3D Neural Network for Detection of ACL Injury in Knee MRI Scans

Author

Abdullah Kamran

Regn Number

00000319860

Supervisor

Dr. Syed Omer Gilani

DEPARTMENT OF BIOMEDICAL ENGINEERING AND SCIENCES

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD, PAKISTAN

September, 2022

3D Neural Network for Detection of ACL Injury in Knee MRI Scans

Author

Abdullah Kamran

Regn Number

00000319860

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Biomedical Engineering

Thesis Supervisor:

Dr. Syed Omer Gilani

Thesis Supervisor's Signature:

_____

DEPARTMENT OF BIOMEDICAL ENGINEERING AND SCIENCES

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD, PAKISTAN

September, 2022

Declaration

I certify that this research work titled **"*3D Neural Network for Detection of ACL Injury in Knee MRI Scans*"** is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student _____

Abdullah Kamran

Regn No. 319860

MS Biomedical Engineering

Copyright Statement

Dedication

Dedicated to my Parents for their prayers and support during my academic career

Acknowledgments

I would like to express my sincere gratitude to my supervisor **Dr. Syed Omer Gilani** for his continuous support and motivation during my MS research work and guiding me throughout the whole research project.

I would also like to thank all my GCE members, Dr. Asim Waris, Dr. Adeeb Shahzad and Dr. Amer Sohail Kashif for their guidance and valuable comments.

I am deeply thankful to all my friends and fellows for making my days at NUST ones I would cherish forever. Finally, I am extremely grateful to my parents for their unconditional love and encouragement throughout my MS degree.

Abstract

Computer aided diagnosis is widely used in medical imaging for the diagnosis of many diseases such as cardiomegaly, brain and kidney tumor, lung cancer, COVID-19 and may more. For the past few decades, computer aided diagnosis has significantly improved due to the development of better architecture used for the diagnosis. Knee injury diagnosis using deep learning techniques is highly popular due its high detection rate and is highly localized. Many state-of-the-art-deep learning models have been used for the detection of abnormalities, meniscus tear and ACL tears in Knee MRI scans. These models include RESNET, Google-Net, VGG19 and VGG16, Alex-Net and many other, all giving significant results. In this study we used a custom 3D CNN model which is light in weight. For training we are using two datasets, one provided by Stanford ML group and the other form Hospital in Croatia. We combined the two dataset and split it into 80-20 ration (80% of the data used for training and remaining for testing purposes). Both the dataset has extreme class imbalance, so we used data augmentation and class weights to rectify its effect on the training process. Further the voxel intensities for the two datasets were different (one dataset was in 8-bit format and the second was in 12-bit format), so we normalized the intensity values using mathematical formulas. For contrast, we performed adaptive histogram equalization Average accuracy and AUC achieved by our model on training set is 97.6 and 99.3 respectively, during 5-fold cross validation.

INDEX TERMS: Deep Learning, MRNet, Croatia Knee MRI, classification, class weights, 3D-CN

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CNN	Convolutional Neural Network
ACL	Anterior Cruciate Ligament
RESNET	Residual Network
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
AUC	Area Under Curve
ML	Machine Learning
ROC	Receiver Operator Characteristic
ROI	Region Of Interest
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
2D	2-Dimensional
3D	3-Dimensional

Chapter 1:Introduction

Among different Knee joint injuries, Anterior cruciate ligament tear is the most common especially when it comes to professional football and basketball players where due to running, Knee joint constantly undergo wear and tear. When joint in human body undergo wear and tear for a significant amount of time the joint eventually fails, causing an injury. Annually, 100 out of 1000 players suffer from ACL ruptures which require reconstruction of ACL through surgery [1]. Reconstruction of ACL is a lengthy process and the person had to undergo at least 8-12 months of recovery period [2]. Also, ACL in Knee helps stabilize the joint by keeping the femur and tibia perfectly aligned and has one connection on the femur bone and other on the tibia, as shown in **Figure 1**.

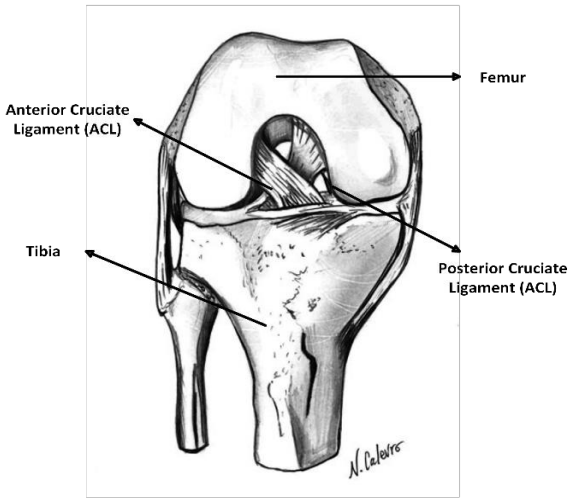

Figure 1. Anatomy of knee joint representing position of anterior cruciate ligament and posterior cruciate ligament in Knee joint.

ACL provides 85 percent anterior stability to knee joint during flexion [3]. Based on the diagnosis, reconstruction of the ligaments grafts and rehabilitation protocols are constructed. There are many techniques to diagnose the ACL injury including both invasive and non-invasive methods. Arthroscopy is an invasive method used for the diagnosis of ACL injuries, but it has potential complication due to its invasive nature. Patients who undergo the surgery develop risk of Pulmonary embolism which happens 2.8 cases for every 10,000 cases [4]. For diagnosis of the disease or an injury, different non-invasive imaging modalities are used such Computed tomography (CT), Magnetic Resonance Imaging (MRI), X-ray etc. MRI is the most widely used imaging technique

because it provides not only better soft tissue contrast and 3d Tomographic images but have better tendency to show physiological changes in scans. Another key significance of magnetic resonance imaging is that it is a radiation free modality unlike CT, X-rays [5, 6].

Humans are the most intelligent beings on the planet and what makes them intelligent is the brains that comprises of billions of neurons [7]. The foundation of the artificial intelligence is based upon the anatomy of the human brain. The importance of Artificial intelligence in disease detection and diagnosis is increasing because they can detect and diagnose the disease precisely and with less computational resources. Every day, researchers are employing modern techniques based on machine learning algorithms for better evaluation and detection of the injuries one injury being ACL tears in Knee joints. Researchers are incorporating machine learning algorithms in medical imaging to obtain better results which are not only precise but also clinically significant. An Artificial Intelligence or a machine learning model consist of 3 layers. First layer is input layer, in which we provide the data on which the training of the machine learning model will be done. The input data is fed to the hidden layer. These layers are responsible for learning key features present in the dataset, for instance in case for image data convolutional neural networks perform better during classification and segmentation of the injury or the disease due to its convolutional layers. In CNN hidden layer consist of three main layers,(1) Convolution layers: uses convolution to form features maps by moving 2-Dimensional feature detector kernel of mainly 3x3 across the entire image, (2) Polling: use to reduce the size of the convolutional layer output by either max-pooling or average pooling, (3) Fully connected layer: activation function is applied to determine the class of the input image (intact or injured) at the end is output layer. See **Figure 2** for graphical representation of basic Convolutional Neural Network.

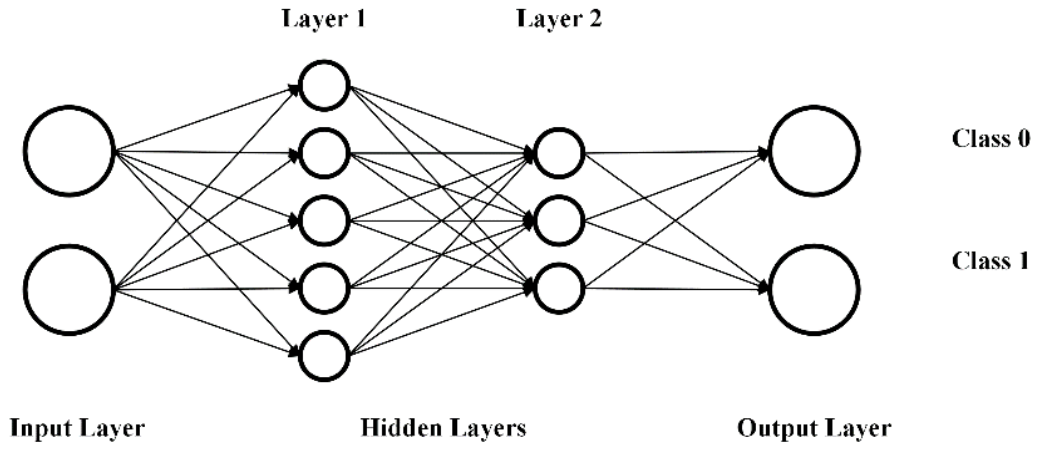

Figure 2. Convolutional Neural network and its fundamental building blocks.

