CNN modeling: Classifying ACL tears, meniscus tears, and abnormalities from MRI exams using AlexNet and ResNet

Nadeem Patel *

4820 N. Central Park Ave, Chicago, IL, 60625

* nadeemp.101@gmail.com

Abstract

While magnetic resonance imaging (MRI) of knees is the preferred method for diagnosing knee injuries, the interpretation of knee MRI is time-intensive and diagnostic error, as well as variability, is possible (Bien et al. 2018). Human-based image interpretation can have pitfalls like subjectivity, distraction, fatigue, and diagnostic uncertainties, which can lead to erratic diagnosis and suboptimal management of knee injuries (Siouras et al. 2022). Deep learning (DL) methods can assist in diagnosing knee injuries since such models can be developed to handle complex relationships between medical images and their interpretations by learning layers of features. More specifically, MRI studies frequently analyze convolutional neural network (CNN) architectures because such models have the ability to learn complex representations and improve pattern recognition from raw data instead of having human engineering and domain expertise to structure data and desire feature extractors (Aggarwal et al. 2021). Based on previous research, this study also examines CNN models, including the Alex Krizhevsky architecture (AlexNet) and residual network (ResNet). Using the MRNet dataset from the Stanford ML Group, the goal is to train CNN models and determine which can be used to assist medical image analysis, leading to improved accuracy and efficiency of diagnostic and treatment processes (Chan et al. 2020).

Introduction

Although knee magnetic resonance imaging (MRI) is the preferred method for diagnosing knee injuries, it is a time-consuming process and there is a possibility of diagnostic errors and variability in interpretation (Bien et al. 2018). When image interpretation is done by humans, there are risks of subjectivity, distraction, fatigue, and diagnostic uncertainties, which can result in inaccurate diagnosis and inadequate treatment of knee injuries (Siouras et al. 2022). Deep learning (DL) has seen immense success in pattern recognition applications in health care. When applied to lesion detection or classification, DL approaches have shown greater performance compared to those by conventional techniques and better than radiologists in certain tasks (Chan et al. 2020). Particularly, MRI investigations frequently utilize CNN designs because these models possess the capacity to acquire intricate representations and enhance pattern identification from unprocessed data, rather than depending on human engineering and expertise in domain to organize data and extract desired features (Aggarwal et al. 2021).

This study analyzes CNN models, including Alex Krizhevsky architecture (AlexNet) and residual network (ResNet), to determine which model provides the best accuracy rate when it comes to analyzing MRI exams. The selection of model architectures is based on previous studies, and through a comparative approach, the models can be compared to determine how each develops its results and why a certain model performs better than others. While addressing which model can best assist medical professionals with MRI diagnostics and treatment processes, the study also aims to analyze the structural details of the models to determine how they can be configured moving forward for optimal results.

Radiologists and other medical professionals rely on MRI because it provides a noninvasive way to examine organs, tissues, and the skeletal system (Mayo Clinical Staff n.d.). However, MRI exams can be misinterpreted by experts, which can lead patients down the incorrect path and delay critical treatment (Expert MRI 2022). Utilizing artificial intelligence (AI), and more specifically, deep learning models, can help to reduce risks of misdiagnosis and improper treatment if done correctly. Determining the best approach for such an implementation is also crucial since a new approach can also be a step backwards. Therefore, analyzing and comparing multiple models can provide insight into what deep learning model works best for MRI analysis.

Literature Review

Application of deep learning has proven to be successful in various studies focused on MRI scans. Such studies highlight the possibilities and drawbacks of using AI compared to traditional methods. A study published to the BMC Musculoskeletal Disorders journal aimed to create a CNN model to detect and classify meniscus tears using coronal and sagittal magnetic resonance (MR) images. The DL model was built using 500 cases with meniscus tears and 449 cases without tears. The model's performance was measured using the training and test sets, and the results showed that the CNN model has potential to diagnose the presence and types of meniscal tears. The area under the curves (AUCs) were calculated for medial meniscal tears, lateral meniscal tears, and medial and lateral meniscal tears. AUCs for horizontal, complex, radial, and longitudinal tears were also collected. The experiment was able to successfully analyze the various angles of an MRI using the AUC metric to determine the performance of the model. The research suggests that the CNN model has potential to be used in diagnosing the presence of tears and different types of tears. Similar to that analysis, this study also aims to utilize metrics in order to determine how well the CNN models can correctly classify MRIs of knees. While AUC is a metric ideal for binary situations, this study aims to use F1 scores for multi-class classification to determine the models' overall performances.

According to "Brain Tumor MRI Image Segmentation Using Deep Learning Techniques," DL models are usually trained to handle necessary tasks using manually developed features derived from raw features or data obtained from other basic machine learning (ML) techniques. CNNs provide an efficient way to learn beneficial features of images, and until the CNNs could be used effectively, representations or features had to be developed by hand or generated by less effective feature learning techniques. Due to CNNs, many of the hand-crafted features were abandoned since the models were able to provide better image features (Chaki 2021, 4-5). A study mentioned in the book used the CNN architecture for brain tumor segmentation and detection in MRIs. According to the study, the model's prediction detection accuracy increased when it came to categorizing benign tumors and malignant tumors. Like the experiment on brain tumor MRIs, this study also focuses on implementing CNN approaches in order to find distinct patterns in knee MRIs (Chaki 2021, 37-57). Based on the literature, CNN models are able to successfully find unique features that may otherwise be missed.

