Jurnal Teknik Informatika (JUTIF)
Vol. 5, No. 1, February 2024, pp. 51-60

DOI: https://doi.org/10.52436/1.jutif.2024.5.1.1150
p-ISSN: 2723-3863
e-ISSN: 2723-3871

GENERATIVE ADVERSARIAL NETWORKS FOR ANTERIOR CRUCIATE LIGAMENT INJURY DETECTION

Sri Hasta Mulyani[1], Mohammad Diqi[*2], Husna Arwa Salsabil[3]

[1,2]Dept. of Informatics, Faculty of Science and Technology, Universitas Respati Yogyakarta, Indonesia
[3]Dept. of Physiotherapy, Faculty of Public Health, Universitas Respati Yogyakarta, Indonesia
Email: [1]hasta@respati.ac.id, [2]diqi@respati.ac.id, [3]arwasalsabil@respati.ac.id

(Article received: June 24, 2023; Revision: July 13, 2023; published: January 31, 2024)

Abstract

This research explores the application of Generative Adversarial Networks (GANs) for detecting and classifying Anterior Cruciate Ligament (ACL) injuries using MRI images. The study utilized a dataset of 917 MRI images, each labeled as healthy, partially injured, or completely ruptured, to train the model. The performance of the GAN model was evaluated using a confusion matrix and a classification report, yielding an overall accuracy of 92%. The model demonstrated high proficiency in identifying healthy ACLs and partially injured ACLs but encountered some challenges in accurately identifying completely ruptured ACLs. Despite this, the results suggest that machine learning techniques, particularly GANs, have significant potential for enhancing the accuracy and efficiency of ACL injury detection. The ability of the model to distinguish between different degrees of injury could potentially aid in treatment planning. However, the study also underscores the need for further refinement of the model, particularly in improving its sensitivity in detecting severe ACL injuries. This research highlights the potential of machine learning in medical imaging and provides a solid foundation for future research in ACL injury detection and classification.

Keywords: *ACL Injury Detection, Anterior Cruciate Ligament, Deep Learning, Generative Adversarial Networks, Medical Image Classification, Unsupervised Learning.*

1. INTRODUCTION

The Anterior Cruciate Ligament (ACL) is a crucial component of the knee joint, significantly maintaining knee stability and function. ACL injuries, particularly ruptures, are prevalent among young and active individuals, often leading to long-term physical and psychological impacts [1]. The incidence of ACL injuries varies by sport type, with contact sports and fixed-object, high-impact rotational landing sports presenting a higher risk, especially among female athletes [2]. These injuries can contribute to residual instability in the knee, even after reconstruction, and if left untreated, may result in further damage and the potential for osteoarthritis [3]. Therefore, accurate detection and effective management of ACL injuries are paramount in preserving knee function and optimizing long-term quality of life.

Detecting ACL injuries presents several challenges, particularly in their early stages. Traditional methods, such as clinical examination and patient history, often fail to diagnose these injuries accurately due to their subjective nature and reliance on patient-reported symptoms [1]. Furthermore, while magnetic resonance imaging (MRI) is commonly used for diagnosis, it has limitations in sensitivity and may not adequately visualize certain types of lesions [4]. These limitations can lead to underdiagnosis or misdiagnosis of ACL injuries, potentially resulting in inappropriate treatment strategies and poorer patient outcomes. Therefore, there is a pressing need for more objective and reliable methods for detecting ACL injuries.

Because of its excellent sensitivity and specificity [4], MRI is frequently utilized for detecting ACL damage]. A few injuries, including those to the Kaplan fiber complex (KFC), have been observed to be initially missed on preoperative MRI scans [5]. This is essential because damage to the KFC can cause ACL-deficient knees to develop anterolateral rotatory instability (ALRI), affecting rehabilitation efforts' success or failure.

Various challenges, including visualization issues, pitfalls of different imaging techniques, and diagnosing different injury types, can impact ACL injury detection and treatment. Visualization problems may arise when assessing ACL injuries. Interpreting MRI images can be difficult for less experienced medical personnel [6]. The complex anatomy of the knee joint and the ACL's small size can hinder injury visualization and identification [7].

Different imaging techniques have their limitations when detecting ACL injuries. Radiography helps rule out fractures but might not provide enough detail for soft tissue injuries like ACL tears [8]. Ultrasound has difficulties visualizing deep structures like the ACL [7]. CT scans can provide detailed bony structure images but may not be effective in evaluating soft tissue injuries like ACL tears [6].

Diagnosing ACL injury types can be challenging. ACL injuries vary in severity from partial tears to complete ruptures. Identifying these injuries is crucial for determining treatment strategies. Meniscal tears often accompany ACL injuries and can complicate diagnosis [7].

Misdiagnosed or late-diagnosed ACL injuries have significant real-world implications, affecting patients, healthcare resources, and economic burdens. Delayed treatment can result in prolonged pain, functional limitations, and increased knee joint damage risk [9]. It can also delay rehabilitation, impact the quality of life, and increase the likelihood of secondary injuries [10].

