ISAKOS CONGRESS 2025 | MUNICH GERMANY June 8–11

The Reliability of Deep Learning Models in Assessing the Shoulder Arthroscopic Field's Visual Clarity in Relation to Bleeding

Son Quang Tran, MD, Vietnam (Presenter)
Thanathep Tanpowpong, Thun Itthipanichpong, Danaithep Limskul, Napatpong Thamrongskulsiri, Bao Nguyen Tu Thai, Minh Cong Bui

Faculty Disclosure Information

- The authors have no conflicts of interest to disclose.

Introduction

- In shoulder arthroscopy, surgical field clarity is a key measure of hemostatic control.
- Researchers have used various scoring scales to quantify this, including numeric rating, visual analog, 3-grade, and 5-grade scales.
- Despite predefined clarity, expert assessments often show inter-rater and intra-rater bias. This inconsistency may affect how the intervention's effectiveness is interpreted, influencing study conclusions.
- Computer vision is a rapidly advancing field with powerful applications, including image classification, object detection, and segmentation.
- Keras, an open-source API library by François Chollet, offers pre-trained models for transfer learning. This approach saves time and computing power compared to training models from scratch.

Research objectives

We aim to develop DL models for the classification of the visual clarity of arthroscopic shoulder images and then evaluate the reliability and agreement of these models with the assessment of raters.

- DenseNet169
- DenseNet201
- Xception
- InceptionResNetV2
- VGG16
- ViT

Materials and Methods

- A retrospective study, Aug 2020 – Aug 2024
- 119 patients undergoing shoulder arthroscopic surgery
- King Chulalongkorn Memorial Hospital.

- Rotator cuff repair
- Bankart repair
- Shoulder debridement arthroscopy

Inclusion criteria

- No recorded video
- Resolution < 1080p
- Significant technical issues
- Prior shoulder surgeries involving implants

