances channel-wise feature analysis, recalibrating each channel to highlight more significant features. Additionally, a custom spatial Attention block in Figure 8 refines the focus on specific areas within the images, using convolutional layers to generate and apply gating signals. This combination effectively distinguishes between injury types like ACL, FCL, MCL, Normal, and PCL, enabling precise and accurate classification.

A. Datasets

In our research work, we have collected two different datasets one is "BMEII-AI RedImageNet" and another is "RedImage". Where the first one contains a total 1021 images of both "ACL" and "NORMAL" classes. On the other hand, the second dataset contains 23,780 images of five classes(ACL, FCL, MCL, Normal, PCL). Both datasets contain RGB images of dimension $256 \times 256 \times 3$.

TABLE III GROUND TRUTH CLASS FOR DATASET-1 (BMEII-AI REDIMAGENET)

Class	Samples		
ACL	512		
Normal	407		

 TABLE IV

 GROUND TRUTH CLASSES FOR DATASET-2 (REDIMAGE)

Class	Samples		
ACL	10,085		
FCL	466		
MCL	7,911		
Normal	4,593		
PCL	725		

B. Data Augmentation

Data augmentation involves artificially increasing the size of a training dataset by transforming the original images in various ways. This method is crucial for mimicking realworld conditions and combating overfitting, particularly when data is scarce. In our model, to enhance its robustness and diversify the training data, we implemented a range of data augmentation techniques. These included rotations, shearing, zooming, and horizontal flipping, all aimed at replicating different real-world situations and reducing overfitting.

C. Evaluation Metircs

To comprehensively assess the performance of our knee injury classification model, we employed a suite of evaluation metrics, each offering a unique perspective on the model's capabilities:

Accuracy: This metric provided a straightforward measure of the model's overall correctness. It quantified the proportion of total predictions that were correct, serving as an initial indicator of the model's efficacy.

$$Accuracy = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total Predictions}}$$
(4)

Precision: Assesses the model's accuracy in identifying positive cases, calculated as the proportion of true positives out of all positive predictions.

$$Precision = \frac{True Positives (TP)}{True Positives (TP) + False Positives (FP)}$$
(5)

Recall (Sensitivity): Determines the model's ability to identify all possible positive cases, defined by the ratio of true positives to the total of true positives and false negatives.

$$Recall = \frac{True Positives (TP)}{True Positives (TP) + False Negatives (FN)}$$
(6)

F1-Score: Represents the harmonic mean of precision and recall, providing a balanced view of the model's performance,



Fig. 8. Attention-based CNN architecture for binary and multi-class classification.

particularly useful for imbalanced class distributions.

$$F1\text{-}Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
(7)

D. Performance Evaluation

According to Table V, it is evident that the integration of attention mechanisms with our Custom CNN model has led to significant improvements in all evaluation metrics for both binary and multi-class knee injury classification. For binary classification, the accuracy, precision, recall, and F1-score all reached a perfect score of 100% with the attention-augmented model, compared to 88% without attention. Similarly, for multi-class classification, there was a notable jump from 77% to 91% in accuracy, and consistent improvements in precision, recall, and F1-score.

The substantial enhancement in performance can be attributed to the attention mechanism's ability to focus on the most relevant features in the input data, thereby allowing the model to make more informed decisions. This selective focus is especially crucial in medical imaging, where subtle details can be the difference between accurate and erroneous diagnosis.

Furthermore, when comparing our model's performance to the state-of-the-art DCNN combined with InceptionV3 [29], as shown in Figure 9, our attention-augmented Custom CNN outperforms the benchmark in both binary and multi-class scenarios. This underscores the efficacy of attention mechanisms in enhancing the discriminative power of convolutional neural networks, especially in complex tasks such as knee injury classification.

Comparison of Both Binary and Multiclass Classification in Figure 9.