The economic impact of misdiagnosed or late-diagnosed ACL injuries is substantial. Surgical intervention, rehabilitation, and long-term follow-up care are often required. Surgery costs, including hospitalization, anesthesia, and postoperative care, can be significant [11]. Moreover, economic burdens continue beyond the initial treatment, as untreated or misdiagnosed patients may need additional medical interventions [10]. These costs strain healthcare resources and increase overall healthcare expenditure.

Additionally, misdiagnosed or late-diagnosed ACL injuries may increase healthcare resource utilization. Patients might need multiple visits to healthcare providers for accurate diagnosis and treatment planning [9]. This increased utilization may result in longer waiting times, impacting healthcare system efficiency.

The potential of machine learning in predicting and detecting ACL injuries has been explored in several studies. Jauhiainen et al. [12] applied machine learning techniques to predict ACL injury risk among female elite handball and soccer players, although the predictive ability was noted to be low from a clinical assessment standpoint. Similarly, Taborri et al. [13] utilized a machine-learning approach to assess ACL injury risk in female basketball players, demonstrating high accuracy and F1-score. In another study, Mazlan et al. [14] employed a support vector machine (SVM) to classify multi-class ACL injury data, achieving an accuracy of up to 100%. Stajduhar et al. [15] investigated the use of machine learning for detecting milder ACL injuries and complete ACL ruptures using MRI data, resulting in the high area under the curve (AUC) scores. Collectively, these studies underscore the potential utility of machine learning as a tool for predicting and detecting ACL injuries.

There has been a growing interest in applying deep learning models for detecting ACL injuries. Several studies have demonstrated the effectiveness of Convolutional Neural Networks (CNNs) in diagnosing and classifying ACL injuries. For instance, Namiri et al. [16] reported high sensitivity and specificity of 2D and 3D CNNs in classifying ACL injuries using MRI images. Similarly, Razali [17] developed a CNN system that achieved an accuracy of 94.7% in classifying ACL injuries based on MRI images. Koga et al. [18] used video analysis to propose a new hypothesis for ACL injury mechanisms, suggesting a focus on specific landing techniques for prevention programs. Carlson et al. [19] also used video analysis to determine that high-risk landing positions significantly influence the likelihood of ACL injuries, emphasizing the need for broad application of preventative training.

In diagnosing ACL tears from knee MRI images, CNNs have shown promising results. Both Shin et al. [20] and Sridhar et al. [21] developed CNN models that accurately diagnose ACL tears. Jaturapisanukul and Pangarad [22] modified the feature extraction module of an existing CNN model to maintain diagnostic accuracy while reducing computational resources. Germann et al. [23] found that the performance of a CNN model in diagnosing ACL tears could approach that of fellowship-trained musculoskeletal radiologists, although performance may decrease with increasing MRI examination heterogeneity. These studies underscore the potential of CNNs as a valuable tool for diagnosing ACL tears from knee MRI images. Meniscus and patellofemoral cartilage lesions in people with osteoarthritis and ACL injuries were studied by Pedoia et al. [4], who assessed the efficacy of 3D convolutional neural networks (3D-CNN) in detecting and grading injury severity. The study showed that deep learning models can effectively detect meniscus lesions, with a sensitivity of 89.81% and a specificity of 81.98%.

Despite the advancements in ACL injury detection, there are still limitations to the current methods. For instance, the sensitivity of MRI in detecting specific injuries such as meniscal ramp lesions is low, leading to underdiagnosis [3]. Additionally, the grading schemes utilized in research are often not used in clinical practice due to their time-consuming nature, further highlighting the need for improved detection methods [4].

Because of these drawbacks, researchers are considering applying sophisticated machine learning methods, including Generative Adversarial Networks (GANs), to the challenge of ACL injury diagnosis. Goodfellow et al. [24] were the first to create AI models called GANs, which are used to generate new data instances similar to the training data. The first component, the generator, generates new data instances, while the second component, the discriminator, determines whether or not these instances are genuine additions to the original training dataset [25].

Deep learning and GANs have been used in medical image analysis to enhance the performance of computer-aided diagnosis systems. They proved their utility in various tasks, including liver lesion classification [26], closed-angle detection in anterior segment optical coherence tomography [25], and melanoma lesion segmentation [27]. Increased accuracy and performance in medical image analysis tasks were reported in these research studies.

GANs are also employed for data augmentation in medical imaging. They can generate synthetic medical images, thereby enlarging the size and diversity of training datasets. This expansion can elevate the performance of deep learning models [26]. Studies on liver lesion classification1 and prostate cancer magnetic resonance imaging [28] have demonstrated this.

Alongside data augmentation, GANs function in image reconstruction within medical imaging. They have been used to escalate the quality of knee plain radiography images [29] and to reconstruct superior-quality artificial prostate cancer magnetic resonance images [28]. The primary aim of deploying GANs in image reconstruction is to upgrade image quality and furnish more precise and comprehensive information for medical diagnosis and treatment planning.

Additionally, GANs are used for modality conversion in medical imaging. This procedure involves converting images from one modality to another, such as transmuting magnetic resonance images to computed tomography images or vice versa [30]. Modality conversion with the help of GANs can mitigate limitations and reinforce the strengths of different imaging modalities, hence providing valuable information for medical professionals.