After model building, we train the machine learning model on specific dataset such as mammographs of breast for cancer detection, CT scans of chest for COVID-19 detection, detection of diabetic retinopathy (DR) using retinal scans. For evaluation of the trained model, it is tested or validated on unseen data. This is done by feeding the trained model unlabeled data and record the true positives and true negative [8].

Machine learning techniques are sometimes unable to classify certain features such as small lesion and organs in human body due to simple nature of the models [9]. Today deep learning models are becoming more efficient, robust, and precise in detection, classification, and segmentation tasks. Deep learning models are more complex, wider networks that gives them better chance to differentiate multiple classes in dataset. In deep learning, the amount of the data and the image quality highly effect the performance of the model. Better image quality and subsequently large amount of data helps the deep learning model to classify more accurately. In this study we developed a custom 3D convolution neural network for the detection ACL tear in Knee MRI scans. The datasets for this study are publicly available. One dataset is provided by the Stanford ML group which contain 1250 Knee MRI scans along with the annotations. Second is provided by the hospital in Croatia which has 917 scans divided into 3 classes (Intact, partial tear and completely ruptured). Before training we applied pre-processing on the given data to improves our model performance. Our model is robust, lightweight having only 153k trainable parameters and takes less computational power to train.

Chapter 2: Literature Review

Researchers are developing new models to achieve better results in diagnosis of disease. In this study [10] used a semi-automated approach which includes extraction of features through manually constructed ROI using Histogram-oriented gradient and gist descriptor. Two machine learning algorithms were used to achieve best performance metrics, support vector machine and random forest. The dataset used for the study is obtained from the Clinical Hospital Center in Rijeka, Croatia which consists of 917 exams divided into 3 classes (intact, partially injured and completely ruptured). The HOG descriptors coupled with support vector machine obtained best results. The performance metrics were calculated for two classes one injured which gave an AUC value of 0.894 and other completely ruptured which gave AUC value of 0.943.

In this study [11] developed MRNet for the classification of ACL injuries in Knee MRI Scans. They used Stanford University medical center dataset containing 1370 knee MRI scans. For training 1130 exams were used, for testing 120 exams and additional 120 exams for validation of the model. The model achieved and accuracy of 0.937, 0.965 and 0.847 for abnormalities, ACL tears and meniscus tear respectively. And AUC of 0.911 on external validation dataset consist of scans from the Croatia Hospital.

In this study [12] employs convolutional neural networks for the detection of ACL tear in Knee MRI scans. The dataset used for this study consisted of 260 MRI scans of Knee joint. For training and testing, 200 cases were used and remaining 60 scans for validation of the model. They used 3 different crop configurations (1) cropped (2) dynamic cropping (3) uncropped. The accuracy achieved for each configuration was 0.68 for uncropped images, 0.72 for cropped images and 0.765 for dynamic cropping.

In this study [13] used multiple state of the art model including GoogleNet [14], AlexNet [15], and RESNET [16] on both Multiview and single view configuration. They used Stanford university medical center dataset that contained 1370 knee MRI scans. The performance metric AUC for single-view model was 0.8387 for AlexNet, 0.7779 for RESNET and 0.7605 for GoogleNet while for multi-view model it was 87.87 for Alex-Net, 85.96 for Google-Net and 85.79 for RESNET. In study, it shows that out of the three classes meniscus tear was the most challenging. For the classification of Knee MRI injuries, [17] used 18-layered residual neural network. The study uses dataset provided by the Stanford ML group that contains 1370 knee MRI

scans. Interpolation technique was used to equalize the number of slices. The average AUC was found to be 0.9557 for ACL tear, 0.9081 for Meniscus tear and 0.9381 for abnormalities.

In this study [18] used custom convolutional neural network to detect ACL tear. The dataset used in this study was the one provided by Stanford ML group contain 1130 scans of Knee MRI and additional 120 scans in validation folder. They combined the two folders and split it for training and testing in 90/10 ratio (1124 exams for training and 125 exams for testing purposes). Further they divided the dataset into two categories, one with the Laplacian filter applied and the other one without the application of Laplacian filter. They achieved significant accuracy gains with Laplacian filter that was 92.8% as compared to non-Laplacian dataset that was 89.2%.

In this study [19] employs a multimodal feature fusion model for the detection of ACL injuries in Knee MRI scans. The dataset used for this study contained MRI scans of 30 patients. The model showed accuracy of 92.17% in detecting ACL injury. The results compared with the invasive procedure for the diagnosis of ACL tear in Knee MRI.

In this study [20] proposed a Deep convolutional neural network-based inception-v3. For training they used transfer learning approach which yields better results in case of limited amount of data. They used dataset provided by the Stanford ML group which contains 1370 knee MRI scans. For training and testing they used 70% (959 scans) and 30% (411 scans) respectively. When tested on the unseen data, the best performing model achieved the highest test accuracy of 95.42% as compared to other deep learning models.

In this study [21] proposed a custom RESNET-14 model for the classification of Knee injuries. The dataset used in this study contained 917 knee MRI scans which were divided into 3 class (intact (690 scans), partial tear (172 scans) and completely ruptured (55 scans)). They used a hybrid technique for class imbalance, in which the minority classes were up-sampled, and majority class was down-sampled to reduce the gap between each class in terms of number of samples. They achieved an average accuracy, sensitivity, and precision of 92%, 91%, and 91% respectively and average AUC and specificity for all three classes was 98% and 95% respectively, with 5-fold cross validation. This study stated that instead of artificially increasing the size of the data, class weights should be used.

In this study [22] proposed a new technique called MRPyrNet which exploits the localizations of injuries in diagnostic scans. The study combines MRPyrNet with state-of-the-art model MRNet and ELNet and compares the improvement in the performance of baseline models. Two datasets were used separately in this study, one provided by Stanford ML group which has 1370 Knee MRI scans (1130 scans for training, 120 for validation and 120 for test set) and second dataset is the one provided by the Clinical Hospital Center in Rijeka, Croatia which has 917 Knee MRI scans annotated into 3 separate classes (intact, partially injured and completely ruptured). The new architecture, combined with state-of-the-art model show significant improvements in performance. the traning was done seperatley on the two datasets. The AUC for ACL tear for MRNet and ELNet with MRPyrNet report gains of 2.2% and 2.1% as compared to baseline models. In case of meniscal tear, the improvement was nearly 5.5% and 2.9% for MRNet and ELNet respectively.

In this study [23] employs a novel RESNET-50 model for the detection of injuries in Knee MRI scans. The dataset used in this study has 1250 knee MRI scans provided by the Stanford ML group. They achieve better performance by coupling RESNET-50 with denoising auto-encoders. The model classified the injuries with respect to individual plane. In this study the dataset has three planes (1) Sagittal Plane, (2) Coronal Plane and (3) Axial plane and in each plane the model classified the given scan in three classes (1) Abnormalities, (2) ACL Tear and (3) Meniscus Tear. The accuracy and AUC value for ACL tear was 0.7881 and 0.8947 in sagittal plane, 0.7583 and 0.8297 in coronal plane, 0.8319 and 0.8721 in axial plane, respectively.

Further detailed information about the accuracy and AUC of the previously done work is described in Table I.

Table I Summary of the related work done regarding classification of knee joint injuries from MRI scans using machine learning algorithms.