Exclusion criteria

ISAKOS CONGRESS 2025 | MUNICH GERMANY June 8–11

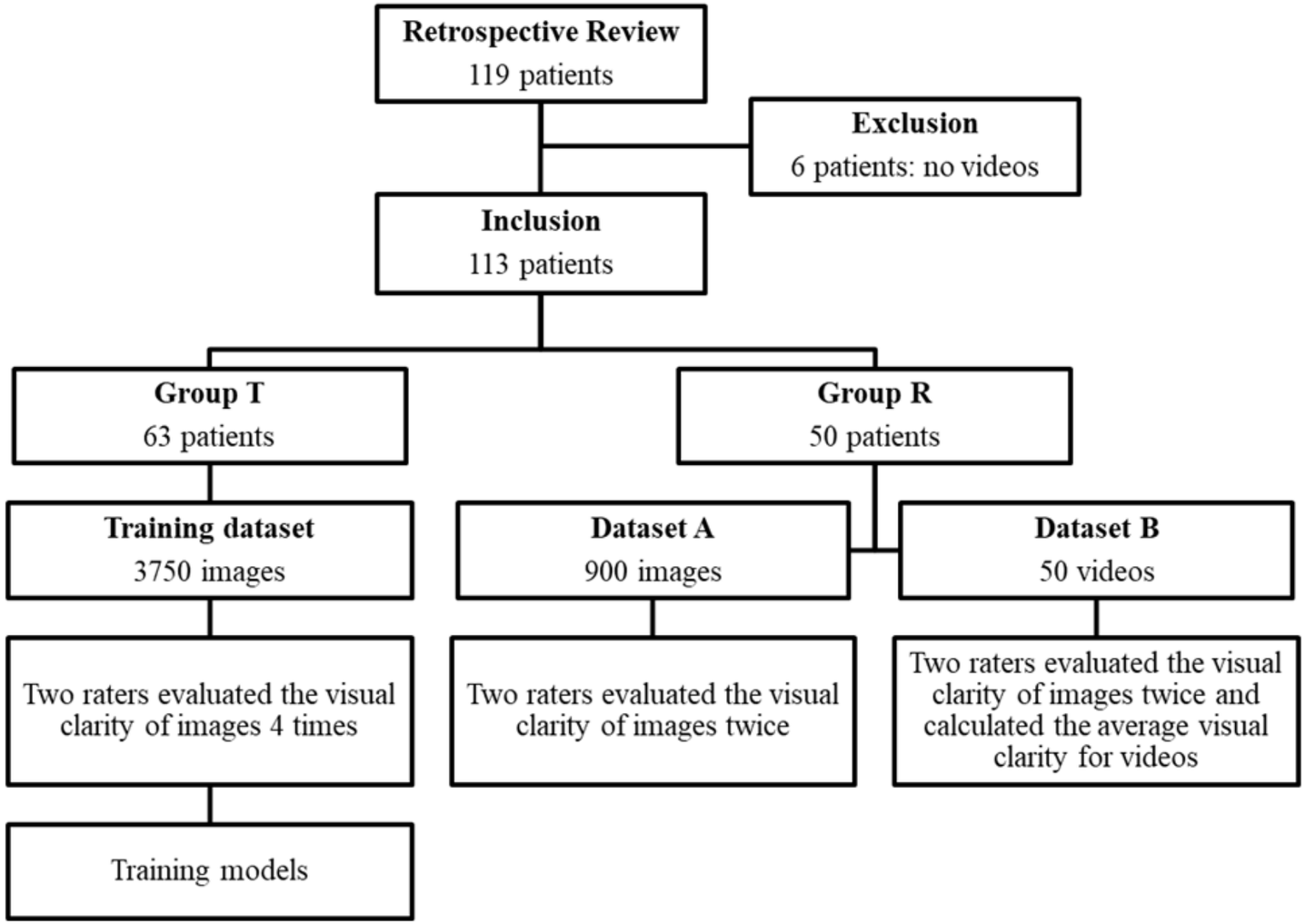


Figure 1. Flowchart of patient review and dataset preparation for training and assessing the deep learning models