V. CONCLUSION

Our research underscores the transformative potential of advanced neural networks in medical diagnostics, particularly for knee injury classification. By employing an Attentionbased CNN architecture, we achieved a remarkable 100%



Fig. 9. Comparison between various models for binary and multi-class classification.

accuracy in binary classification, outperforming the traditional CNN's 88%. Utilizing the Attention-based CNN architecture for multi-class classification, we attained an accuracy of 91%, surpassing the conventional CNN's accuracy rate of 77%. This success is attributed to the model's ability to prioritize diagnostically relevant features within MRI images. As the medical domain progresses, the integration of such tailored deep-learning techniques will be crucial, offering promising avenues for future research and improved diagnostic precision.

REFERENCES

- V. Musahl and J. Karlsson, "Anterior cruciate ligament tear," N. Engl. J. Med., vol. 380, pp. 2341–2348, 2019.
- [2] I. Ahmed et al., "Protocol: Meniscal tear outcome Study (METRO Study): A study protocol for a multicentre prospective cohort study exploring the factors which affect outcomes in patients with a meniscal tear," BMJ Open, vol. 10, pp. e038681, 2020.
- [3] C.J. Darrow, C.L. Collins, E.E. Yard, and R.D. Comstock, "Epidemiology of severe injuries among United States high school athletes: 2005–2007," Am. J. Sports Med., vol. 37, pp. 1798–1805, 2009.
- [4] B.E. Gage, N.M. McIlvain, C.L. Collins, S.K. Fields, and R.D. Dawn Comstock, "Epidemiology of 6.6 million knee injuries presenting to United States emergency departments from 1999 through 2008," Acad. Emerg. Med., vol. 19, pp. 378–385, 2012.
- [5] G. Merkely, J. Ackermann, and C. Lattermann, "Articular cartilage defects: Incidence, diagnosis, and natural history," Oper. Tech. Sports Med., vol. 26, pp. 156–161, 2018.

TABLE V						
COMPARISON BETWEEN MODELS FOR BINARY AND MULTI CLASS CLASSIFICATI	ION					

	Classification	Model	Metrics			
			Accuracy	Precision	Recall	F1-Score
-	Binary	Custom CNN	0.88	0.88	0.88	0.88
		Custom CNN + Attention	1.00	1.00	1.00	1.00
	Multi Class	Custom CNN	0.77	0.79	0.77	0.77
		Custom CNN + Attention	0.91	0.91	0.91	0.91