Authors	**Dataset**	**Metrics**	
		Accuracy	**AUC**
[10]	Croatia Knee MRI Dataset (917 cases)	-	Injured: 0.894 Completely Ruptured: 0.943

[11]	MRNet (1130)	-	Abnormalities: 0.937; ACL tears: 0.965 Meniscus tears: 0.847 External validation: 0.824
[12]	260 MRI scans	Single slice (uncropped: 0.68, cropped: 0.72 and cropped + dynamic: 0.765) 3 slices: 0.865 5 slices: (CrossValidation:0.915; Test: 0.967)	-
[13]	MRNet 1370 scans	AlexNet: Single-view: 0.7639; Muti-view: 0.8166 Resnet-18: Single-view: 0.7537; Multi-view: 0.8167 GoogleNet: Single-view:0.7046; Multi-view: 0.7806	AlexNet: Single-view: 0.8387; Muti-view: 0.8787 Resnet-18: Single-view: 0.7779; Multi-view: 0.8579 GoogleNet: Single-view:0.7605; Multi-view: 0.8596
[17]	MRNet 1370	-	Average AUC for best performing configuration: ACL tear: 0.9557 Meniscus Tear: 0.9081; Abnormal: 0.9381
[18]	MRNet 1249	Laplacian flter applied: 0.928 Without Laplacian filter: 0.892	-

[19]	30 Knee MRI scans	ACL Tear detection: 0.9628 (Sagittal plane) Meniscus Tear: 0.7537 (Sagittal plane)	ACL Tear detection: 0.9726 (Sagittal plane) Meniscus Tear: 0.923 (Sagittal plane)
[20]	MRNet 1370	ACL Tear: VGG16: 0.8545; VGG19:0.8790 Inception ResNet-v28: 0.8991 Xception: 0.9225; Proposed Inception-v3: 0.9542	-
[21]	Croatia Knee MRI Dataset (917 cases)	Average 5-fold cross validation: 0.9200	Average 5-fold cross validation for all three classes: 0.9800
[22]	MRNet 1370	(MRPryNet + MRNet) ACL Tear: 0.886 Meniscus Tear: 0.808 (ELNet + MRPryNet) ACL Tear: 0.881 Meniscus Tear: 0.761	(MRPryNet + MRNet) ACL Tear: 0.976 Meniscus Tear: 0.889 (ELNet + MRPryNet) ACL Tear: 0.960 Meniscus Tear: 0.895
	Croatia Knee MRI Dataset (917 cases)	(MRPryNet + MRNet) ACL Tear: 0.834 (ELNet + MRPryNet) ACL Tear: 0.851	(MRPryNet + MRNet) ACL Tear: 0.914 (ELNet + MRPryNet) ACL Tear: 0.900

[23]	MRNet 1250	Saggital Plane: Abnormalities: 0.8898 ACL tears: 0.7881 Meniscus tears: 0.7712	Saggital Plane: Abnormalities: 0.9316 ACL tears: 0.8947 Meniscus tears: 0.7987
		Coronal Plane: Abnormalities: 0.8667 ACL tears: 0.7583 Meniscus tears: 0.75	Coronal Plane: Abnormalities: 0.8029 ACL tears: 0.8297 Meniscus tears: 0.7393
		Axial Plane: Abnormalities: 0.8992 ACL tears: 0.8319 Meniscus tears: 0.6891	Axial Plane: Abnormalities: 0.8596 ACL tears: 0.8721 Meniscus tears: 0.7075

Chapter 3:Methods and Material

3.1 Dataset

In this study two datasets are being used for the classification of ACL injuries in Knee MRI scans. One from Stanford ML group and the other from Clinical Hospital Centre Rijeka in Croatia. Total dataset is divided into two parts 80 percent to train the model and remaining for testing the performance of the trained neural network. The dataset provided by Stanford ML Group was primarily divided into three separate views (coronal, sagittal and axial) and the dataset provided by Clinical Hospital Centre Rijeka, Croatia had only one view (sagittal).

3.1.1 Stanford Knee MRI Dataset

The dataset consists of 1130 knee MRI scans provided by Stanford ML group. There is also a separate validation dataset containing 120 scans (66 normal scans and 54 injured scans). Dataset also contains .csv file which specify the labels associated with each scan (0 for normal ACL and 1 for ruptured ACL). There are total of 922 normal ACL scans and 208 injured scans. See **Figure 3.**

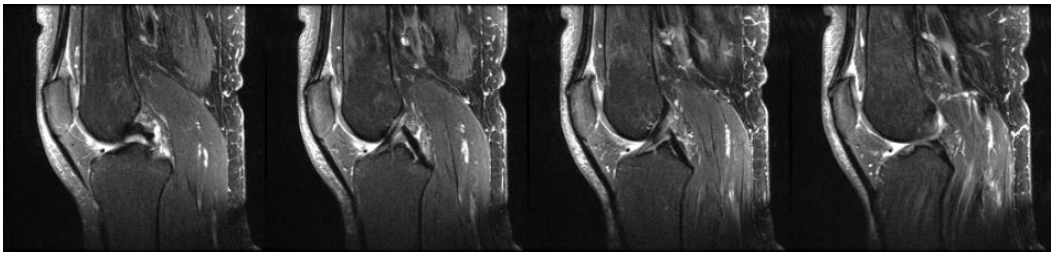

Figure 3. Sample image from the Stanford Knee MRI Dataset

3.1.2 Croatia Knee MRI Dataset

The dataset consists of 917 total knee MRI scans provided by the Clinical Hospital Centre Rijeka, Croatia performed between 2006 and 2014. Form 917 total scans, 690

are normal scans, 172 partially injured scans and 55 completely ruptured scans. For this study we are including only normal and completely ruptured scans. See **Figure 4.**

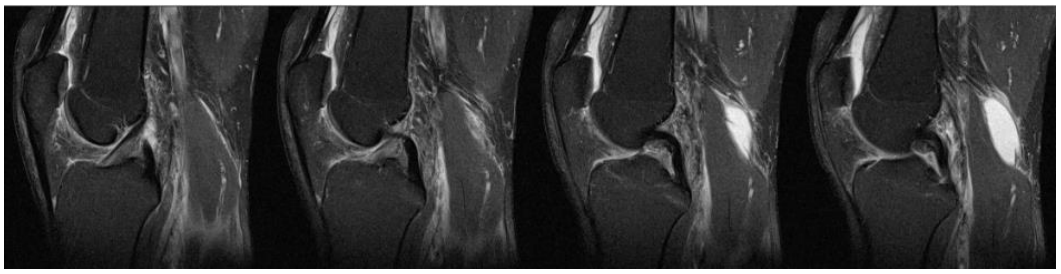

Figure 4. Sample image from the Stanford Knee MRI Dataset

3.2 Pre-processing

3.2.1 Equalization of Number of slices

We extracted the desired region of interest for slices for both the dataset. A .csv file was provided by the Clinical Hospital in Rijeka, Croatia containing ROI regions for anterior cruciate ligament. Based on that we obtained ROI number of slices varying from 2 to 6 slices. According to the anatomy of knee joint, ACL is located almost at the center from medial and lateral side of the knee, so we decided that the ideal size would be 4 number of slices (1 slices before the middle frame and 3 slices after) to

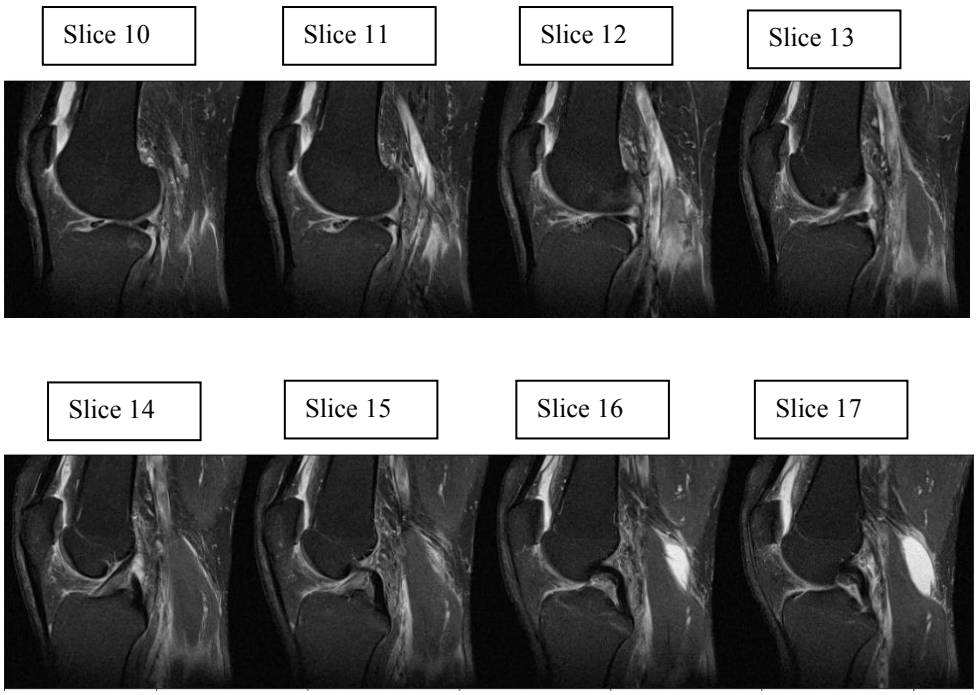

Figure 5. Slices from the original MRI scan before pre-processing

obtain maximum information. Similar technique was used by [21] which showed significant improvement in model performance. **Figure 5** shows the number of slices before ROI extraction. The selected ROI captures maximum information about ACL without losing and features. As shown in **Figure 6**.