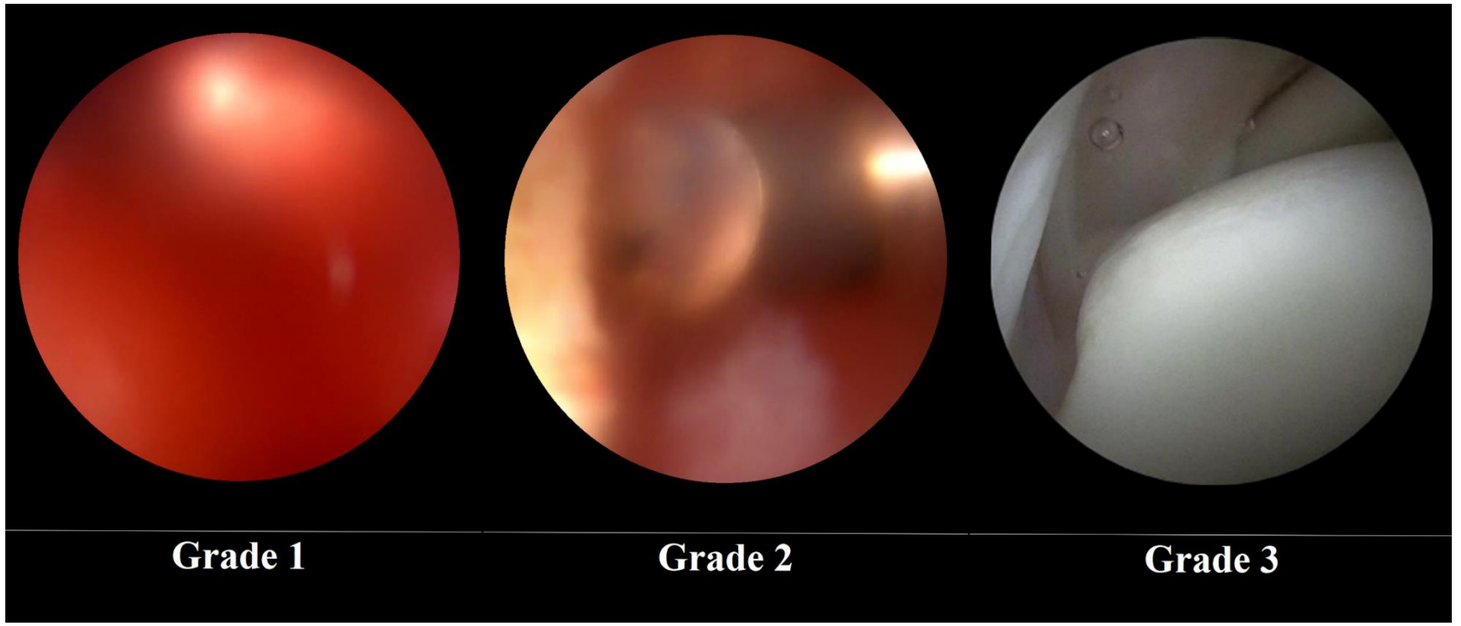


Figure 2. Three-grade scale for classifying the visual clarity of shoulder arthroscopic image relative to bleeding

- **Grade 1** indicated poor visual clarity, and the anatomical structure is seen as poor because bleeding disrupts the arthroscopic field, and the surgeon needs hemostasis.
- **Grade 2** was considered to moderate visual clarity, causing partial obstacles in surgical procedures due to the presence of diluted blood in irrigation fluid during arthroscopy.
- **Grade 3** corresponded to good visual clarity with slight or no bleeding, allowing a surgeon to observe structures in the arthroscopic field well.