Deep learning methods have provided better performance compared to conventional techniques and medical experts. According to a chapter from "Advances in Experimental Medicine and Biology," the potential for applying deep-learning-based medical image analysis to computer-aided diagnosis (CAD) can provide support to clinicians and improve accuracy, as well as efficiency, when it comes to diagnostics and treatment processes. Deep CNN techniques like AlexNet demonstrated pattern recognition capability of multiple layers of a deep structure, and such techniques provided immense success in deep learning studies. However, there are challenges in the development and implementation of CAD or AI tools in clinical practices. Large amounts of data is necessary to test and validate deep learning models. Additionally, the preferred mode of assistance needs to be provided to clinicians or the implementation will not provide sufficient value (Chan et al. 2020). This study that looks at different CNN models, including AlexNet, further acknowledges what is mentioned in the chapter. The amount and type of data available can make it difficult to properly conduct experiments. While this study uses a dataset provided by the Stanford ML Group, collecting and cleaning the data would otherwise be costly.

A study tested the hypothesis that AI techniques can aid in identifying and assessing lesion severity in the cartilage, bone marrow, meniscus, and ACL in the knee, improving overall MRI intergrader agreement. The study was conducted on 1,435 knee MRIs, and three-dimensional (3D) CNNs were developed to detect the regions of interest within MRI studies and grade abnormalities of the cartilage, bone marrow, meniscus, and ACL. The CNNs were evaluated based on sensitivity, specificity, and Cohen linear-weighted K, and the impact of the AI-aided grading in intergrader agreement was assessed on a separate dataset. The study also looked at the AUC for the tissues of interest. Based on the analysis, the 3D CNNs had high sensitivity, specificity, and accuracy for knee-lesion-severity scoring and increased intergrader agreement when used on an external dataset. The research provides insight into how various angles of knee MRIs should be taken into consideration when it comes to image assessment (Astuto et al. 2021). Similarly, the dataset used in this study provides three different planes, and they are all analyzed in order to accurately assess and classify the knee injuries.

Data

The data used in this research was from the Stanford ML Group that provided a dataset consisting of 1,370 knee MRI exams performed at Stanford University Medical Center. The dataset contains 1,104 (80.6 percent) abnormal exams, including 319 (23.3 percent) ACL tears and 508 (37.1 percent) meniscal tears. According to the Stanford ML Group, the labels were obtained through manual extraction from clinical reports. The MRI examinations are split into a training set (1,130 exams, 1,088 patients) and a validation set (120 exams, 111 patients), and a hidden test set (120 exams, 113 patients). Stratified sampling was used to make sure that at least 50 positive examples of each label (abnormal, ACL tear, and meniscal tear) were present in each set, and exams from each patient were put in the same split (Stanford ML Group n.d.).

Exploratory data analysis (EDA) showed that CSV files contained the case number of the MRI and whether the case fit the category. For example, case 0001 in the training set of the ACL file is abnormal and is labeled with a one [see Table 3 in Appendices]. The dataset also provides images from three different sides, including sagittal, axial, and coronal (as .npy files). Access to such images provides the models with the ability to predict and classify injuries while taking into consideration the same type of information medical professionals would view during their assessments. Additionally, the MRI scans can be used to view the injuries from different angles to determine features that may be unique specific injuries. Creating a widget further helped to view the MRI slices within a given place, and manipulating the splices can better highlight the different injuries [see Figure 2 in Appendices]. When looking at the positive and negative cases in the training set, there were 217 positive cases and 913 cases for abnormalities, 922 positive cases and 208 negative cases for ACL tears, and 733 positive cases and 397 negative cases for meniscus tears [see Table 6 in Appendices]. Understanding how many positive and negative cases can help to determine the accuracy of the models when viewing how many positive and negative cases were determined by those models.

Before preprocessing and merging the validation set, the training set was analyzed to determine how many cases fall into more than one category. For example, there were cases where an MRI indicated an injury to be all cases, including abnormal, an ACL tear, and a meniscus tear. In the training set, the highest frequency of injuries was abnormal only injuries (38.3 percent) while the lowest frequency was of ACL and abnormal injuries combined (7.3

percent) [see Table 4 in Appendices]. In the validation set, the distribution was more even with the lowest number of cases was for abnormal injuries only (16.2 percent) and highest frequency was of injuries with meniscus tears, ACL tears, and abnormality (25.8 percent) [see Table 5 in Appendices]. The EDA showed that there were 433 abnormal cases, 83 ACL and abnormal cases, 125 meniscus, ACL, and abnormal cases, 272 meniscus and abnormal cases, and 217 cases with nothing detected [see Table 4 in Appendices]. The validation set had 20 abnormal cases, 23 ACL and abnormal cases, 31 meniscus, ACL, and abnormal cases, 21 meniscus and abnormal cases, and 25 cases with nothing detected [see Table 5 in Appendices]. Such numbers of cases indicate that there were no cases with just ACL or meniscus tears in the training or validation sets. Insight from the EDA not only helped to gain a better understanding of the dataset, but it also helped to determine how the models need be configured to properly account for all cases. This means the planes of each case, as well as the type of injury, need to be properly parsed through during the training process.

Methods

The methods applied in this study are based on literature review that indicates CNN models are ideal for image analysis. One of the models used is the AlexNet architecture, which allows for multi-GPU training by putting half the model's neurons on one GPU and the other half on another GPU (Wei 2019). That provides the ability to train a bigger model and cut down on the training time. AlexNet has eight layers with learnable parameters. The model has five layers with a combination of max pooling, which helps to reduce spatial dimensionality of the input data, and three fully connected layers that help to extract high-level features from the input data.

ReLU activation is used in each layer except the output layer in order to handle vanishing gradients (Saxena 2021).

Another model that is used in this study is ResNet, a neural network that helps to address the problem of the vanishing or exploding gradient. Skip connections is a technique used in a ResNet model, which helps to connect activations of a layer to further layers by skipping some layers in between. When applying skip connection, a layer hurting the performance of the architecture is skipped by regularization, so a deep neural network is trained without problems caused by vanishing or exploding gradient (Pawangfg 2023).