- [6] L. Zhang et al., "Deep Learning Approach for Anterior Cruciate Ligament Lesion Detection: Evaluation of Diagnostic Performance Using Arthroscopy as the Reference Standard," J. Magn. Reson. Imaging, vol. 52, pp. 1745–1752, 2020.
- [7] C.C. Kaeding, B. Léger-St-Jean, and R.A. Magnussen, "Epidemiology and diagnosis of anterior cruciate ligament injuries," Clin. Sports Med., vol. 36, pp. 1–8, 2017.
- [8] P.D. Chang, T.T. Wong, and M.J. Rasiej, "Deep Learning for Detection of Complete Anterior Cruciate Ligament Tear," J. Digit. Imaging, vol. 32, pp. 980–986, 2019.
- [9] D.S. Logerstedt et al., "Knee pain and mobility impairments: Meniscal and articular cartilage lesions: Clinical practice guidelines linked to the international classification of functioning, disability, and health from the orthopaedic section of the American Physical Therapy Association," J. Orthop. Sports Phys. Ther., vol. 40, pp. A1–A35, 2010.
- [10] R.C. Mather III et al., "Societal and economic impact of anterior cruciate ligament tears," J. Bone Jt. Surg. Am. Vol., vol. 95, pp. 1751, 2013.
- [11] K.L. Cameron et al., "Normative values for the KOOS and WOMAC in a young athletic population: History of knee ligament injury is associated with lower scores," Am. J. Sports Med., vol. 41, pp. 582–589, 2013.
- [12] G.R. Huffman et al., "Normative SF-36 values in competing NCAA intercollegiate athletes differ from values in the general population," JBJS, vol. 90, pp. 471–476, 2008.
- [13] K.C. Lam et al., "Previous knee injury and health-related quality of life in collegiate athletes," J. Athl. Train., vol. 52, pp. 534–540, 2017.
- [14] M. Koller et al., "Deep convolutional neural network-based detection of meniscus tears: Comparison with radiologists and surgery as standard of reference," Skelet. Radiol., vol. 49, pp. 1207–1217, 2020.
- [15] E.R. Garwood, R. Tai, and G. Joshi, "The Use of Artificial Intelligence in the Evaluation of Knee Pathology," in *Seminars in Musculoskeletal Radiology*, New York, NY, USA: Thieme Medical Publishers, 2020, pp. 21–29.
- [16] S. Palermi et al., "Indirect Structural Muscle Injuries of Lower Limb: Rehabilitation and Therapeutic Exercise," J. Funct. Morphol. Kinesiol., vol. 6, no. 75, 2021.
- [17] F. Sirico, S. Palermi, B. Massa, and B. Corrado, "Tendinopathies of the hip and pelvis in athletes: A narrative review," *J. Hum. Sports Exerc.*, vol. 15, pp. S748–S762, 2020.
- [18] D. McGonagle, W. Gibbon, and P. Emery, "Classification of inflammatory arthritis by enthesitis," *Lancet*, vol. 352, pp. 1137–1140, 1998.
- [19] I. Hetsroni et al., "Symptomatic pulmonary embolism after outpatient arthroscopic procedures of the knee: The incidence and risk factors in 418 323 arthroscopies," J. Bone Jt. Surg. Br. Vol., vol. 93, pp. 47–51, 2011.
- [20] H.O. Alanazi, A.H. Abdullah, and K.N. Qureshi, "A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care," *J. Med. Syst.*, vol. 41, no. 69, 2017.
- [21] W.D. Prickett, S.I. Ward, and M.J. Matava, "Magnetic resonance imaging of the knee," *Sports Med.*, vol. 31, pp. 997–1019, 2001.
- [22] W. Krampla et al., "MRI of the knee: How do field strength and radiologist's experience influence diagnostic accuracy and interobserver correlation in assessing chondral and meniscal lesions and the integrity of the anterior cruciate ligament?," *Eur. Radiol.*, vol. 19, pp. 1519–1528, 2009.
- [23] R. Mohankumar, L.M. White, and A. Naraghi, "Pitfalls and pearls in MRI of the knee," Am. J. Roentgenol., vol. 203, pp. 516–530, 2014.
- [24] K.L. Cameron et al., "Normative values for the KOOS and WOMAC in a young athletic population: History of knee ligament injury is associated with lower scores," Am. J. Sports Med., vol. 41, pp. 582–589, 2013.

- [25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
- [26] A. Vaswani et al., "Attention is all you need," in Advances in neural information processing systems, vol. 30, 2017.
- [27] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.
- [28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141, 2018.
- [29] S. Sridhar, J. Amutharaj, P. Valsalan, B. Arthi, S. Ramkumar, S. Mathupriya, T. Rajendran, and Y. A. Waji, "A Torn ACL mapping in knee MRI images using deep convolution neural network with Inception-v3," in *Journal of Healthcare Engineering*, 2022.
- [30] A. Wahid, J. A. Shah, A. U. Khan, M. Ullah, and M. Z. Ayob, "Multi-layered basis pursuit algorithms for classification of MR images of knee ACL tear," in *IEEE Access*, vol. 8, pp. 205424-205435, 2020.
 [31] Huang Orthopaedics, "Anterior Cruciate Ligament Tear," [Online].
- [31] Huang Orthopaedics, "Anterior Cruciate Ligament Tear," [Online]. Available: https://www.huangortho.com/anterior-cruciate-ligament-tear. [Accessed: Aug. 9, 2023].
- [32] "PCL Injury: Posterior Cruciate Ligament Tear- Knee Pain Explained," [Online]. Available: https://www.knee-pain-explained.com/ pcl-injury.html. [Accessed: Aug. 9, 2023].
- [33] J. Yau, "MCL Tear: Knee Specialist: Santa Barbara, Santa Maria, Ventura CA," [Online]. Available: https://santabarbarasportsorthopedic.com/ medial-collateral-ligament-injury-mcl-tear-santa-barbara-santa-maria-ventura-ca/. [Accessed: Aug. 9, 2023].