In the context of ACL injury detection, GANs could be used to augment existing imaging datasets, thereby improving the robustness of the detection models. Furthermore, GANs could be used to generate synthetic yet realistic images of ACL injuries, which could aid in training other machine learning models for injury detection. While the application of GANs in this area is still in its early stages, preliminary studies in related fields suggest they hold significant promise for improving the accuracy and efficiency of ACL injury detection [31].

The primary objective of this study is to explore the application of GANs for the detection of ACL injuries. Specifically, we aim to develop a GAN model that can accurately classify ACL injuries into three categories: healthy, partially injured, and completely ruptured, based on data from a Kaggle dataset. We also aim to evaluate the performance of our model using a confusion matrix and classification report metrics. Through this research, we hope to contribute to the existing knowledge on ACL injury detection and potentially pave the way for more accurate and efficient diagnostic methods in the future.

The remainder of this paper is organized as follows: The dataset, preparation methods, GAN model architecture, training procedure, and assessment measures will all be described in the Methods section. In the Results section, we will show our confusion matrix and classification report, along with how well our model performed on the test data. In the next section, we will provide an analysis of the data, a discussion of its implications for ACL injury detection, an admission of the study's shortcomings, and recommendations for future investigation. Finally, we will present a summary of the significant findings of our study, an analysis of their potential implications for ACL injury identification, and some concluding remarks in the Conclusion section.

2. RESEARCH METHODOLOGY

2.1. Dataset

The KneeMRI dataset [15] was downloaded from Kaggle and used in this study. Exams performed on a Siemens Avanto 1.5T MR scanner at the Clinical Hospital Centre Rijeka, Croatia, between 2006 and 2014 were used to compile this dataset. Proton density-weighted fat suppression imaging was utilized for this purpose. There are 917 volumes of data, each containing a single image of the left or right knee in 12-bit grayscale. Each volume record in the dataset was double-blindly labeled with a diagnostic about the health of the ACL. This means that two evaluators, blind to one another's ratings, examined each volume record.

The labels assigned to each volume record correspond to the condition of the ACL and are categorized as follows: (1) healthy, indicating no injury to the ACL; (2) partially injured, indicating some degree of injury to the ACL but not a complete rupture; and (3) completely ruptured, indicating a full tear of the ACL. These labels serve as the ground truth for training and evaluating our machine learning model.

2.2. Preprocessing

Before feeding the data into the GAN model, several preprocessing steps were undertaken to ensure the data was in a suitable format for analysis.

Data Cleaning: The first step involved cleaning the data to remove any inconsistencies or errors. This included checking for and handling missing values and ensuring that all volume records were correctly labeled according to the condition of the ACL.

Normalization: The grayscale images in the dataset were normalized to ensure that the pixel intensity values fell within a standard range. This is an important step as it helps reduce the data's scale and prevent the model from being biased towards features with higher magnitudes. The pixel intensities, originally ranging from 0 to 4095 due to the 12-bit depth, were scaled to fall within the range of 0 to 1.

Data Augmentation: Data augmentation methods were used to boost the quantity and variety of the dataset. Specifically, this required applying transformations such as rotations, translations and flips to the source images to generate new data instances. As a result, the model can improve its generalization capacity by learning to recognize ACL situations in various orientations and positions and benefiting from the more significant amount of available training data.

Data Splitting: Finally, we divided the data into training, validation, and test sets. The model is trained using the training set, validated using the validation set to fine-tune the model parameters and prevent overfitting, and tested using a separate dataset.

These preprocessing steps are crucial in preparing the data for analysis and ensuring that the GAN model can learn effectively from the data.

2.3. GAN Architecture

The architecture of the GAN used in this study consists of two main components: the generator and the discriminator, as shown in Figure 1.

