TISSUE ENGINEERING: Part A
Volume 00, Number 00, 2023

DOI: 10.1089/ten.tea.2023.0206

termis
Tissue Engineering
& Regenerative Medicine
International Society

Deep Learning Augmented Osteoarthritis Grading Standardization

Lacksaya Nagarajan, M.Tech,[1] Aadyant Khatri, B.Tech,[1] Arnav Sudan, B.Tech,[1]
Raju Vaishya, MS, MCh (L'pool), FRCS (London), FACS (USA),[2] and Sourabh Ghosh, PhD[1]

Manual grading of cartilage histology images for investigating the extent and severity of osteoarthritis (OA) involves critical examination of the cell characteristics, which makes this task tiresome, tedious, and error prone. This results in wide interobserver variation, causing ambiguities in OA grade prediction. Such drawbacks of manual assessment can be overcome by implementing artificial intelligence–based automated image classification techniques such as deep learning (DL). Hence, we present the feasibility of training a deep neural network with cartilage histology images, which can grade the extent and severity of knee OA based on modified Mankin scoring system. The grading system used here for automating OA grading was simplified and modified based on the microscopic observations from the histology images, where three parameters (Safranin-O staining intensity, chondrocyte distribution and arrangement, and morphology) were considered for evaluating the OA progression. The histology images were tiled, labeled, and grouped together based on the developed grading system (Grade 0–3). Four different DL architectures were tried for image classification and the best performing model was selected by fivefold validation method. With a validation accuracy of $\sim$84%, 0.632 Cohen's kappa score, and an excellent receiver operating characteristic (ROC)–area under the ROC curve ranging between 0.89 and 0.99, DenseNet121 was selected among the four models as the best performing model, and was used for inferencing on new data. Final grades obtained from the models were in accordance with the grades provided by the medical experts. We hereby demonstrate that a DL architecture can be taught to interpret the degree of cartilage degradation, with excellent discriminatory ability across all four classes of OA severity. Unlike other works where radiographic images have been considered for grading of OA, we have considered histology images, which is a fundamental approach for grading OA extent and severity. This would bring a paradigm shift in histology-based assessment of OA, making this automated approach to be explored as an option for OA grading standardization. Ethical approval number-IAH-BMR-018/10-19.

Keywords: osteoarthritis, cartilage histopathology, OA grading, OA severity, deep learning, image classification, grading standardization

Impact Statement

1. Feasibility of training a deep neural network with human knee osteoarthritic histology images to grade the extent and severity of osteoarthritis (OA) has not been reported yet anywhere to the best of our knowledge.

[1]Department of Textile and Fibre Engineering, Indian Institute of Technology Delhi, New Delhi, India.
[2]Department of Orthopaedics, Indraprastha Apollo Hospital, New Delhi, India.

2. Among the four deep learning models tried ResNet50, DenseNet121, VGG16, and InceptionV3, DenseNet121 was found to be the best performing model with a validation accuracy of ~84%, and the grades presented were in accordance with the grades given by medical experts (pathologists and orthopedic surgeons).
3. This work would be of great interest to the researchers who work on distinguishing the subsets of OA, defining the end points for clinical trials, evaluation of characteristic features and disease markers of OA, and evaluation of animal OA models.
4. Such automated grading systems help in transforming the tiresome and error-prone grading process to swift, precise, and unbiased grading, thereby reducing the demand for pathologists with years of experience

Introduction

Osteoarthritis (OA), brought on by mechanical wear and tear of the articular cartilage is a common degenerative joint disease with a global incidence of ~203 per 10,000 person-years.[1] Being an age-dependent and slowly progressive disease, it results in decreased mobility owing to joint pain causing chronic disability.[2] Cartilage has limited self-repairing capacity.[3] Even when the cartilage starts to regenerate, further hypertrophic differentiation offers a major problem to develop phenotypically stable cartilage.[4–6] Several research groups have investigated the conditions of chondrogenesis and hypertrophy for developing an understanding about the cartilage microenvironment,[7–9] but identification of hypertrophy in OA cartilage biopsies is still debatable. Studies also reveal that the correlation between symptoms and observable radiographic changes at different stages of the disease has not been properly understood yet.[10] Thus with a high global prevalence, OA necessitates the investigation of its contingency, risk factors, and preventive strategies much in great detail.

One of the major challenges associated with OA is that the disease cannot be diagnosed at early stages, because it cannot be apparently seen clinically. Only at the progressed states, it can be diagnosed with pronounced alterations in the joint that can lead to chronic pain, functional restrictions in the affected joints, and some other radiographically detectable changes. Assessment of OA involves physical examination, radiographs, and MRI (including few other new T2 and T1 contrast-enhanced image techniques like dGEMRIC),[11,12] which provides only a semiquantitative diagnosis of the extent and severity of the disease. Patients with early stages of OA might not necessarily manifest any evident radiographic changes, whereas patients with observable radiographic changes might not exhibit any severe symptoms. This makes these methods unbefitting for determining the degree of cartilage destruction. The most common way of extracting information about the existing state, features, and characteristics of cartilage is by histopathology.[13]

Histological image analysis offers the advantage of providing detailed information on the entire tissue section, where the cell characteristics like cell morphology, cell density, distribution, and so on, can be investigated in detail. Although histological images provide enough information for detecting the extent of the disease progression, the success of detection and diagnosis relies entirely on the pathologist who grades those images. The primary issue with such a method of manual OA grading is that there is no universal or standardized method available yet for the pathologists to follow, which is why there is a huge variation in the grades given for a single case by different pathologists around the world, although the grading system used for diagnosis is the same.