Both the ResNet50 and AlexNet used in this study were based on the architecture for their respective models. However, additional layers were added to the models that were the same for both models. A dropout layer was removed from the AlexNet model since the original model structure already has two dropout layers after each fully connected layer [see Table 7 in Appendices]. Adding similar additional layers can help to not only enhance feature learning and extraction but also determine how each model behaves. Such layers also provide a future approach of combining models and modifying architectures to improve generalization capabilities and noise toleration, as well as avoid biases. As seen in the model summary [see Table 6 in Appendices], most factors for both models were kept consistent in order to not only understand the impact of the dataset itself but also determine why a model is providing certain results. The input image (image resolution), epochs (number of iterations through training data), loss function (cost function), batch size (number of training examples processed together in each iteration), dropout layers (technique to prevent overfitting in neural networks), and early stopping monitor (technique to prevent overfitting and find optimal number of epochs) were kept consistent for both models [see Table 6 in Appendices]. However, ResNet50 is a pre-trained model in Tensorflow that is trained on ImageNet with 1.2 million images belonging to 1,000 classes (LearnOpenCV n.d.).

The two models are compared using metrics like training loss and accuracy, as well as the F1 scores. The training loss can help to determine how well the model fits the training set while the accuracy can help to determine how well the model fits the testing set. The F1 score is ideal for multi-class classification tasks since it can be used to determine the performance of the models by combining precision and recall into a single value. Metrics like accuracy (overall performance based on correct predictions) and F1 scores, which includes precision (number of true positives among all positive predictions), recall (number of true positives among all actual positives) can ultimately help to understand the performances of the two models in this comparative study.

Results

Table 1 – ResNet50 model

Plane (injury type)	Train loss	Train accuracy	F1 score
Axial (Meniscus)	2.29	0.59	0.30
Axial (ACL)	0.77	0.70	0.54
Axial (Abnormal)	4.27	0.40	0.36
Coronal (Meniscus)	1.23	0.63	0.45
Coronal (ACL)	1.02	0.73	0.54
Coronal (Abnormal)	5.26	0.27	NaN
Sagittal (Meniscus)	1.42	0.63	0.35
Sagittal (ACL)	2.01	0.74	0.39
Sagittal (Abnormal)	1.10	0.68	0.71

Table 2 – AlexNet model

Plane (injury type)	Train loss	Train accuracy	F1 score
Axial (Meniscus)	0.67	0.57	NaN
Axial (ACL)	0.71	0.55	NaN
Axial (Abnormal)	0.93	0.21	NaN
Coronal (Meniscus)	0.70	0.57	NaN
Coronal (ACL)	0.87	0.55	NaN
Coronal (Abnormal)	0.97	0.21	NaN

Sagittal (Meniscus)	0.70	0.57	NaN
Sagittal (ACL)	0.91	0.55	NaN
Sagittal (Abnormal)	0.92	0.21	NaN

Figure 1 – ResNet50 (axial – meniscus)

Figure 2 – ResNet50 (axial – ACL)

Figure 3 – ResNet50 (axial – abnormal)

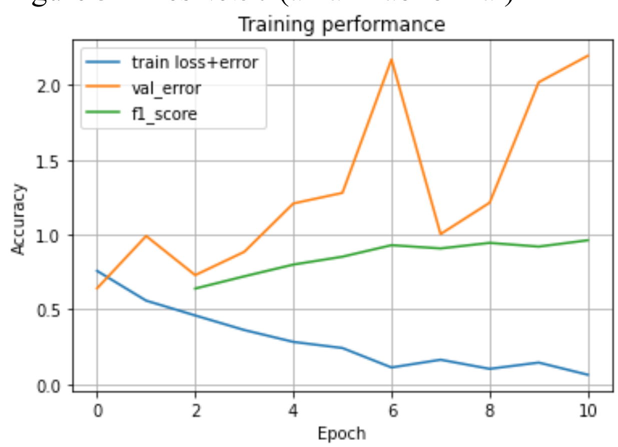


Figure 4 – ResNet50 (coronal – meniscus)

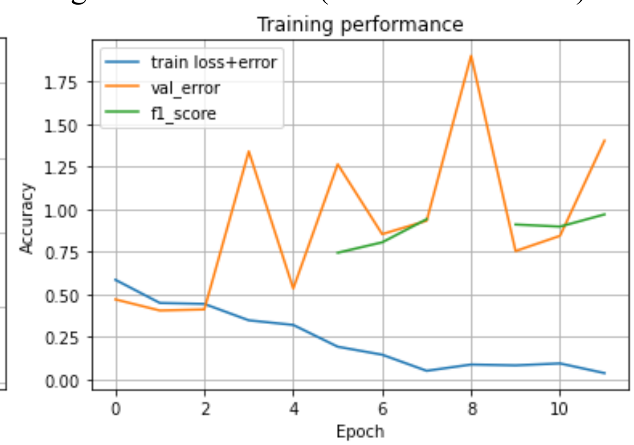


Figure 5 – ResNet50 (coronal – ACL)

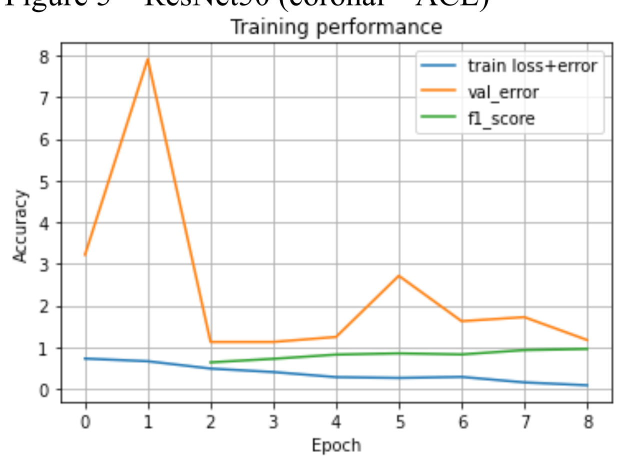


Figure 6 – ResNet50 (coronal – abnormal)