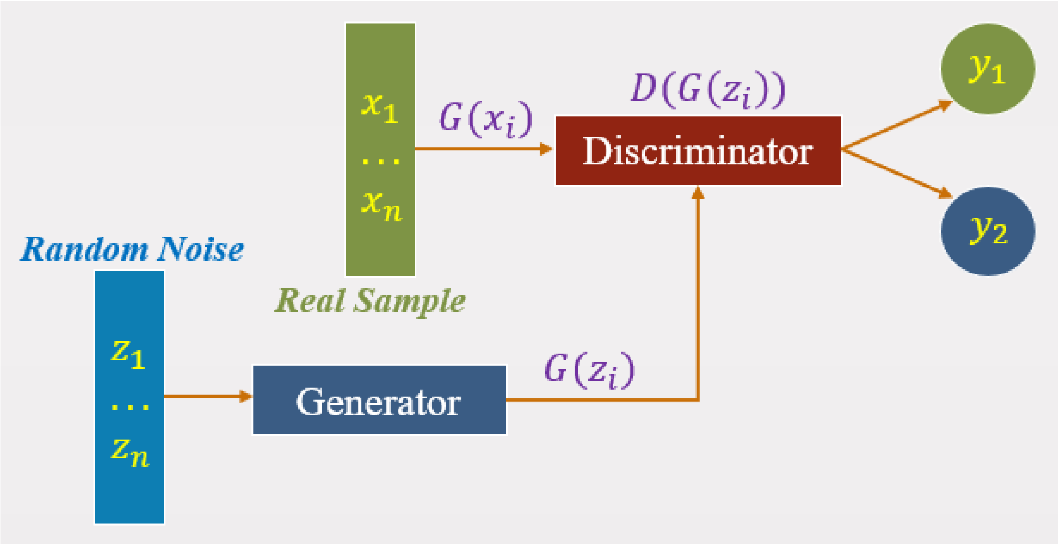

Figure 1. GAN Architecture

In order to create fresh data instances similar to the training data, the generator employs a neural network. Using deconvolutional (or transposed convolutional) layers, it accepts a random noise vector as input and outputs data instances. In order to fool the discriminator, the generator must come up with fake data that looks just like the genuine thing. As seen in Equation 1, the loss function of a generator is commonly defined as the negative log-likelihood of fooling the discriminator.

$$\mathcal{L}_G = -\log\left(D\left(G(z)\right)\right) \qquad (1)$$

where D is the discriminator, G is the generator, and z is the random noise vector.

The discriminator is a second neural network that tries to tell the difference between training set instances of genuine data and generator-created fakes. It takes in a data instance and spits out the likelihood that it represents the real data distribution. The discriminator is taught to prioritize accuracy over speed when determining whether an input is genuine. Equation 2 expresses the discriminator's loss function.

$$\mathcal{L}_D = -\log\left(D(x)\right) - \log\left(1 - D\left(G(z)\right)\right) \qquad (2)$$

where x is a real data instance.

In a two-player minimax game, the generator and discriminator are trained simultaneously, with the generator attempting to mislead the discriminator while the discriminator strives to categorize actual and fake data instances accurately. In Equation 3, we may express the GAN's overarching goal as a single number.

$$\min_G \max_D \mathcal{L}_D + \mathcal{L}_G \qquad (3)$$

2.4. Training and Testing

This study's GAN was trained using a two-step iterative process involving the discriminator and the generator. The training was performed over 1000 epochs, each consisting of several mini-batches of size 128.

The first step of each cycle involved training the discriminator. This was accomplished by feeding the generator a stream of random noise in images. These synthetic images were then fused with authentic images from the dataset used for training. Using the binary cross-entropy loss function, the discriminator was taught to determine whether the photos were authentic. For binary classification problems, this loss function quantifies the discrepancy between the discriminator's predictions and the correct labels.

Each iteration's second phase involved training the generator. In order to accomplish this, we needed to generate another set of images using noise. In contrast to the discriminator training phase, the images in this phase had the 'real' label applied to them. We next used the same images and labels to train the combined generator-discriminator model using the binary cross-entropy loss function. The goal at this stage was adjusting the generator weights to generate images that the discriminator would correctly label as real.

This two-stage procedure was performed on the training dataset every time a new epoch began. Each epoch's batch-specific discriminator and generator losses were computed and displayed. The alternating discriminator and generator training stabilized the training process to prevent one component from becoming overly potent. This method is frequently employed to keep the generator and discriminator in check during GAN training. Following the training procedure, the generator was then used to produce images for each class.

The subsequent stage involves deploying the trained model on the testing data. The model will examine the features of the testing data and generate predictions based on the knowledge acquired during the training stage. These predictions are subsequently compared to the actual values for each test data record, commonly referred to as the "ground truth." This step aims to assess the accuracy of the model's predictions.

2.5. Evaluation Metrics

The evaluation of our model will be based on two key metrics: the confusion matrix and the classification report. These metrics were chosen because they provide a comprehensive overview of the model's performance across all classes (healthy, partially injured, and completely ruptured).

The performance of an algorithm can be visualized using the confusion matrix, which is a table layout. Instances belonging to one predicted class are represented by rows in the matrix, while columns represent those belonging to another actual class. The name comes from the fact that it is simple to tell if the system is mixing up two types of data (in other words, if it frequently mistakes one for the other).

Each class's precision, recall, F1-score, and support are detailed in the classification report. The F1-score can be understood as a weighted harmonic mean of the accuracy and recall, where precision is the classifier's ability to avoid incorrectly labeling a sample as positive, and recall is the classifier's ability to locate all positive samples. The number of instances of each class in the actual response values is called support.

In mathematical terms, for binary classification:

a. Precision (P) is defined as

$$P = \frac{TP}{TP+FP} \tag{4}$$

where TP is the number of true positives, and FP is the number of false positives.

b. Recall (R) is defined as

$$R = \frac{TP}{TP+FN} \tag{5}$$

where FN is the number of false negatives.

c. F1-score is the harmonic mean of precision and recall, defined as

$$F1 = 2 \cdot \frac{P \cdot R}{P+R} \tag{6}$$

These metrics will be extended to the multi-class case in our study, considering each class against the rest.

All stages of this research can be visualized in Figure 2.