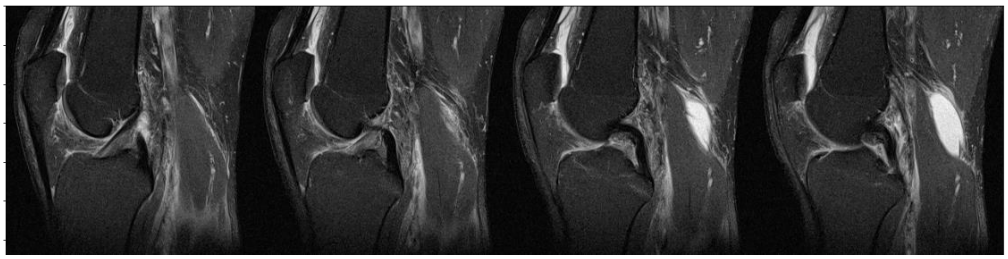

Figure 6. Desired ROI number of slices extracted from the original image

3.2.2 Size of the Image

In the dataset provided by the Stanford ML group, the size of the single 2D image from 3D stack has standard size of 256 x 256 while the dataset from Croatia has 2D image size ranges between 290 x 300 and 320 x 320.So, we resized all the images to a standard size of 200 x 200 x 4 using linear interpolation. The image is zoom in z-direction keeping the features of the image in x-y direction intact. We performed this step to obtain our desired cropped ROI with cutting any information out of our final image.

3.2.3 Intensity Normalization

The intensity values for both datasets were different. For Stanford dataset, the intensity values were between 0 to 255 while the scans from Croatia Knee MRI dataset had intensities between 0 to 1024 (12-bit). From the Figure I, the image from the Stanford Dataset is on the brighter side as compared to image from the Croatia Knee MRI

dataset. To overcome this problem, we applied standardize z-score normalization to scale the intensities between 0 and 1. As shown in **Figure 7.**

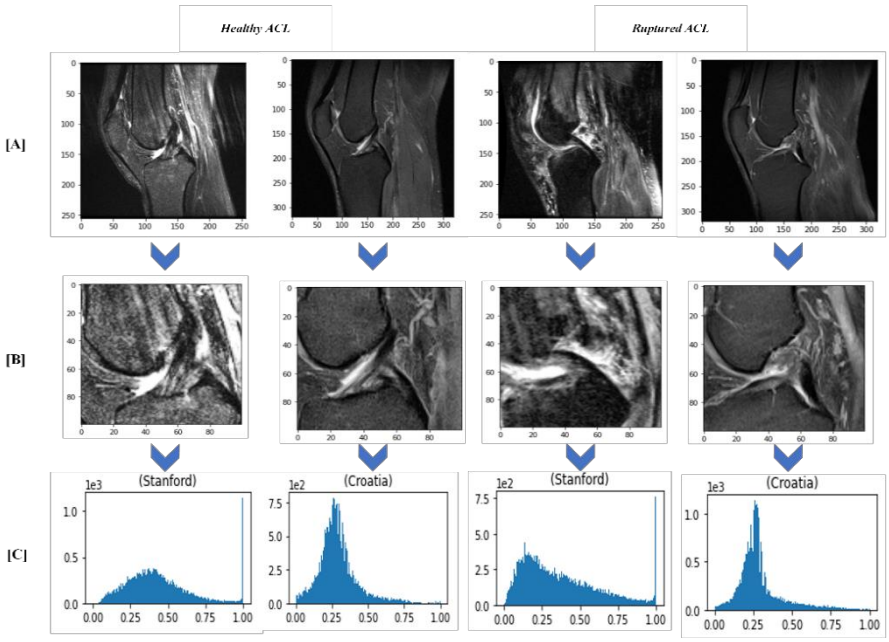

Figure 7. This figure shows the steps invovled in the preprocessing of Knee MRI scans before feeding it to the CNN.

3.2.4 Data Augmentation

We are using a hybrid technique to address the major different in number of cases in each class. We have increased the cases in minority class through data augmentation techniques. The data augmentation includes up-sampling of minority class, rotate the volumes randomly between [-15 15] degrees and translate them randomly between [-10 10] pixels. As shown in **Figure 8-9** respectively.

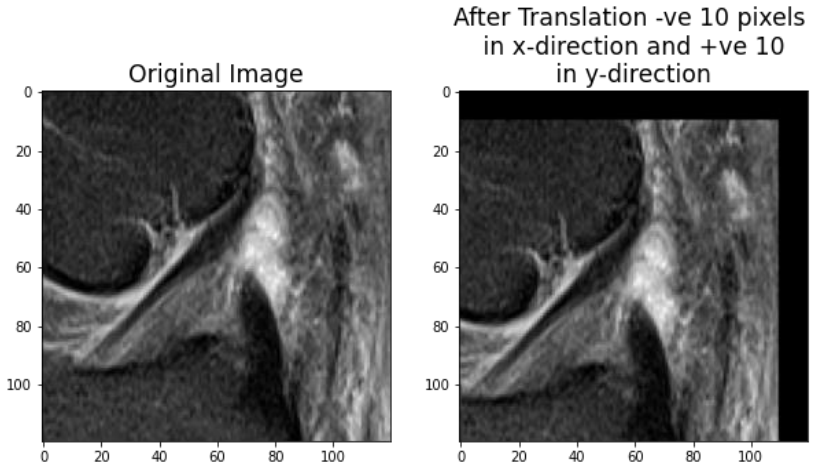

Figure 8. Translation of Image during data augmentation

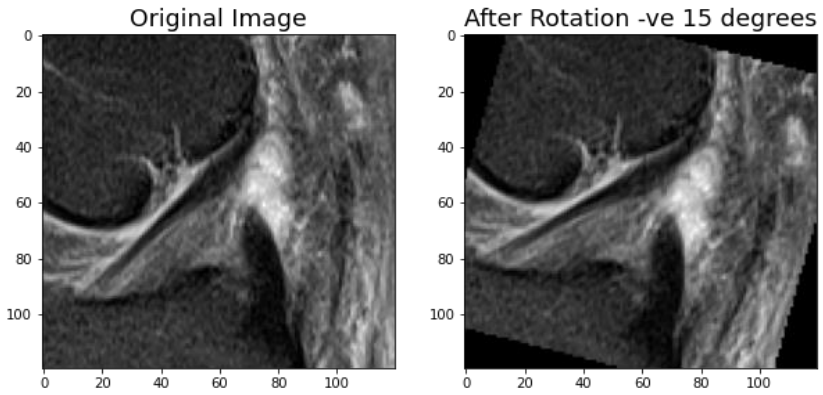

Figure 9. Rotation of Image during data augmentation

3.2.5 Class Imbalance

Both datasets have severe type of class imbalance. For instance, Stanford dataset has 922 normal and 208 injured ACL scans roughly 4 to 5 times more normal scans as compared to injured ACL scans. In case of Croatia Knee MRI dataset, it has 12 times more normal scans as compared to injured scans. As shown in **Table II**. There are multiple methods to overcome the class imbalance problem like over sampling and under sampling, data augmentation of minority class and by using class weightage. We opted for hybrid method to rectify the effect of class imbalance on training of our model. For this we used data augmentation to decrease the difference in number of cases between normal class and injured class as a result, the number of cases in injured

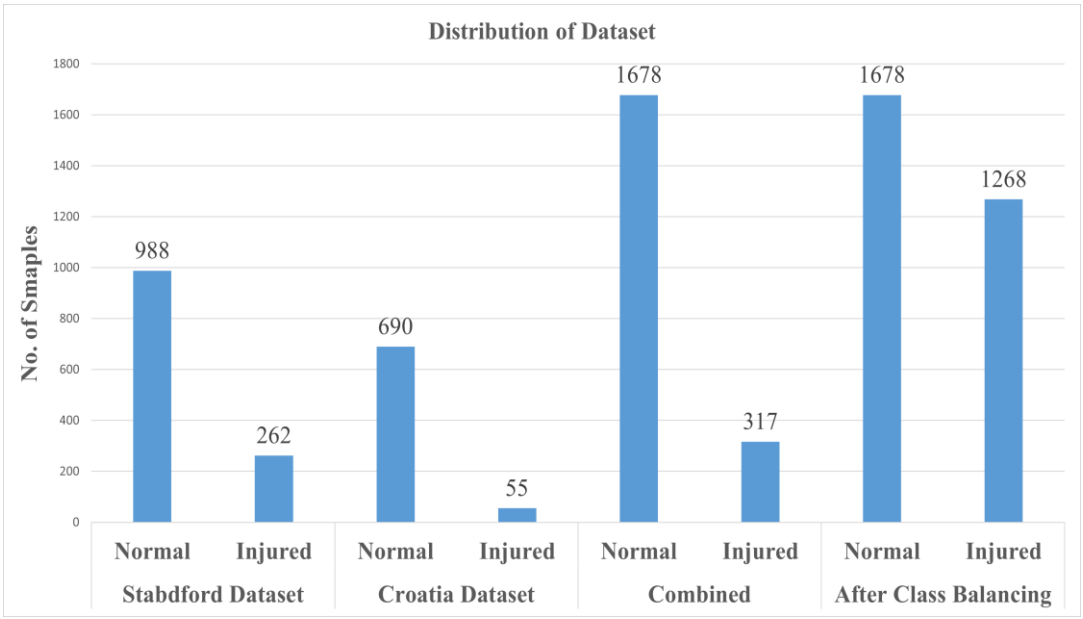

Figure 10. Distribution of Injured and normal classes in each configuration of dataset.

class became 1268 and in normal class 1678, as shown in **Figure 10**. After augmentation we calculated the class weights to eliminate the effect of remaining imbalance present in the dataset. The details for the class weights are given in **Table III**.