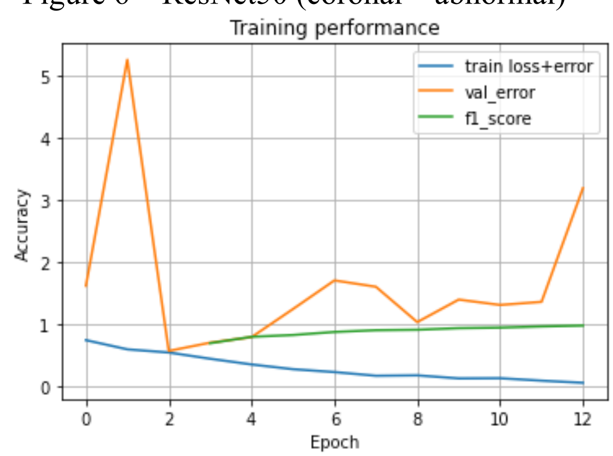


Figure 7 – ResNet50 (sagittal – meniscus)

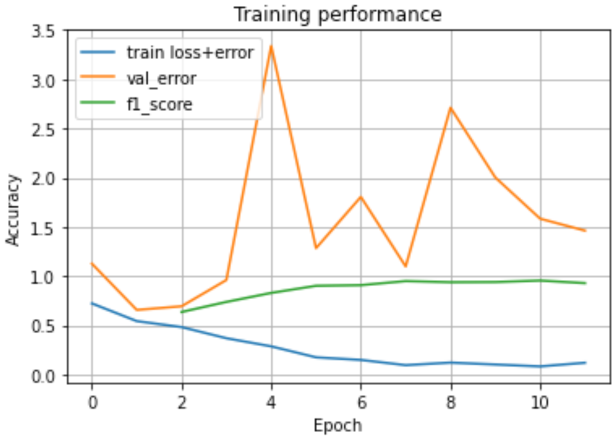


Figure 8 – ResNet50 (sagittal – ACL)

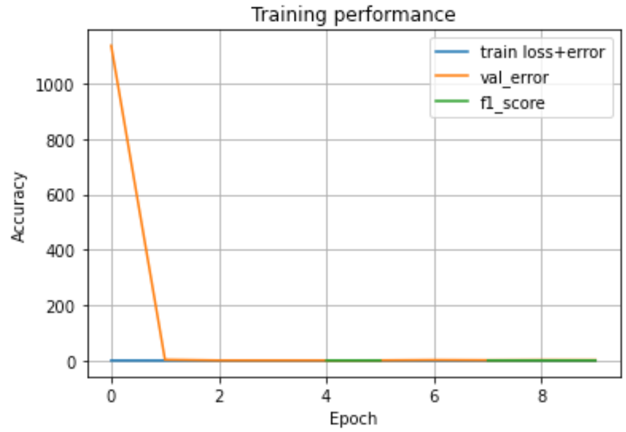


Figure 9 – ResNet50 (sagittal – abnormal)

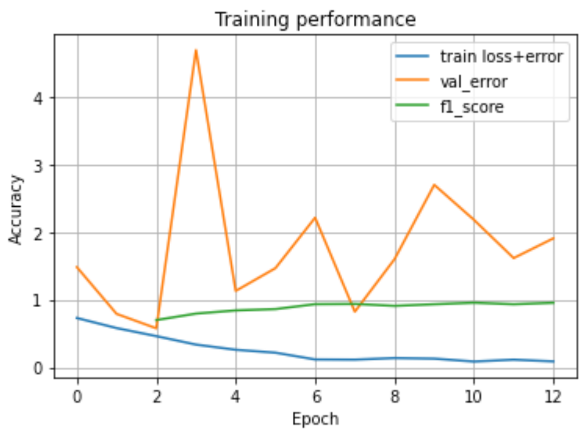


Figure 10 – AlexNet (axial – meniscus)

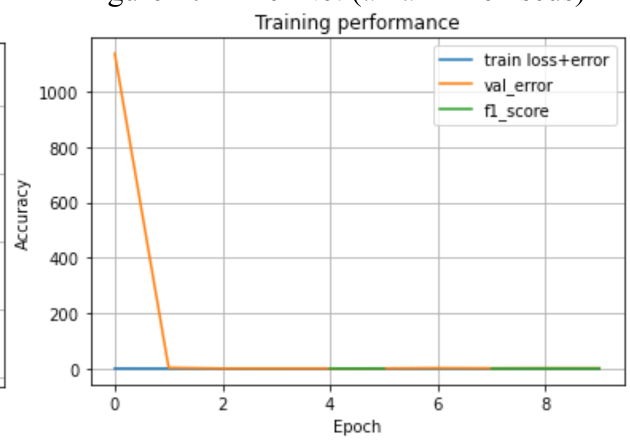


Figure 11 – AlexNet (axial – ACL)

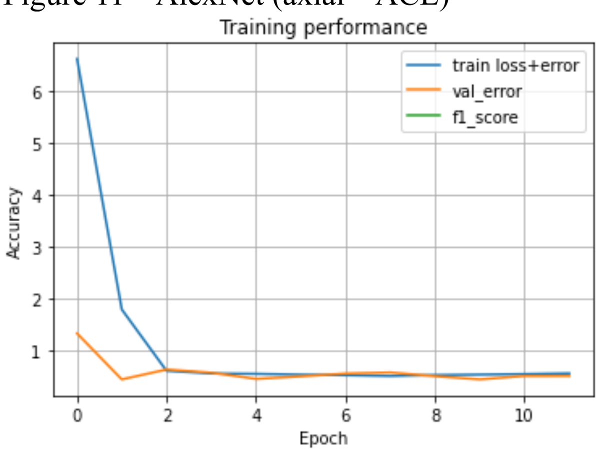


Figure 12 – AlexNet (axial – abnormal)