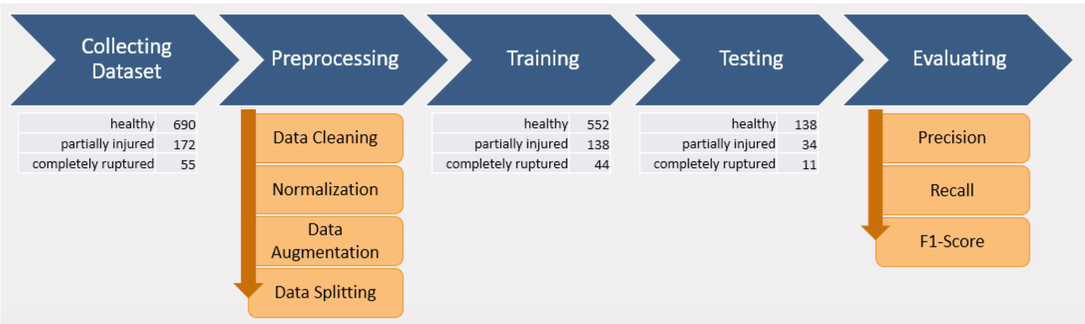

Figure 2. The stages of the research

3. RESULTS

3.1. Initial Dataset

The initial dataset contains 917 grayscale images representing either left or right knees. Each volume record includes various information about specific features such as examId, seriesNo, aclDiagnosis, kneeLR, roiX, roiY, roiZ, roiHeight, roiWidth, roiDepth, and volumeFilename. The aclDiagnosis feature indicates the diagnosis assigned to each volume record regarding the condition of the anterior cruciate ligament, which can be categorized as (1) healthy, (2) partially injured, or (3) completely ruptured.

After the data cleaning, normalization, and data augmentation processes, we split 80% of the data for training and 20% for testing, with the distribution shown in Table 1.

Table 1. Data distribution for training and testing

Class	Total	Training	Testing
healthy	690	552	138
partially injured	172	138	34
completely ruptured	55	44	11

3.2. Model Performance

The performance of our GAN model on the test data was evaluated using a confusion matrix and a classification report. The confusion matrix for our model is shown in Figure 3.

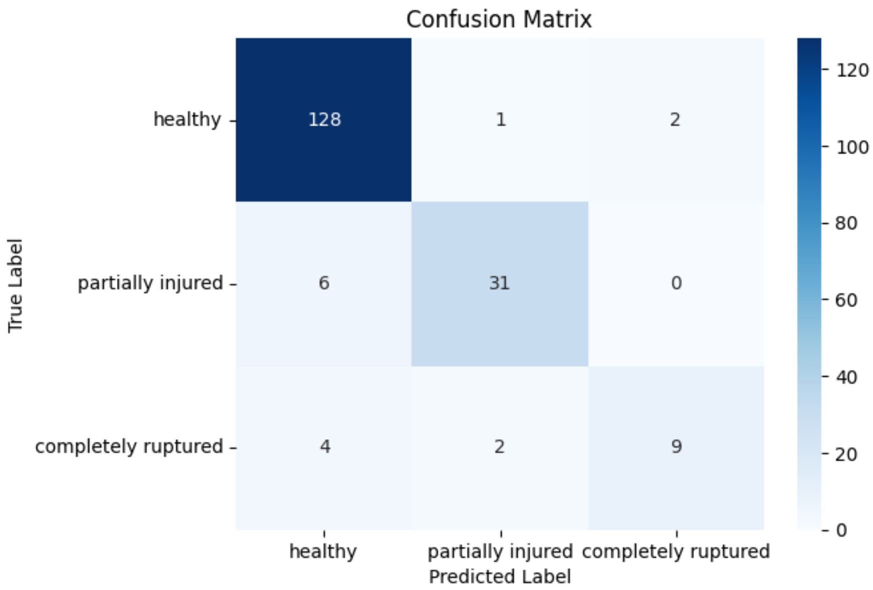

Figure 3. Confusion Matrix

The classification report for our model is described in Table 2.

Table 2. Classification report

class	precision	recall	f1-score	support
healthy	0.93	0.98	0.95	131
partially injured	0.91	0.84	0.87	37
completely ruptured	0.82	0.60	0.69	15
accuracy			0.92	183
macro avg	0.89	0.80	0.84	183
weighted avg	0.92	0.92	0.91	183

4. DISCUSSIONS

4.1. Result Analysis

The confusion matrix shows that our model is highly accurate in predicting healthy ACLs with 128 correct predictions out of 131 healthy cases. The model correctly predicted 31 out of 37 cases of partially injured ACLs. The model correctly predicted 9 out of 15 cases of completely ruptured ACLs.

The model shows promising results in detecting healthy and partially injured ACLs but has trouble correctly identifying severely torn ACLs, as shown by the precision, recall, and F1-score for each class. The model accurately categorized 92% of the instances in the validation set, as indicated by the overall accuracy of 0.92. Both the overall average and the weighted average scores are very indicative of the model's superior performance across all classes.

4.2. Interpretation

The model results indicate a high degree of accuracy in predicting the ACL condition from the MRI images. The model's overall accuracy is 92%, suggesting that it correctly classified the ACL condition in 92% of the cases in the test data.

