Machine Learning Models for Interpretation of Knee MRI

Alexandru Licuriceanu
Computer Science Department
Politehnica University of Bucharest
Bucharest, Romania
alicuriceanu@stud.acs.upb.ro

Nelu-Rareș-Ionuț Cărăușu
Computer Science Department
Politehnica University of Bucharest
Bucharest, Romania
nelu_rares.carausu@stud.acs.upb.ro

I. Introduction

Knee injury classification from MRIs is a challenging task due to the complex nature of soft tissue structures and subtle variations between pathological and normal cases. Traditional approaches rely heavily on handcrafted feature extraction and shallow learning models, which often fail to generalize across diverse datasets.

To address these challenges, we utilized transfer learning using a pretrained EfficientNet-B0 model [1]. EfficientNet uses compound scaling to optimize the trade-off between network depth, width, and resolution, making it computationally efficient and well-suited for medical imaging tasks. The EfficientNet-B0 variant, being the most lightweight version and less computationally intensive than other networks, was selected to ensure good results at a reasonable computational cost.

II. Data Preprocessing

Since the MRI images are not standardized, the number of slices can vary significantly depending on the imaging protocol and the patient. By analyzing the dataset, we found out that the number of images per plane would range anywhere from a minimum of 17 to a maximum of 61. Therefore, we selected a fixed subset of slices for analysis from each patient, specifically, we extracted the 8 most central slices from each plane (axial, coronal, and sagittal). The justification for this is that by choosing the central slices, which typically contain the most critical information for diagnosis, we focus more on relevant images and discard the less relevant images that wouldn't provide any informational gain, and secondly, standardize the number of slices for each patient. Experiments were also conducted with 17 slices per plane instead of 8, with the results being similar, the training time took almost 4 times longer.

After performing exploratory data analysis (EDA) on the dataset, we observed that the class distribution was highly imbalanced. To counter wrong predictions, enhance the diversity of the training data and improve the generalization of the model, we applied various data augmentation techniques to the MRI slices. These augmentations included random rotations, horizontal and vertical flips, translations, and intensity scaling. By introducing variability in the orientation, position, and brightness of the slices, we simulated real-world conditions such as different patient positions and variations in image acquisition settings of the MRI machine and also helped to prevent overfitting. The augmentation steps used on the training dataset are the following:

TABLE I
Transformation sequence for the training dataset

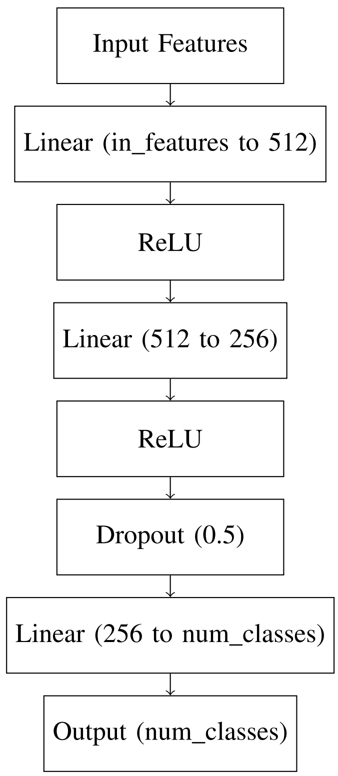
Transformation	Parameters
Resize	Width 224, Height 224
RandomRotation	15 Degrees
HorizontalFlip	50% Probability
VerticalFlip	50% Probability
ColorJitter	Brightness 0.2, Contrast 0.2
RandomPerspective	Distortion Scale 0.1, 50% Probability

On the validation and testing datasets, the only transformation we applied was resizing the images from a resolution of 256x256 to the input size of EfficientNet-B0, 224x224.

III. Own Contribution

Our approach introduces a novel methodology for analyzing knee MRI scans using machine learning by addressing a significant gap in existing research. While most prior studies have focused on diagnosing specific injuries, such as abnormalities or isolated conditions like anterior cruciate ligament (ACL) tears, they often fail to provide a comprehensive analysis of multiple injury types. This narrow scope limits their applicability in clinical settings where there is a need for a complete understanding of various knee injuries. Therefore, our work aims to encompass a broader range of injuries, integrating the detection of ACL tears, meniscus tears, and general abnormalities into a unified model.