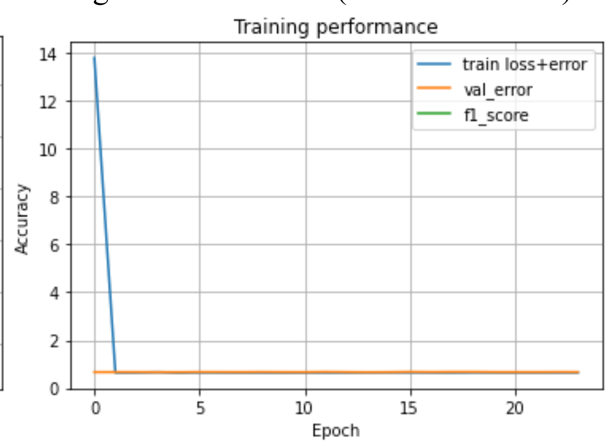


Figure 13 – AlexNet (coronal – meniscus)

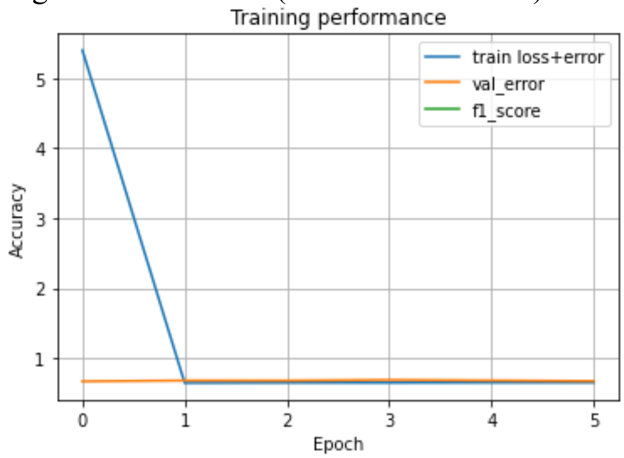


Figure 14 – AlexNet (coronal – ACL)

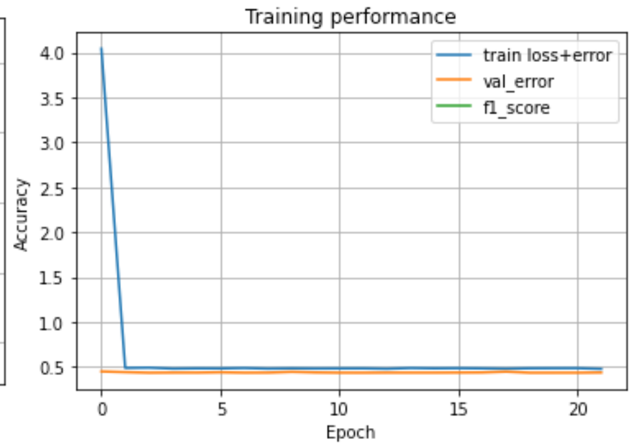


Figure 15 – AlexNet (coronal – abnormal)

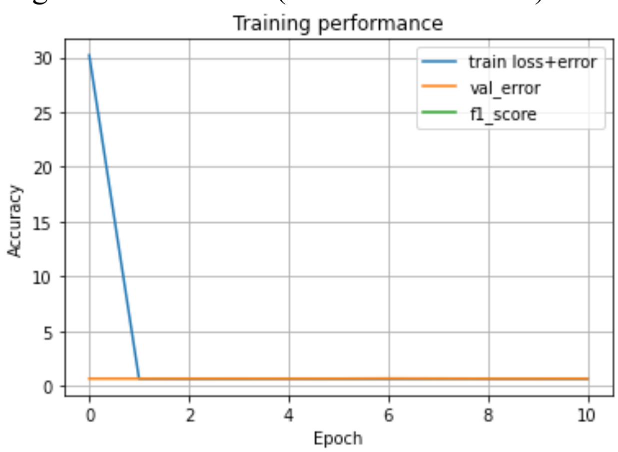


Figure 16 – AlexNet (sagittal – meniscus)

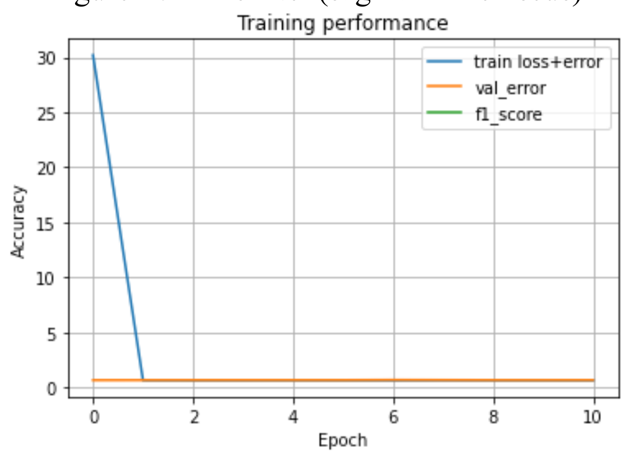


Figure 17 – AlexNet (sagittal – ACL)