Looking at the confusion matrix, the model shows a strong performance in identifying healthy ACLs, with 128 correct predictions out of 131 actual healthy cases. This suggests that the model effectively distinguishes healthy ACLs from injured ones.

The model correctly predicted 31 out of 37 cases of partially injured ACLs. While this is a relatively high accuracy, the six misclassifications indicate that there may be some overlap in the features of healthy and partially injured ACLs that the model is struggling to distinguish.

The model had the most difficulty correctly identifying completely ruptured ACLs, predicting only 9 out of 15 cases. This suggests that the model may struggle to distinguish between partially injured and completely ruptured ACLs, possibly due to similarities in their appearance in MRI images.

The classification report provides further insights into the model's performance. The precision, recall, and F1-score for each class indicate that the model performs well in identifying healthy and partially injured ACLs but has some difficulty with completely ruptured ACLs. This is reflected in the lower recall and F1-score for the completely ruptured class, indicating that the model is less sensitive and accurate in identifying this condition.

Overall, these results suggest that while the model performs well in general, there may be room for improvement in its ability to distinguish between different degrees of injury. Future work could focus on refining the model or incorporating additional features to improve its performance in this area.

4.3. Implications of findings

The findings of this study have several important implications for detecting ACL injuries.

Firstly, the high accuracy of the model in identifying healthy ACLs suggests that machine learning techniques, specifically GANs, can be effectively used to rule out ACL injuries in patients. This could potentially reduce the need for unnecessary treatments or interventions in patients with healthy ACLs.

Secondly, the model's ability to distinguish between partially injured and completely ruptured ACLs, although imperfect, indicates that machine learning models can potentially be used to grade the severity of ACL injuries. This could aid in treatment planning, as different grades of injury may require different treatment approaches.

However, the model's lower performance in identifying completely ruptured ACLs suggests that further work is needed to improve the sensitivity of machine learning models in detecting severe ACL injuries. This could involve refining the model architecture, incorporating additional features, or using larger and more diverse training datasets.

These findings suggest that machine learning models hold significant promise for improving the accuracy and efficiency of ACL injury detection. However, it is essential to note that these models should complement, rather than replace, clinical judgment and other diagnostic tools. Further research and validation studies are needed to realize these models' potential in clinical practice fully.

4.4. Limitations

While our study presents promising results, it is important to acknowledge several limitations that may have influenced our findings.

Dataset Limitations: The dataset used in this study was retrospectively collected from a single hospital center, which may limit the generalizability of our findings. The MRI images were all obtained using the same scanner and imaging technique, which may not reflect the variability in imaging quality and techniques used in other settings. Furthermore, the dataset is relatively small and imbalanced, with fewer partially injured and completely ruptured ACLs than healthy ACLs. This could potentially bias the model towards predicting the majority class (healthy ACLs).

GAN Architecture Limitations: While GANs have shown promise in various applications, they are not without their limitations. GANs can be difficult to train due to the challenging nature of the minimax optimization problem, which can lead to issues such as mode collapse, where the generator produces a limited variety of samples. Additionally, the black-box nature of GANs makes it difficult to interpret their decision-making process, which can be a critical factor in medical applications.

Evaluation Metrics Limitations: While often utilized in classification issues, the evaluation measures employed in this study may not accurately reflect the model's actual performance. Precision, recall, and F1-score are all binary measures extended to the multi-class scenario in our research; accuracy, on the other hand, might be deceptive in imbalanced datasets.

Lack of Clinical Validation: Finally, while our model showed high accuracy in a test set, it was not validated in a clinical setting. The model's performance in real-world conditions, with varying image quality and patient populations, remains to be seen.

These limitations highlight future work and improvement areas, including using larger and more diverse datasets, exploring different GAN architectures or other machine learning models, using additional or alternative evaluation metrics, and conducting clinical validation studies.

4.5. Future Work

The results of this study open several avenues for future research and potential improvements to the model.

Larger and More Diverse Datasets: Future studies could use larger and more diverse datasets, including MRI images from different hospital centers using other scanners and imaging techniques, to improve the model's generalizability. This could also help to address the imbalance in the dataset, providing more examples of partially injured and completely ruptured ACLs for the model to learn from.

Improved GAN Architecture: While the GAN architecture used in this study showed promising results, there is always room for improvement. Future work could explore different GAN architectures or modifications to the existing architecture to improve performance. This could include conditional GANs, which could improve the model's ability to distinguish between different degrees of injury.

Alternative Machine Learning Models: Besides GANs, other machine learning models could be explored for ACL injury detection. This could include different deep learning models, such as Convolutional Neural Networks (CNNs), or traditional machine learning models, such as Support Vector Machines (SVMs).

Additional Evaluation Metrics: Additional or alternative evaluation criteria could be used in future research to provide a more in-depth assessment of the model's performance. This may involve measurements like the Matthews Correlation Coefficient (MCC) or the Area Under the Receiver Operating Characteristic Curve (AUROC).

Clinical Validation: Finally, a critical step in future research will be to validate the model in a clinical setting. This would involve testing the model's performance in real-world conditions, with varying image quality and patient populations, to determine its suitability for clinical use.