Table II. Class imbalance ratios in dataset

Dataset	**Normal Scans**	**Injured scans**	**Imbalance Ratio (Normal to Injured)**
Stanford Knee MRI dataset	988	262	3.7:1
Croatia Knee MRI dataset	690	55	12.5:1
Total	1678	317	5.3:1

Table III Distribution of Dataset and Class Weights are given after Data Augmentation

	Normal Scans	**Injured Scans**
Combined dataset	1678	317
After data augmentation	1678	1268
Class Weights	0.87	1.16

3.2.6 Convolutional Neural Network

For training we used Custom 3D convolutional neural network which takes 4-D tensor as an input with shape 120 x 120 x 4 x 1. We used Conv3D layer with maxpooling3D and batch normalization. As it is a binary classification model, so we used sigmoid activation function. The Network we are using is small, approximately 2MB and have 153K trainable parameters. We tuned the hyper-parameters to get the best out of our model. Graphical representation of our custom 3D convolutional neural networks and complexity of our model is shown in **Figure 11-12**.

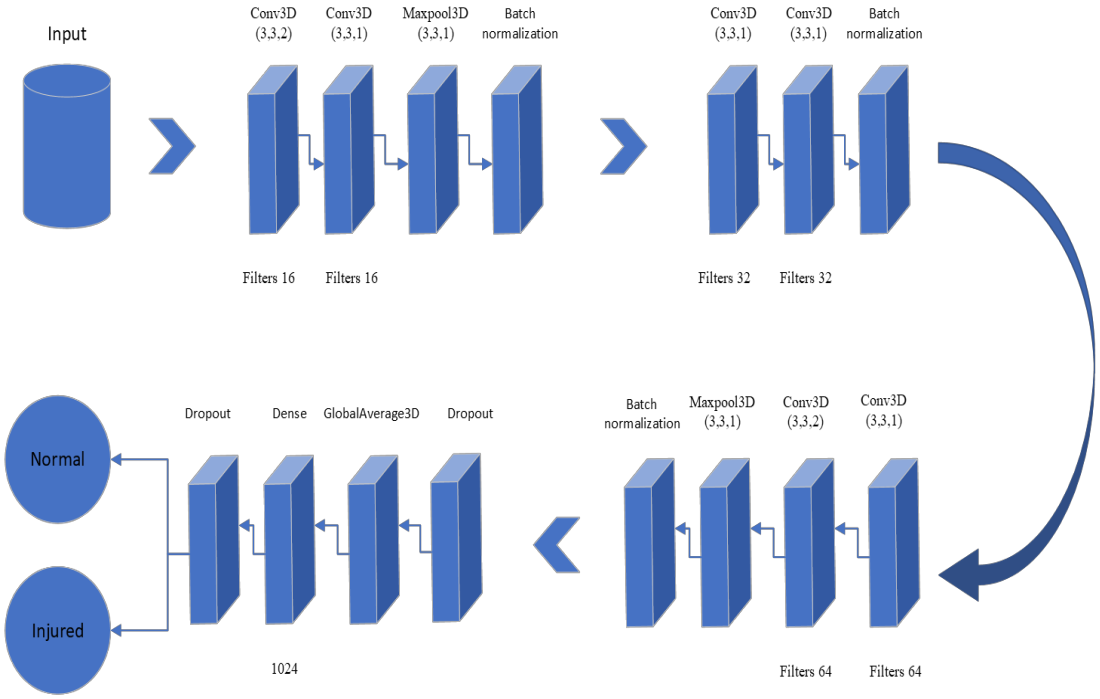

Figure 11. This figure shows the detail of the model used for the classification of ACL injureies in knee MRI scans.

```
Model: "3dcnn"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 120, 120, 4, 1)]  0
_________________________________________________________________
conv3d (Conv3D)              (None, 118, 118, 3, 16)   304
_________________________________________________________________
conv3d_1 (Conv3D)            (None, 116, 116, 3, 16)   2320
_________________________________________________________________
max_pooling3d (MaxPooling3D) (None, 38, 38, 3, 16)     0
_________________________________________________________________
batch_normalization (BatchNo (None, 38, 38, 3, 16)     64
_________________________________________________________________
conv3d_2 (Conv3D)            (None, 36, 36, 3, 32)     4640
_________________________________________________________________
conv3d_3 (Conv3D)            (None, 34, 34, 3, 32)     9248
_________________________________________________________________
batch_normalization_1 (Batch (None, 34, 34, 3, 32)     128
_________________________________________________________________
conv3d_4 (Conv3D)            (None, 32, 32, 2, 64)     36928
_________________________________________________________________
conv3d_5 (Conv3D)            (None, 30, 30, 2, 64)     36928
_________________________________________________________________
max_pooling3d_1 (MaxPooling3 (None, 10, 10, 2, 64)     0
_________________________________________________________________
batch_normalization_2 (Batch (None, 10, 10, 2, 64)     256
_________________________________________________________________
dropout (Dropout)            (None, 10, 10, 2, 64)     0
_________________________________________________________________
global_average_pooling3d (Gl (None, 64)                0
_________________________________________________________________
dense (Dense)                (None, 1024)              66560
_________________________________________________________________
dropout_1 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 1025
=================================================================
Total params: 158,401
Trainable params: 158,177
Non-trainable params: 224
_________________________________________________________________
```

Figure 12. Summary of our proposed light weight convolutional neural network

3.2.7 Hyper-parameters

The hyper-parameters that effect the learning process greatly are learning rate, number of epochs, batch size and dropout regularization method. For classification, we are using sigmoid function and binary-cross entropy as loss function to classify ACL tear in MRI scans. We trained the model from scratch for 120 number of iterations or

epochs. Other details regarding hyper-parameter tuning of our model training is given in **Table IV**. We are using different attributes to better evaluate our model. These attributes include precision, recall, f1-score, and AUC. We cross validated the model using 5-fold cross validation and each fold was extracted using stratified sampling technique. The model was trained on 8-Gigabyte RTX2060Super with 32-GB DDR4 memory. It took 10 minutes to train the model. **Figure 13.** shows a brief overall methodology of our proposed study to detect ACL injuries in Knee MRI scans.

Table IV Hyperparameter for Model Training

Tuning Parameter	Description
Initial Learning Rate	3e-5
Epochs	120
Batch size	32
Optimizer	Adam
Loss	Binary cross-entropy

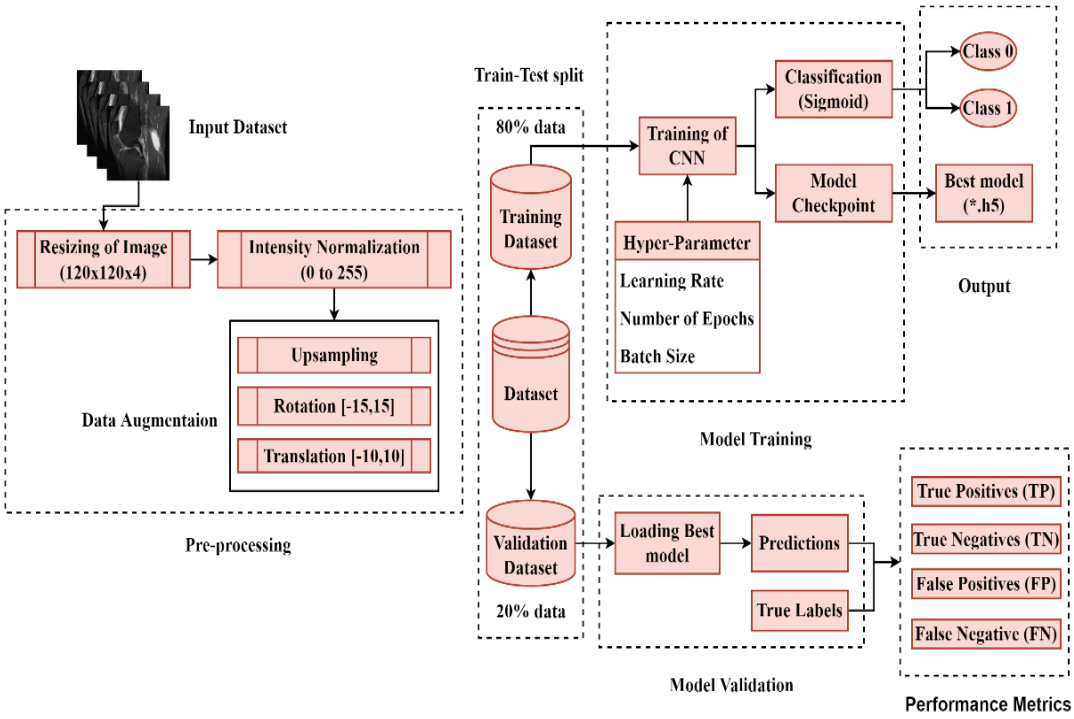

Figure 13. Brief overview of our proposed methodology.