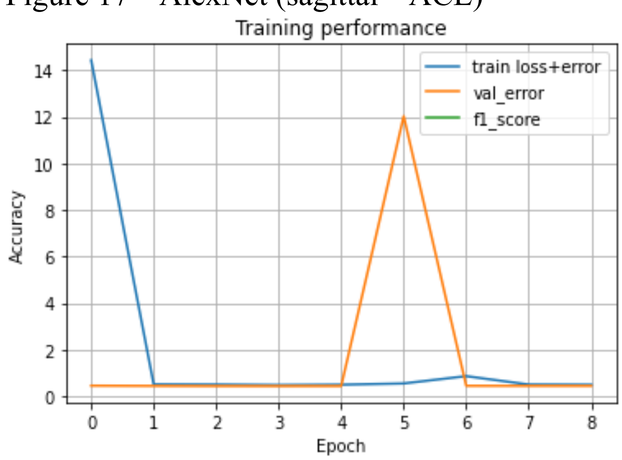
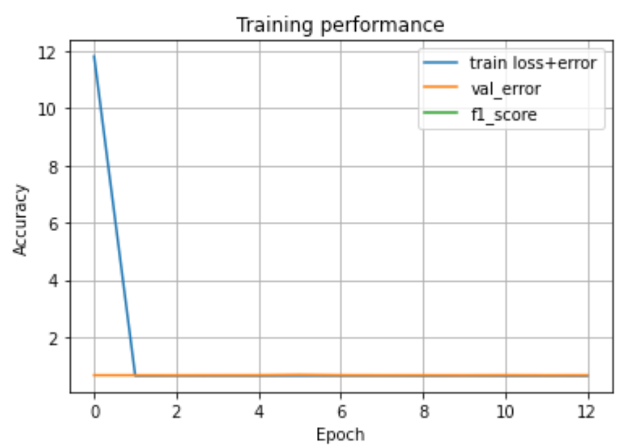


Figure 18 – AlexNet (sagittal – abnormal)



Analysis and Interpretation

Results show that the pre-trained ResNet50 model had a train loss that was between 0.75 and 4.30 [see Table 1 in Results]. Generally, lower train loss values indicate better model performance during training since those values suggest that the model fits the training data closely. However, there are various factors to consider, including the dataset, which in this case includes 1,370 cases, the model architecture, the complexity of the problem, and the training process. When compared to the custom AlexNet model created, the train loss was higher on average for the ResNet50 model. The custom AlexNet model had a loss between 0.65 and 0.95, which would indicate the model should provide better accuracy compared to the ResNet50 model. However, the accuracy was much lower for AlexNet with values between 0.20 and 0.60 while the ResNet50 model had accuracy values between 0.25 and 0.75 [see Tables 1 and 2 in Results]. A major reason behind the higher accuracy scores for all planes (axial, coronal, and sagittal) across all injury types (meniscus, ACL, and abnormal) was likely because the Tensorflow model was pre-trained on 1.2 million labeled images (LearnOpenCV n.d.).

Considering the ResNet50 results, the AlexNet model with additional data and extensive finetuning can help to improve results. The values for training accuracy also suggest that the relationship between loss and accuracy in CNN models is not necessarily straightforward, and there may be situations where models with higher loss can provide higher accuracy compared to models with lower train loss. This could be due to overfitting for the AlexNet model as it became too specific to the training data but could not properly generalize to the testing data, which led to the lower accuracy despite the low loss values when compared to the ResNet50 model's loss values.

When looking at the F1 scores for the ResNet50 model, the scores were typically between 0.30 and 0.75 with a score indicating “NaN” while the F1 scores for all AlexNet runs were “NaN” across all planes and injuries [see Tables 1 and 2 in Results]. This is also reflective across every epoch, which shows that F1 scores are not available as the values remain low throughout the run [see Figures 10-18 in Results]. While a higher F1 score indicates a better-quality classifier, there are factors to consider when looking at the models in this study. Within the context of the study that does not have extremely high F1 scores, a score above 0.50 can be considered reasonably beneficial since this likely means that the model is reaching a balance between precision and recall. Additionally, the F1 score is dependent on the dataset, which in this case may not contain enough cases for each variation to provide solid precision and minimize false positives, as well as solid recall that should identify positive cases. Even though stratified sampling was used to ensure the training and testing sets had the correct proportion of cases, the total number of exams had 80.6 percent of cases that are considered abnormal exams while 23.3 percent are ACL tears and 37.1 percent are meniscal tears (Stanford ML Group n.d.). Therefore, imbalanced classes could have led to “average” F1 scores for the ResNet50 model and “NaN” values for the AlexNet model.

While AlexNet models provide quality results for other studies, the AlexNet architecture used to build the simple model in this study may not be suitable for MRI analysis, which requires capturing patterns and complexities of data. Removing a dropout layer meant to prevent overfitting helped to improve results slightly, but the accuracy across all planes for each type of injury was still not sufficient compared to the ResNet50 model. Additionally, the angle of the MRI (axial, coronal, and sagittal) might play a role in the individual accuracy scores, but there was no clear indication that one plane provides better results than another.

Conclusion

The experiments ran in this study showed that training a model on an efficient amount of data can provide the best results. In this case, the pre-trained ResNet50 model provided better accuracy scores compared to the custom AlexNet model. While additional data can help the models to better identify various MRI cases for knees, the models themselves also require modification in order to meet realistic expectations. The dataset in this study included far more abnormal cases compared to ACL tears and meniscus tears. While that may have been a factor in the low accuracy scores and even “NaN” F1 scores, the dataset is reflective of what medical professionals likely handle on a common basis. With that in mind, further research is necessary to validate the models’ capabilities and determine how valuable they are in a clinical setting to support professionals in MRI assessments where certain cases are more common than others.

Directions for Future Work

This study can be further enhanced by taking various directions of the existing experiments. Firstly, the ResNet50 and AlexNet models can be analyzed after adding additional data, which can help to improve accuracy, as well as reduce overfitting and unexpected variability that occurs with unbalanced datasets. Another direction for future work includes conducting in-depth research into how AlexNet models are applied to image analysis, and more specifically, MRIs, in order to understand why the simple AlexNet model built in this study failed to provide results similar to the pre-trained ResNet50 model. Part of this process also includes testing the existing models on other datasets to determine how the complexity of the problem in this study requires modification and finetuning of the models.