By addressing these areas, future research can continue to advance the field of ACL injury detection and contribute to the development of more accurate and efficient diagnostic tools.

5. CONCLUSION

Our research explored the application of GANs for detecting ACL injuries. Using a dataset of MRI images, our model was trained to classify ACL conditions into three categories: healthy, partially injured, and completely ruptured. The model's performance was evaluated using a confusion matrix and a classification report, demonstrating an overall accuracy of 92%.

The high accuracy of our model suggests that machine learning techniques, specifically GANs, hold significant promise for improving the accuracy and efficiency of ACL injury detection. The ability of our model to distinguish between different degrees of injury could potentially aid in treatment planning, as various grades of injury may require different treatment approaches. However, the model's lower performance in identifying completely ruptured ACLs indicates that further work is needed to improve the sensitivity of machine-learning models in detecting severe ACL injuries.

While our findings are promising, they also highlight the challenges in applying machine learning models in medical imaging, such as the need for extensive and diverse datasets, the difficulty in training GANs, and the need for clinical validation. Despite these challenges, our work represents a significant step forward in applying machine learning for ACL injury detection. We hope our research will inspire further studies in this area, contributing to developing more accurate and efficient diagnostic tools for ACL injuries.

REFERENCES

[1] S. R. Filbay and H. Grindem, “Evidence-based recommendations for the management of anterior cruciate ligament (ACL) rupture,” *Best Practice & Research Clinical Rheumatology*, vol. 33, no. 1, pp. 33–47, Feb. 2019, doi: 10.1016/j.berh.2019.01.018.

[2] A. M. Montalvo *et al.*, “Anterior Cruciate Ligament Injury Risk in Sport: A Systematic Review and Meta-Analysis of Injury Incidence by Sex and Sport Classification,” *Journal of Athletic Training*, vol. 54, no. 5, pp. 472–482, Apr. 2019, doi: 10.4085/1062-6050-407-16.

[3] B. Sonnery-Cottet, R. Serra Cruz, T. D. Vieira, R. A. Goes, and A. Saithna, “Ramp Lesions: An Unrecognized Posteromedial Instability?,” *Clinics in Sports Medicine*, vol. 39, no. 1, pp. 69–81, Jan. 2020, doi: 10.1016/j.csm.2019.08.010.

[4] V. Pedoia, B. Norman, S. N. Mehany, M. D. Bucknor, T. M. Link, and S. Majumdar, “3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects,” *Journal of Magnetic Resonance Imaging*, vol. 49, no. 2, pp. 400–410, Feb. 2019, doi: 10.1002/jmri.26246.

[5] D. P. Berthold *et al.*, “High Rate of Initially Overlooked Kaplan Fiber Complex Injuries in Patients With Isolated Anterior Cruciate Ligament Injury,” *Am J Sports Med*, vol. 49, no. 8, pp. 2117–2124, Jul. 2021, doi: 10.1177/03635465211015682.

[6] K.-H. Chen, C. Y. Yang, H.-Y. Wang, H.-L. Ma, and O. K. Lee, “Artificial Intelligence–Assisted Diagnosis of Anterior Cruciate Ligament Tears From Magnetic Resonance Images: Algorithm Development and Validation Study,” *Jmir Ai*, 2022, doi: 10.2196/37508.

[7] F. R. Espinosa-Leal and K. V. Rodriguez-Alanis, “MR Imaging of Meniscal Tears Associated With Traumatic Anterior Cruciate Ligament Injury,” *Journal of the Mexican Federation of Radiology and Imaging*, 2023, doi: 10.24875/jmexfri.m23000043.

[8] L. Stijak and P. Schai, “Is There an Influence of the Tibial Slope of the Lateral Condyle on the ACL Lesion?,” *Knee Surgery Sports Traumatology Arthroscopy*, 2007, doi: 10.1007/s00167-007-0438-1.

[9] M. C. van der Steen, M. Reijman, and R. P. A. Janssen, “Diagnostic Values of History Taking, Physical Examination and KT-1000 Arthrometer for Suspect Anterior Cruciate Ligament Injuries in Children and Adolescents: A Prospective Diagnostic Study,” *BMC Musculoskeletal Disorders*, 2022, doi: 10.1186/s12891-022-05659-1.

[10] M. H. Hagmeijer *et al.*, “Secondary Meniscal Tears in Patients With Anterior Cruciate Ligament Injury: Relationship Among Operative Management, Osteoarthritis, and Arthroplasty at 18-Year Mean Follow-Up,” *The American Journal of Sports Medicine*, 2019, doi: 10.1177/0363546519844481.

[11] K. M. Sunbol *et al.*, “Epidemiology of Recurrent Anterior Cruciate Ligament Injuries According to Sport Type,” *International Journal of Community Medicine and Public Health*, 2022, doi: 10.18203/2394-6040.ijcmph20222087.

[12] S. Jauhiainen, J.-P. Kauppi, T. Krosshaug, R. Bahr, J. Bartsch, and S. \" Ayr\" am\" o, “Predicting ACL Injury Using Machine Learning on Data From an Extensive Screening Test Battery of 880 Female Elite Athletes,” *The American Journal of Sports Medicine*, vol. 50, no. 11, pp. 2917–2924, Aug. 2022, doi: 10.1177/03635465221112095.