Since the size of the MRI dataset is relatively small, training a model from scratch would likely lead to overfitting and suboptimal performance. The approach we selected was to use transfer learning with the pre-trained EfficientNet-B0 model on the large and diverse ImageNet dataset. This technique should improve the performance of the model on the MRI dataset, by using the model's ability to capture general characteristics, such as edges, shapes, and textures. First of all, we removed the fully-connected layer of EfficientNet-B0 and replaced it with the following architecture (num_classes is 3: ACL, meniscus, abnormalities):



Secondly, we froze all the layers of the network, except the fully-connected layers and the last 3 (out of 9 total) convolutional blocks of the EfficientNet model, and started the training on 10 epochs, measuring the loss and Receiver Operating Characteristic Area Under the Curve (ROC AUC) for both train and validation datasets:

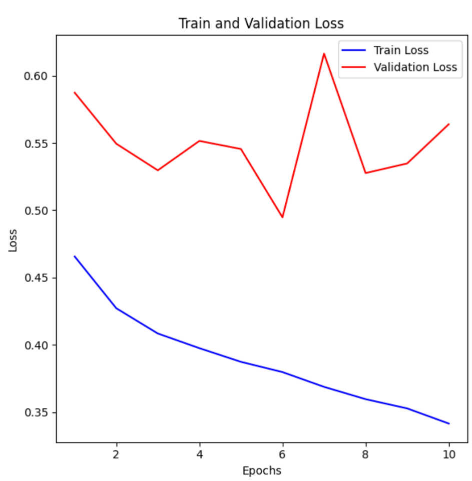


Fig. 1. Train and validation loss vs epoch.

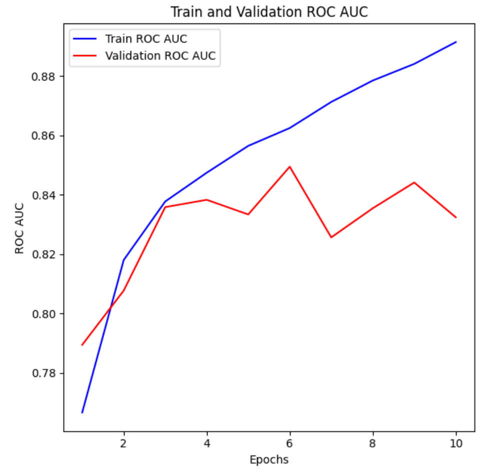


Fig. 2. ROC AUC vs epoch.

The best model achieved a ROC AUC score of 0.84 on the validation dataset. Finally, we also tested this best model on the testing dataset, measuring a ROC AUC score of 0.81 on it. We also displayed the classification report for the individual labels:

TABLE II
CLASSIFICATION REPORT (INDIVIDUAL LABELS)

Label	**Precision**	**Recall**	**F1-Score**	**Support**
ACL	0.80	0.32	0.45	912
Meniscus	0.61	0.46	0.52	1032
Abnormal	0.85	0.89	0.87	2352
Micro Average			0.79	4296
Macro Average	0.76	0.55	0.61	4296
Weighted Average	0.78	0.66	0.70	4296
Samples Average	0.62	0.52	0.54	4296

A key takeaway from this table is that the class imbalance is clearly visible, the model predicts the "abnormal" label better than the others because of the nature of the dataset.

IV. Final Comparison

Our approach demonstrates an improvement over the baseline model, with an ROC AUC of 0.84 on the validation set and 0.81 on the training set, compared to the baseline's ROC AUC of 0.77 on validation and 0.78 on training, highlighting the performance of our approach. When compared to state-of-the-art methods, which typically focus on diagnosing specific knee injuries, instead of a full comprehensive view, our method stands out by addressing a broader range of injuries, maintaining good performance, while also surpassing other models' ROC AUC scores, but not achieving exceptional performance such as other implementations that focus only on a single type of injury (for example: 0.98 ROC AUC on validation set for abnormalities [2], 0.78-0.96 ROC AUC on validation set for each type of injury [3] or 0.97 ROC AUC on validation set for ACL tears [4])

REFERENCES

[1] Tan, M. and Le, Q.V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.

[2] Yin Dai, Yifan Gao and Fayu Liu, TransMed: Transformers Advance Multi-Modal Medical Image Classification, " Machine Learning for Computer-Aided Diagnosis in Biomedical Imaging", July 31, 2021.

[3] Benjamin Fritz, Giuseppe Marbach, Francesco Civardi, Sandro F. Fucentese and Christian W.A. Pfirrmann, Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference, "Skeletal Radiology", Volume 49, pages 1207–1217, 2020.

[4] Peter D. Chang, Tony T. Wong and Michael J. Rasiej, Deep Learning for Detection of Complete Anterior Cruciate Ligament Tear, "Journal of Digital Imaging", Volume 32, pages 980–986, 2019.