3.2.8 Performance metrics

The performance of a deep learning model is estimated by calculating the following four attributes:

Suppose positive case are those in which patient have intact ACL or Normal ACL and Negative cases are those in which there is tear in the ACL ligament.

- True Positive (TP): Model predicts the Normal class as Normal
- False Positive (FP): Model predicts the Injured class as Normal
- True Negative (TN): Model predicts the Injured class as Injured
- False Negative (FN): Model predicts the Normal class as Injured

By using the attributes mentioned above we can find the accuracy, precision, recall and F1-score of the trained model. The predictions are created on the test set.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$specificity = \frac{TN}{TN + FP}$$

$$sensitivity/recall = \frac{TP}{TP + FN}$$

$$f1 - score = \frac{2 \times (precision \ \ times \, recall)}{precision + recall}$$

Chapter 4: Results

The results were obtained by training the model on 80% of the total data and was validated on the remaining 20%. After training we obtained average accuracy of 97.6% and average AUC of 99.3 on the validation set, which is better than the previously purposed networks. For further performance evaluation of our model, we plot confusion matrix for each fold. It shows how many scans were predicted accurately by our model. **Figure 14-18** represents AUC-ROC curve of our trained model, having true positive rate on the y-axis and false positive rate on the x-axis. The highest value for validation accuracy in each fold was 0.9796, 0.9643, 0.9728, 0.9762, and 0.9915 and the minimum value for the validation loss was 0.0843, 0.1152, 0.0906, 0.0802, and 0.0213 respectively. These values are less than any of the previously proposed and state of the art models. **Figure 19** shows the training process of our CNN model with respect to each epoch.

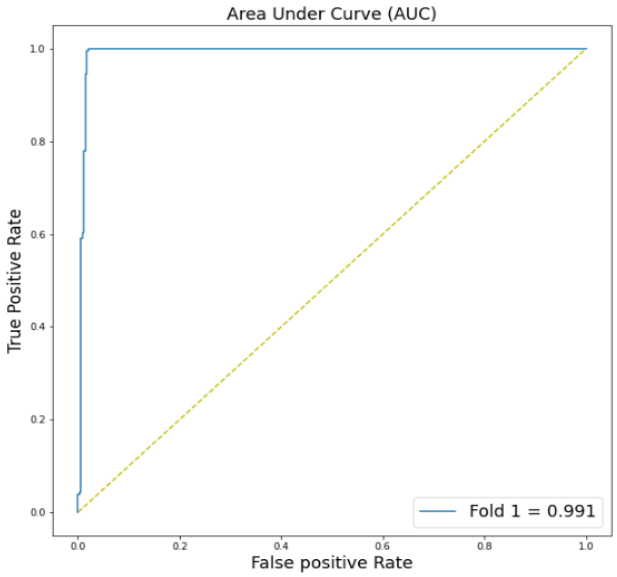

Figure 14. Area under the Curve for First fold during 5-fold cross validation.

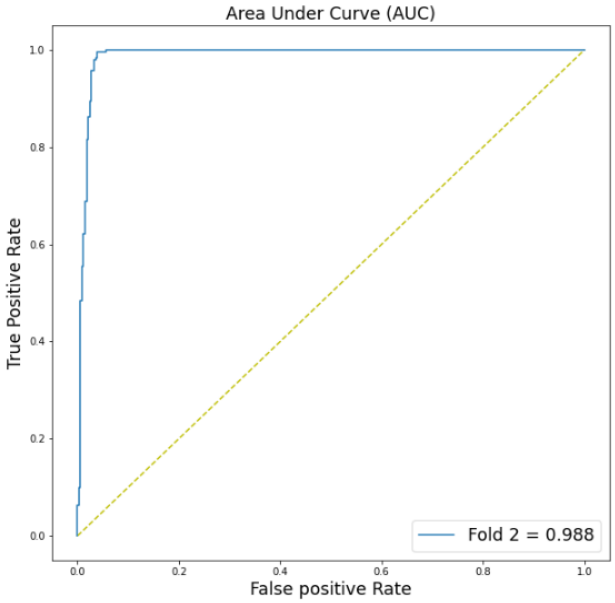

Figure 15. Area under the Curve for second fold during 5-fold cross validation.

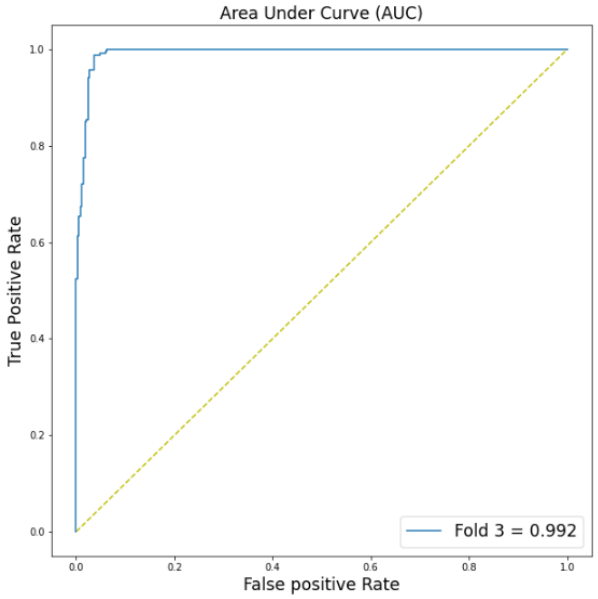

Figure 16. Area under the Curve for Third fold during 5-fold cross validation.

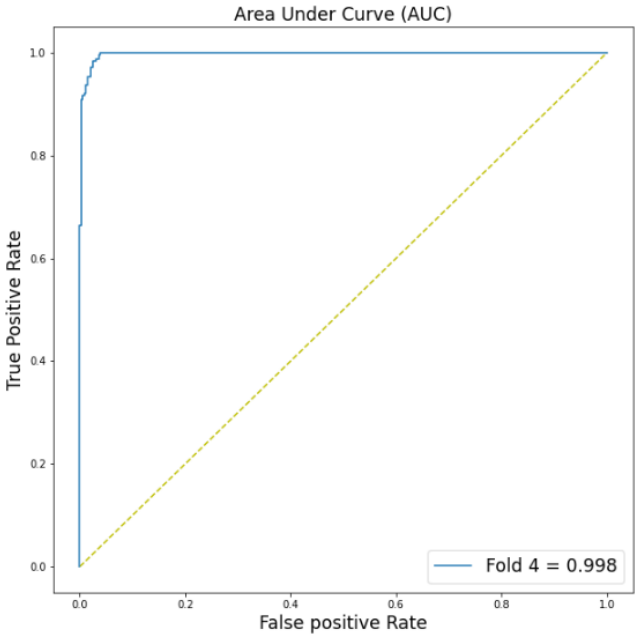

Figure 17. Area under the Curve for fourth fold during 5-fold cross validation.

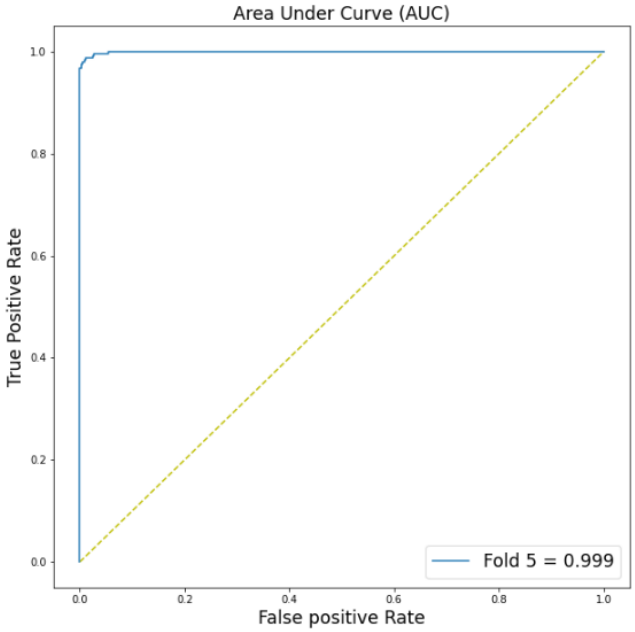

Figure 18. Area under the Curve for fifth fold during 5-fold cross validation.