[13] J. Taborri, L. Molinaro, A. Santospagnuolo, M. Vetrano, M. C. Vulpiani, and S. Rossi, “A machine-learning approach to measure the anterior cruciate ligament injury risk in female basketball players,” *Sensors*, vol. 21, no. 9, 2021, doi: 10.3390/s21093141.

[14] S. S. Mazlan, M. Z. Ayob, and Z. A. Kadir Bakti, “Anterior cruciate ligament (ACL) injury classification system using support vector machine (SVM),” in *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)*, IEEE, Sep. 2017. doi: 10.1109/ice2t.2017.8215960.

[15] I. Štajduhar, M. Mamula, D. Mileti\’ c, and G. ozde \" Unal, “Semi-automated detection of anterior cruciate ligament injury from MRI,” *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 151–164, Mar. 2017, doi: 10.1016/j.cmpb.2016.12.006.

[16] N. K. Namiri *et al.*, “Deep Learning for Hierarchical Severity Staging of Anterior Cruciate Ligament Injuries from MRI,” *Radiology: Artificial Intelligence*, vol. 2, no. 4, p. e190207, Jul. 2020, doi: 10.1148/ryai.2020190207.

[17] M. H. Razali, S. M. Sazwan, M. Mahmood, D. Nazri, J. Ali, and M. Z. Ayob, “Anterior cruciate ligament (ACL) coronal view injury diagnosis system using convolutional neural network,” in *ACM International Conference Proceeding Series*, 2019. doi: 10.1145/3362752.3365196.

[18] H. Koga, T. Muneta, R. Bahr, L. Engebretsen, and T. Krosshaug, “ACL Injury Mechanisms: Lessons Learned from Video Analysis,” in *Rotatory Knee Instability*, Springer International Publishing, 2016, pp. 27–36. doi: 10.1007/978-3-319-32070-0_3.

[19] V. R. Carlson, F. T. Sheehan, and B. P. Boden, “Video Analysis of Anterior Cruciate Ligament (ACL) Injuries,” *JBJS Reviews*, vol. 4, no. 11, Nov. 2016, doi: 10.2106/jbjs.rvw.15.00116.

[20] H. Shin, G. S. Choi, and M. C. Chang, “Development of convolutional neural network model for diagnosing tear of anterior cruciate ligament using only one knee magnetic resonance image,” *Medicine*, vol. 101, no. 44, p. e31510, Nov. 2022, doi: 10.1097/md.0000000000031510.

[21] S. Sridhar *et al.*, “A Torn ACL Mapping in Knee MRI Images Using Deep Convolution Neural Network with Inception-v3,” *Journal of Healthcare Engineering*, vol. 2022, pp. 1–9, Feb. 2022, doi: 10.1155/2022/7872500.

[22] P. Jaturapisanukul and A. Pangarad, “Reduction of computing resources in convolutional neural network for knee MRI of ACL tears by feature-based method,” in *Twelfth International Conference on Digital Image Processing (ICDIP 2020)*, H. Fujita and X. Jiang, Eds., SPIE, Jun. 2020. doi: 10.1117/12.2573234.

[23] C. Germann *et al.*, “Deep Convolutional Neural Network-Based Diagnosis of Anterior Cruciate Ligament Tears: Performance Comparison of Homogenous Versus Heterogeneous Knee MRI Cohorts with Different Pulse Sequence Protocols and 1.5-T and 3-T Magnetic Field Strengths,” *Investigative Radiology*, vol. 55, no. 8, 2020, doi: 10.1097/RLI.0000000000000664.

[24] I. Goodfellow *et al.*, “Generative Adversarial Networks,” *Communications of the Acm*, 2020, doi: 10.1145/3422622.

[25] C. Zheng *et al.*, “Assessment of Generative Adversarial Networks for Synthetic Anterior Segment Optical Coherence Tomography Images in Closed-Angle Detection,” *Translational Vision Science & Technology*, 2021, doi: 10.1167/tvst.10.4.34.

[26] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “GAN-based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification,” *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2018.09.013.

[27] Z. Ali, S. Naz, H. Zaffar, J. Choi, and Y. Kim, “An IoMT-Based Melanoma Lesion Segmentation Using Conditional Generative Adversarial Networks,” *Sensors*, 2023, doi: 10.3390/s23073548.

[28] I. R. L. Xu *et al.*, “Generative Adversarial Networks Can Create High Quality Artificial Prostate Cancer Magnetic Resonance Images,” *Journal of Personalized Medicine*, 2023, doi: 10.3390/jpm13030547.

[29] G. Ahn *et al.*, “High-resolution Knee Plain Radiography Image Synthesis Using Style

Generative Adversarial Network Adaptive Discriminator Augmentation," *Journal of Orthopaedic Research®*, 2022, doi: 10.1002/jor.25325.

[30] K. Koshino *et al.*, "Narrative Review of Generative Adversarial Networks in Medical and Molecular Imaging," *Annals of Translational Medicine*, 2021, doi: 10.21037/atm-20-6325.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.