Extract images from videos

Screen and exclude the error images

Crop images in square shape 1080x1080

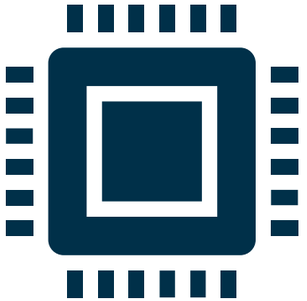
Preprocessed images

Import necessary libraries

Label for dataset and split 70:30

Load pre-trained models

Train models with 50 epochs



Dataset A
900 images

Dataset B
50 videos

Reliability evaluation

Weighted kappa and percent agreement

Intraclass correlation coefficient, bias and limits of agreement

ISAKOS CONGRESS 2025

MUNICH GERMANY

June 8–11

Results

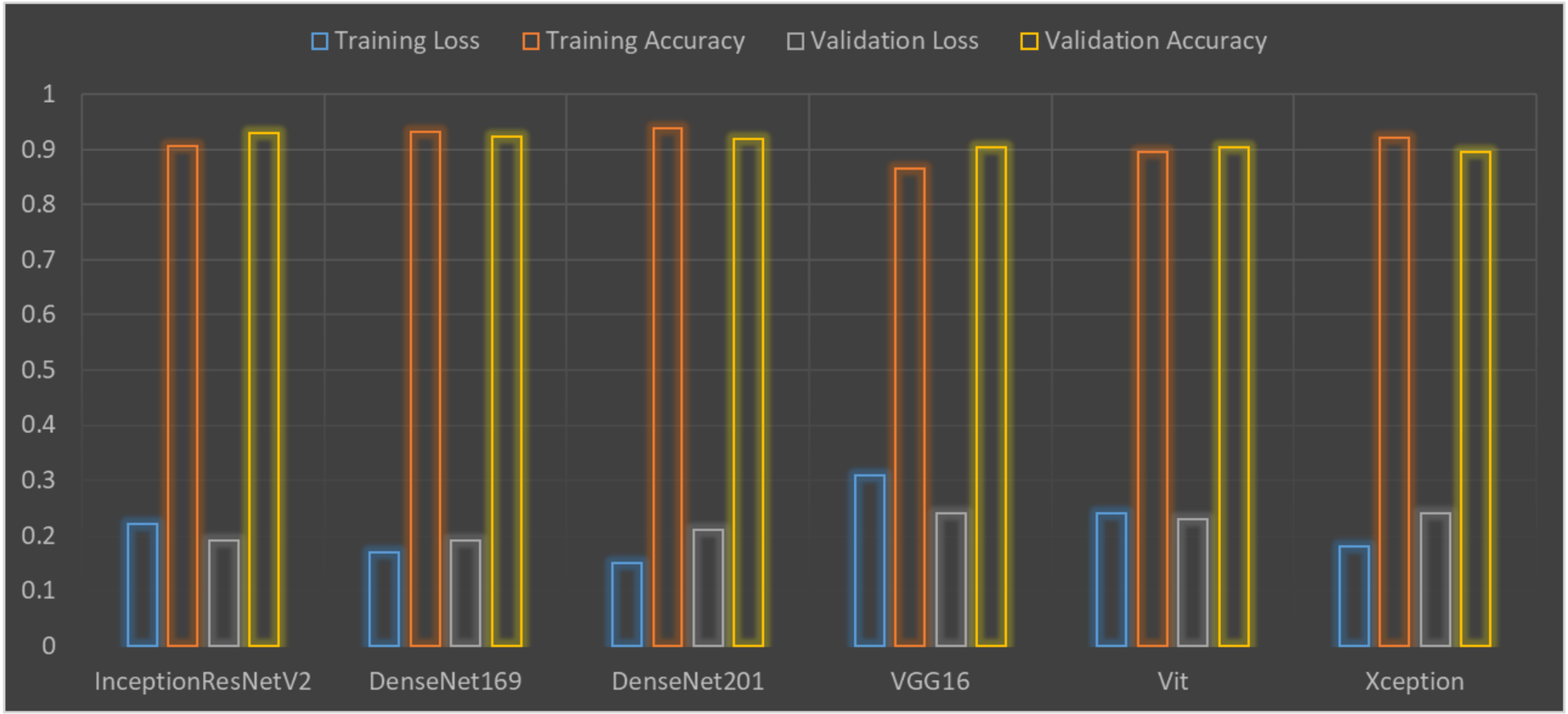


Figure 3. Loss and accuracy metrics of the most optimally trained predictive models

Results

Comparison to raters	Classification of visual clarity for SA images		Quantification of visual clarity for SA videos	
	Percent agreement	Kappa (95% CI)	ICC (95% CI)	Bias (LoA)
DenseNet169	79.2%	0.86 (0.84 – 0.88)	0.90 (0.73 – 0.96)	0.027 (-0.086 – 0.197)
InceptionResNetV2	82.3%	0.87 (0.85 – 0.89)	0.86 (0.74 – 0.92)	-0.017 (-0.352 – 0.095)
ViT	77.8%	0.82 (0.80 – 0.85)	0.84 (0.71 – 0.92)	-0.014 (-0.313 – 0.133)
Xception	73%	0.78 (0.75 – 0.80)	0.84 (0.44 – 0.93)	-0.051 (-0.291 – 0.070)
VGG16	70.7%	0.72 (0.68 – 0.75)	0.83 (0.57 – 0.92)	0.051 (-0.101 – 0.203)
DenseNet201	67%	0.67 (0.63 – 0.71)	0.81 (0.68 – 0.89)	0.026 (-0.294 – 0.216)

Conclusion

“The trained DenseNet169 models demonstrated the most optimal agreement and reliability compared to the evaluation of raters about the visual clarity of shoulder arthroscopic images and videos. These models can be utilized in further research as objective measurements”

Read the full text

ISAKOS CONGRESS 2025

MUNICH GERMANY

June 8–11