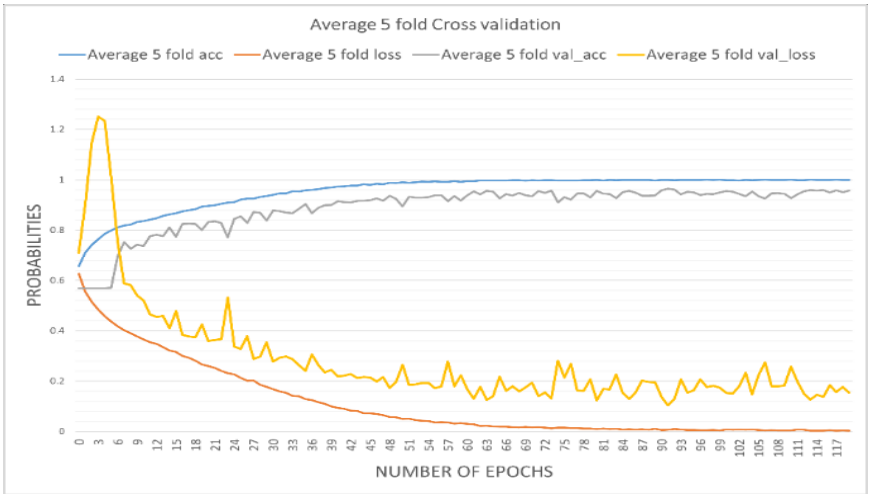

Figure 19. This figure shows the graphical representation of the training of our convolutional neural network. The graph shows the value of training accuracy, training loss, validation accuracy, validation loss with respect to each epoch.

We recorded performance metrics which includes training accuracy, training loss, validation accuracy and validation loss. The validation accuracy and loss for average 5-fold cross validation was 97.6 and 0.07832, respectively. **Figure 20-24** shows the confusion metrics drawn for each fold representing the correct and incorrect predictions made by our trained model by comparing it with true labels. The obtained performance is better that all the state-of-the-art models which includes VGG16, VGG19, Inception- Resnet, Xception, etc. and recently proposed models. See **Table V** for comparison with the previously used models. The comparisons shows that not only our model outperforms previously applied models but also is robust and lightweight. Also, **Table VI** shows the classification report of each fold of our trained model.

Table V Comparison with Previous best performing CNN model

Authors	Accuracy	Precision	Recall	F1-score	AUC
[20]	95.42	95.02	95.13	94.83	-
[24]	-	-	98.1		98.3
Proposed	97.6	97.6	97.6	97.4	99.3

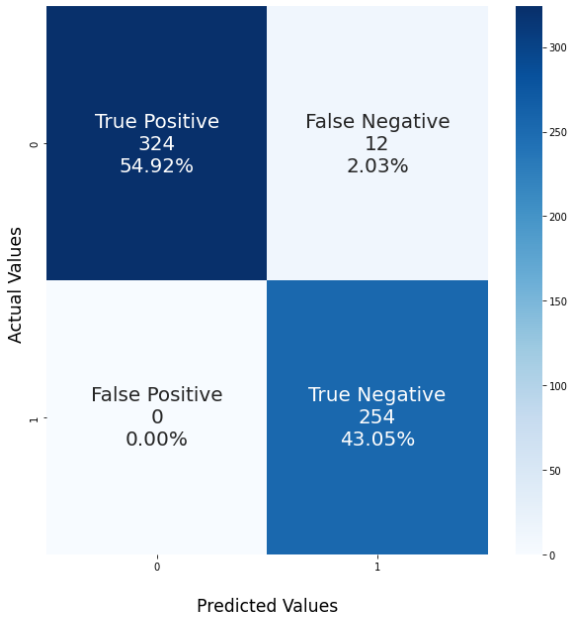

Figure 20. Confusion matrix for fold 1 shows that the model predicted 97.69% of the cases correctly given in the validation set

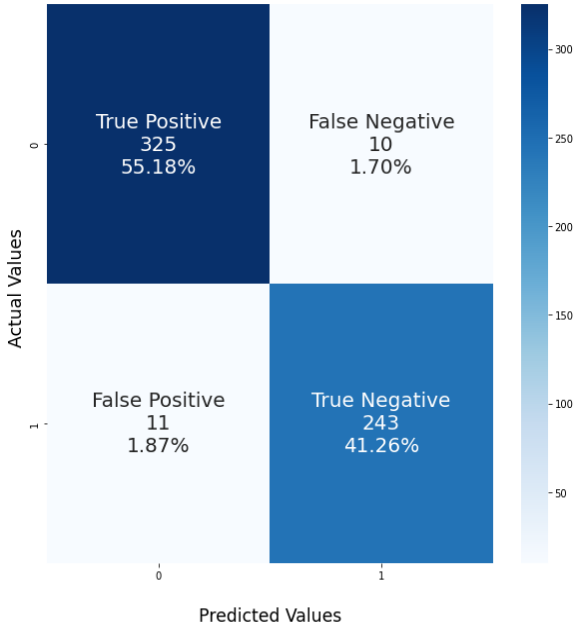

Figure 21. Confusion matrix for fold 2 shows that the model predicted 96.43% of the cases correctly given in the validation set.

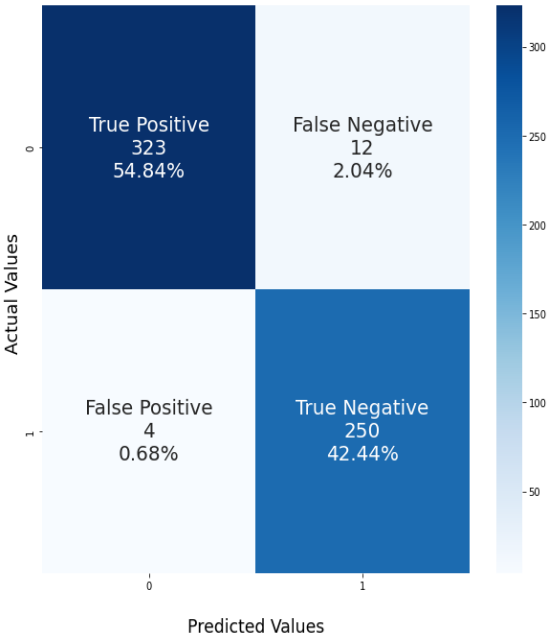

Figure 22. Confusion matrix for fold 3 shows that the model predicted 97.28% of the cases correctly given in the validation set.

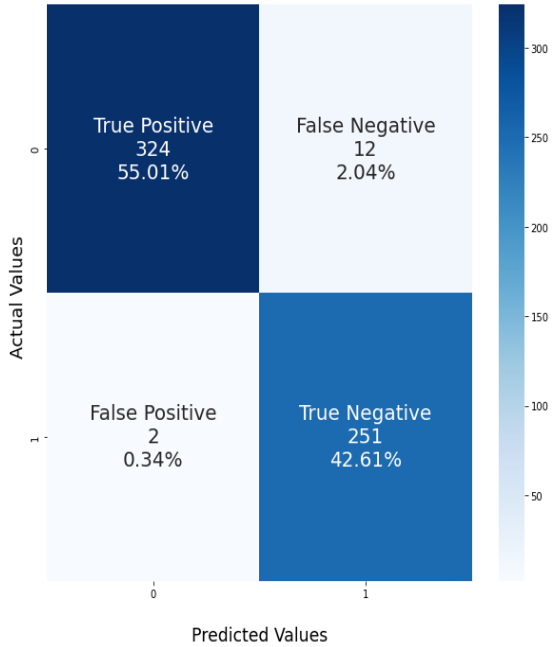

Figure 23. Confusion matrix for fold 4 shows that the model predicted 97.62% of the cases correctly given in the validation set.