Thirdly, analyzing and comparing another model can help to determine where the current two models stand. The visual geometry group (VGG) model is commonly used in image analysis. VGG is a standard CNN architecture with multiple layers, and more specifically, this study uses VGG-16, which consists of 16 convolutional layers (13 convolutional layers and three fully connected layers). The hidden layers in the VGG network use ReLU to address the issues of the vanishing gradient. The first two fully connected layers have 4,096 channels each, and the third layer has 1,000 channels, one for each class (Boesch n.d.).

Finally, this study focused on analyzing each plane and injury individually. Realistically, medical professionals analyze MRI scans and consider all angles to diagnose an injury. Future models should combine all three MRI scan planes to analyze injuries holistically.

Code and Data Availability

I am truly grateful to the Stanford ML Group for making the knee MRI data publicly available.

Acknowledgements

I would like to express my gratitude to Dr. Alianna Marena and Robert Guenther for providing guidance used to properly conduct this study and construct the paper.

Appendices

Table 3 – Training set for ACL tears (CSV file)

	Case	Abnormal
0	0000	0
1	0001	1
2	0002	0
3	0003	0
4	0004	0

Table 4 – Training set frequency count for types of injuries

	diagnostic_label	case	abnormal	acl	meniscus	%_freq
0	ACL + Abnormal	83	83	83	83	7.3
1	Abnormal Only	433	433	433	433	38.3
2	Meniscus + ACL + Abnormal	125	125	125	125	11.1
3	Meniscus + Abnormal	272	272	272	272	24.1
4	Nothing Detected	217	217	217	217	19.2

Table 5 – Validation set frequency count for types of injuries

	diagnostic_label	case	abnormal	acl	meniscus	%_freq
0	ACL + Abnormal	23	23	23	23	19.2
1	Abnormal Only	20	20	20	20	16.7
2	Meniscus + ACL + Abnormal	31	31	31	31	25.8
3	Meniscus + Abnormal	21	21	21	21	17.5
4	Nothing Detected	25	25	25	25	20.8

Table 6 – Positive and negative cases of knee injuries

Type of injury	**Positive (1)**	**Negative (0)**
Abnormal	913	217
Meniscus	397	733
ACL	922	208

Figure 1 – Training set MRI scans (sagittal, axial, and coronal)

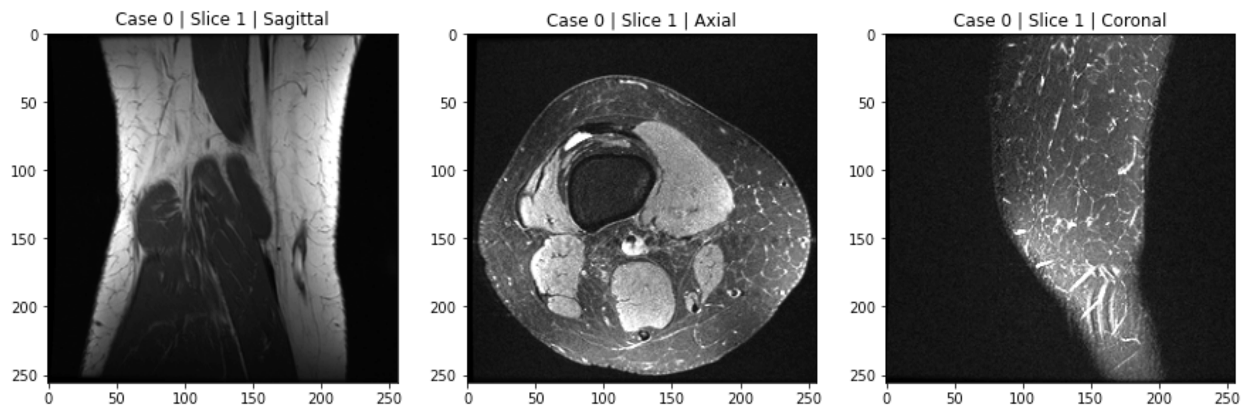
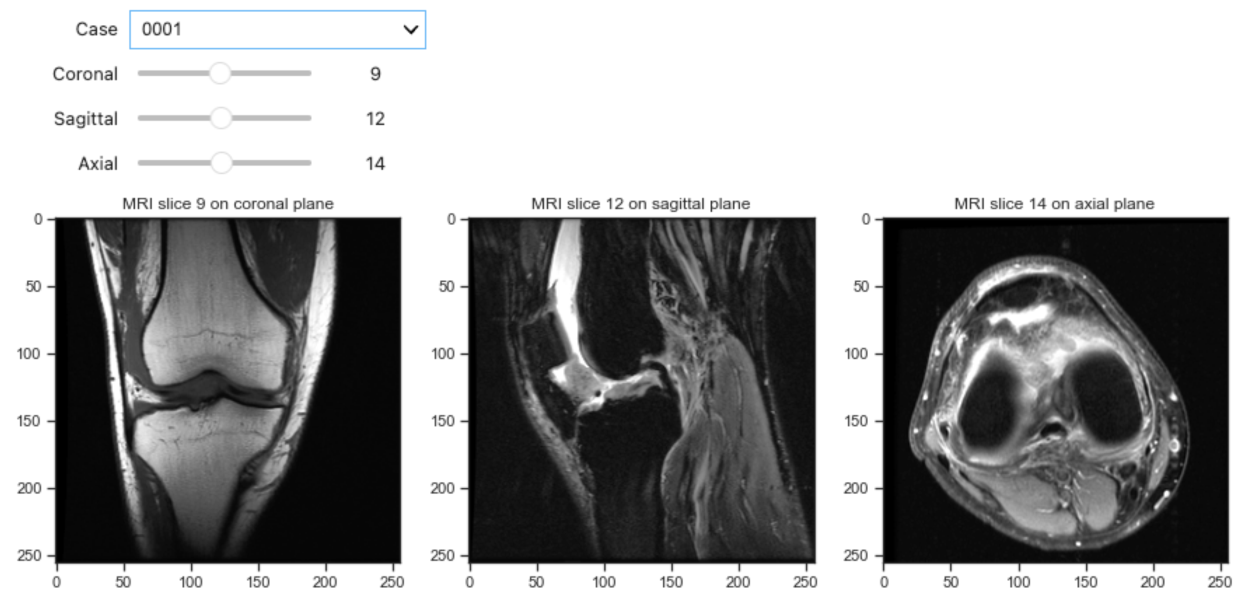