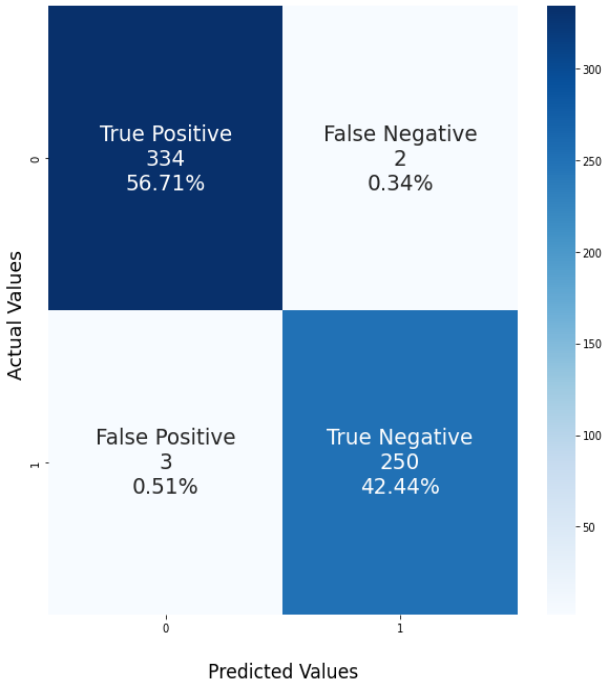

Figure 24. Confusion matrix for fold 5 shows that the model predicted 99.15% of the cases correctly given in the validation set.

Table VI Classification report of individual folds

No. of Folds	Class	Precision	Recall	F1-score
Fold 1	Normal	0.97	0.97	0.97
	Injured	0.96	0.96	0.96
	Weighted Average	0.96	0.96	0.96
Fold 2	Normal	0.98	0.96	0.97
	Injured	0.95	0.98	0.97
	Weighted Average	0.97	0.97	0.97
Fold 3	Normal	0.98	0.96	0.97
	Injured	0.95	0.98	0.97
	Weighted Average	0.97	0.97	0.97

Fold 4	Normal	0.99	0.96	0.98
	Injured	0.95	0.99	0.97
	Weighted Average	0.98	0.98	0.98
Fold 5	Normal	0.99	0.99	0.99
	Injured	0.99	0.98	0.99
	Weighted Average	0.99	0.99	0.99

Chapter 5:Discussion

In this study we developed a custom 3D-CNN and trained it from scratch. The model we developed has very small number of trainable parameters (153K) because the knee MRI scans have very small number of distinguishable feature so a light- weight model can provide better performance than a large and complex model which tends to over-fitting. [24] developed a custom lightweight 3D deep learning model which had only 43 thousand trainable parameters. Model that achieved the best accuracy has 161,921 trainable parameters, achieving an AUC-ROC of 98 and 98.3 on two different datasets thus proving that a lightweight model can be used to detect ACL tear in Knee MRI scans. Over-fitting happens when the equipped data is simple as compared to the complexity of the model. So, rather learning the features in an image the model memorizes it. This causes the training accuracy to be very high as compared to validation accuracy because model fails to generalize on validation dataset or unseen MRI scans. To avoid this, we equipped a model that is more robust and able to generalize the classification information learned during training on the validation set or un-seen data. To avoid Overfitting of our model we artificially increased the size of our data for better training. For this we applied data augmentation and class imbalance technique to help the model learn better. Data-augmentation includes up-sampling, rotation, and translation. There are two publicly available datasets for Knee MRI diagnosis, one provided by Stanford ML group which contains 1370 knee MRI scans in 3 different plans (coronal, sagittal and axial) and 3 categories (abnormal, ACL and meniscus) with additional 120 scans for validation separately, and the other by Clinical Hospital Centre Rijeka in Croatia with contains 917 Knee MRI scans divided into 3 categories Normal, partially injured and completely ruptured. For classification, we are using sigmoid function and binary-cross entropy as loss function to classify ACL tear in MRI scans. We trained the model from scratch for 120 number of iterations or epochs with the batch size of 32 and learning rate of 3e-5 using 5-fold cross validation. Each fold took approximately 10 minutes to train. To save the best performing model we used model-checkpoints.

Chapter 6:Conclusion

In this study we developed a lightweight 3D CNN that classifies ACL injuries in Knee MRI scans. We combined two datasets, one provided by Stanford ML group which contain 1130 knee MRI scans with additional 120 scans and the other obtained from Clinical Hospital Centre Rijeka which contains 917 knee MRI scans. We used eighty percent of our data for the training of our model and remaining twenty percent for the validation purposes. for better training we increased the augmented the dataset and deal with the class imbalance through class weights [25]. Our model accurately classifies the scans with 0.976 accuracy on the validation set and 0.0783 validation loss. This offers an opportunity for automated detection of ACL injuries in Knee MRI scans and help in better management of the patients in clinical practice. We are using precision, recall, f1-score, and AUC for the evaluation of our model. We cross validated the model using 5-fold cross validation with stratified sampling technique to have equaled proportion of normal and injured scans.

References

1. Norris, R., P. Thompson, and A. Getgood, *Suppl 3: The Effect of Anterior Cruciate Ligament Reconstruction on the Progression of Osteoarthritis.* The Open Orthopaedics Journal, 2012. **6**: p. 506.
2. Zaffagnini, S., et al., *Return to sport after ACL reconstruction: how, when and why? A narrative review of current evidence.* Joints, 2015. **3**(1): p. 25.
3. Vaienti, E., et al., *Understanding the human knee and its relationship to total knee replacement.* Acta Bio Medica: Atenei Parmensis, 2017. **88**(Suppl 2): p. 6.
4. Hetsroni, I., et al., *Symptomatic pulmonary embolism after outpatient arthroscopic procedures of the knee: the incidence and risk factors in 418,323 arthroscopies.* J Bone Joint Surg Br, 2011. **93**(1): p. 47-51.
5. Brody, H., *Medical imaging.* Nature, 2013. **502**(7473): p. S81-S81.
6. Prickett, W.D., S.I. Ward, and M.J. Matava, *Magnetic resonance imaging of the knee.* Sports Medicine, 2001. **31**(14): p. 997-1019.
7. Roth, G. and U. Dicke, *Evolution of the brain and intelligence.* Trends Cogn Sci, 2005. **9**(5): p. 250-7.
8. Krogh, A., *What are artificial neural networks?* Nature Biotechnology, 2008. **26**: p. 195-197.
9. Suzuki, K., *Overview of deep learning in medical imaging.* Radiol Phys Technol, 2017. **10**(3): p. 257-273.
10. Štajduhar, I., et al., *Semi-automated detection of anterior cruciate ligament injury from MRI.* Computer methods and programs in biomedicine, 2017. **140**: p. 151-164.
11. Bien, N., et al., *Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet.* PLoS medicine, 2018. **15**(11): p. e1002699.
12. Chang, P.D., T.T. Wong, and M.J. Rasiej, *Deep Learning for Detection of Complete Anterior Cruciate Ligament Tear.* J Digit Imaging, 2019. **32**(6): p. 980-986.
13. Irmakci, I., et al. *Deep learning for musculoskeletal image analysis*. in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. 2019. IEEE.
14. Szegedy, C., et al. *Going deeper with convolutions*. in *Proceedings of the IEEE*

conference on computer vision and pattern recognition. 2015.

15. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks*. Advances in neural information processing systems, 2012. **25**.
16. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
17. Azcona, D., K. McGuinness, and A.F. Smeaton. *A comparative study of existing and new deep learning methods for detecting knee injuries using the MRNet dataset*. in *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*. 2020. IEEE.
18. Kumar, R. and R. Bhansali, 2021.
19. Li, Z., et al., *Deep Learning-Based Magnetic Resonance Imaging Image Features for Diagnosis of Anterior Cruciate Ligament Injury.* J Healthc Eng, 2021. **2021**: p. 4076175.
20. Sridhar, S., et al., *A Torn ACL Mapping in Knee MRI Images Using Deep Convolution Neural Network with Inception-v3.* J Healthc Eng, 2022. **2022**: p. 7872500.
21. Awan, M.J., et al., *Efficient Detection of Knee Anterior Cruciate Ligament from Magnetic Resonance Imaging Using Deep Learning Approach.* Diagnostics (Basel), 2021. **11**(1).
22. Dunnhofer, M., N. Martinel, and C. Micheloni. *Improving MRI-based knee disorder diagnosis with pyramidal feature details*. in *Medical Imaging with Deep Learning*. 2021. PMLR.
23. Kara, A.C. and F. Hardalaç, *Detection and Classification of Knee Injuries from MR Images Using the MRNet Dataset with Progressively Operating Deep Learning Methods.* Machine Learning and Knowledge Extraction, 2021. **3**(4): p. 1009-1029.
24. Jeon, Y., et al., *Interpretable and Lightweight 3-D Deep Learning Model for Automated ACL Diagnosis.* IEEE J Biomed Health Inform, 2021. **25**(7): p. 2388-2397.
25. Johnson, J.M. and T.M. Khoshgoftaar, *Survey on deep learning with class imbalance.* Journal of Big Data, 2019. **6**(1).