Figure 2 – Interactive widget for given cases



The interactive widget provides the ability to view the different planes (coronal, sagittal, and axial) from different angles, which can help to highlight the injuries.

Table 6 – Model summary

Model Type	**AlexNet**	**ResNet50**
Model concepts	Custom model	Pre-trained
Input image	256x3 (RBG)	256x3 (RBG)
Epochs	100	100
Weights	-	ImageNet
Loss function	Binary crossentropy	Binary crossentropy
Optimizer	Adam	SGD

Learning rate	0.01	0.001
Batch size	16	16
Dropout	Yes	Yes
Early stopping monitor	Yes	Yes

The model summary provides an overview of the models, including the training approach, optimizer, loss function, learning rate, batch size, use of dropout layers, and early stop monitoring.

Table 7 – Model architecture

AlexNet	**ResNet50**
Input layer	Input layer
Convolutional layers (5)	Convolutional layers/batch normalization layers/skip connections
MaxPooling layers (3) – In 3 out of 5 convolutional layers	Global average pooling layer
Flatten layer	Fully connected layer
Fully connected layers (2)/Dropout layer	Softmax activation layer
Output layer	-

Additional layers
Global average pooling layer
Dropout layer (AlexNet – 2) (ResNet50 – 3)
Dense layer (2)

The model architecture provides an overview of the layers used in the models.

References

Astuto, Bruno, Io Flament, Nikan K. Namiri, Rutwik Shah, Upasana Bharadwaj, Thomas M. Link, Matthew D. Bucknor, Valentina Pedoia, and Sharmila Majuumdar. 2021. "Automatic Deep Learning–assisted Detection and Grading of Abnormalities in Knee MRI Studies." *Radiology: Artificial Intelligence* 3, no. 3 (January). https://doi.org/10.1148/ryai.2021200165.

Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, Ashrafian H, Darzi A. 2021. "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis." *NPJ Digit Medicine*, 4, no. 65. https://doi.org/10.1038/s41746-021-00438-z.

Awati, Ragul. 2023. "Convolutional neural network (CNN)." *TechTarget*, April 2023. https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network.

Bien, N., Rajpurkar, P., Ball, R. L., Irvin, J., Park, A., Jones, E., Bereket, M., Patel, B. N., Yeom, K. W., Shpanskaya, K., Halabi, S., Zucker, E., Fanton, G., Amanatullah, D. F., Beaulieu, C. F., Riley, G. M., Stewart, R. J., Blankenberg, F. G., Larson, D. B., Jones, R. H., Langlotz, C. P., Ng, A. Y., and Lungren, M. P. 2018. "Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet." *PLOS Medicine* 15, no. 11 (November). https://doi.org/10.1371/journal.pmed.1002699.

Boesch, Gaudenz. n.d. "VGG Very Deep Convolutional Networks (VGGNet) – What you need to know." *Viso.ai*. https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/.

Chaki, Jyotismita. 2021. *Brain Tumor MRI Image Segmentation Using Deep Learning Techniques*. New Delhi: Elsevier.

Chan, Heang-Ping, Ravi K. Samala, Lubomir M. Hadjiiski, Chuan Zhou. "Deep Learning in Medical Image Analysis." In *Advances in Experimental Medicine and Biology*. Berlin: Springer Nature.

LearnOpenCV. n.d. "Image Classification using Pre-Trained ImageNet Models in TensorFlow & Keras." Accessed April 21, 2023. https://learnopencv.com/image-classification-pretrained-imagenet-models-tensorflow-keras/.

Stanford ML Group. "MRnet Dataset: A Knee MRI Dataset And Competition." Accessed April 5, 2023. https://stanfordmlgroup.github.io/competitions/mrnet/.

Pawangfg. 2023. "Residual Networks (ResNet) – Deep Learning." *Geeks for Geeks*. https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/

PTVN-S. 2022. "Knee-Abnormality-and-Common-Disorders." *GitHub*, March 3, 2022. https://github.com/ptnv-s/Knee-Abnormality-and-Common-Disorders.

Saxena, Shipra. 2021. “Introduction to The Architecture of Alexnet.” *Analytics Vidhya*, March 19, 2021. https://www.analyticsvidhya.com/blog/2021/03/introduction-to-the-architecture-of-alexnet/#:~:text=The%20Alexnet%20has%20eight%20layers,layers%20except%20the%20output%20layer.

Shin, H., Choi, G.S., Shon, OJ. *et al.* “Development of convolutional neural network model for diagnosing meniscus tear using magnetic resonance image.” *BMC Musculoskelet Disorders* 23, 510 (2022). https://doi.org/10.1186/s12891-022-05468-6.

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 (Basel)* 12, no. 2. (February):537. https://doi.org/10.3390/diagnostics12020537.

Wei, Jerry. 2019. “AlexNet: The Architecture that Challenged CNNs.” *Towards Data Science*, July 2, 2019. https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951#:~:text=AlexNet%20allows%20for%20multi%2DGPU,Overlapping%20Pooling.