As far as manual grading of OA is concerned, it is an error-prone and cumbersome process. It always has the drawbacks of having bias, poor interuser agreement, poor reproducibility, and reliability, despite demanding highly experienced and qualified pathologists.[14] These common manual histopathologic assessment methods fail to detect the early or mild phases of the disease, as it is highly subjective. Hence in this study, an attempt has been made to automate the grading of OA by developing a deep learning (DL) model that would outperform the manual grading system in terms of faster readouts, scoring bias, reliability, and reproducibility.[15] This model is also expected to replace the manual grading system, thereby eliminating the vast interuser variation in determining the grade of cartilage degradation.

Being a subset of artificial intelligence (AI), DL has proved to be a much better and popular tool for biomedical image analysis.[16–18] It can find patterns in data or fit predictive models to it, and is notably helpful for automating the analysis of data that is too massive or complicated for human analysis in a time-efficient and repeatable manner.[19,20] To surmount these mentioned drawbacks of manual grading, a DL-assisted automated grading model has been developed, that grades the human cartilage specimens based on the severity and extent of the disease progression. This would potentially reduce the scoring bias, and improve the reliability, reproducibility and interuser agreement regarding the grades presented. To the best of our knowledge, this is the first work that reports the automation of OA grading of human cartilage samples, which would thereupon act as a gold standard for histology-based OA grading.

Materials and Methods

Workflow

DL-based supervised image classification for OA grading requires data collection, sorting, labeling, training, and validation. Figure 1 provides the workflow for automating the grading of human OA samples by a DL model. The image dataset was obtained from human healthy cartilage (HC) and osteoarthritic cartilage (OAC) samples collected from Indraprastha Apollo hospital (New Delhi), following proper and essential approval. All these samples were stained with Safranin-O and hematoxylin and eosin (H&E) and these stained samples were then taken for imaging. Detailed information on sample collection and ethical approval, and histological analysis has been provided in Sections 1 and 2 in Supplementary Data, respectively. For improving the heterogeneity of the data set, images of the stained slides were captured by two different microscopes (as mentioned in Supplementary Section 2).

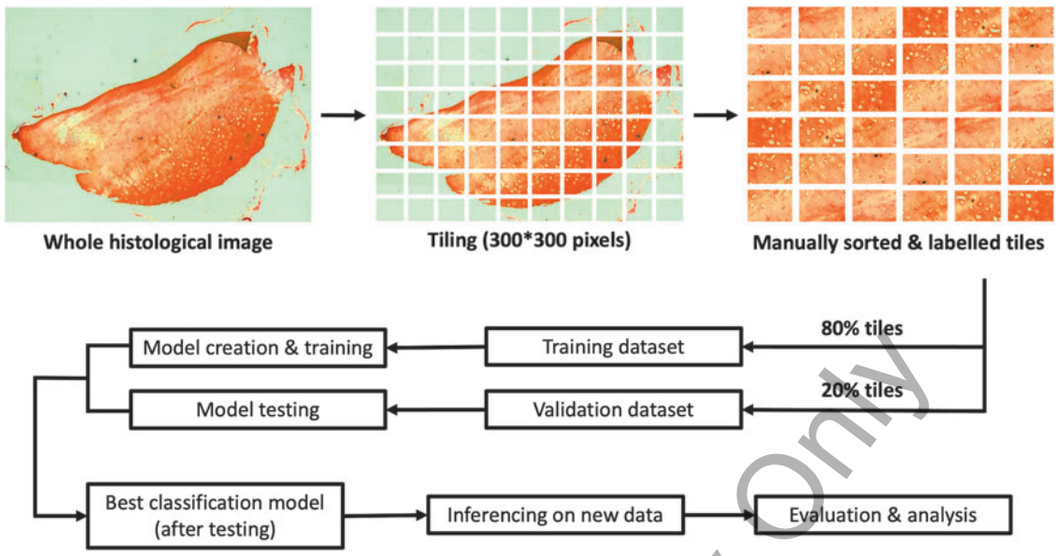


FIG. 1. Flowchart depicting how the training and testing of a supervised deep learning model works for our dataset. Tiling of the whole histological image results in a 300×300 px tile, which undergoes sorting and grading based on the different parameters selected. The tiles are graded by an experienced user (user 1) and 80% of those tiles are fed as the training dataset to the image classification models with the user 1 grade labels. The rest 20% of tiles are graded by both user 1 and user 2, and this forms the test dataset. After validation, the best performing model is selected based on validation accuracy, ROC-AUC score, and Cohen's kappa score. Finally, the model decision and its performance are visualized by generating class activation maps. AOC, area under the ROC curve; ROC, receiver operating characteristic.

All the TIFF images were considered as different layers of a single image for stitching and blending them together by Adobe photoshop (2022, Version 23.5.1) under photomerge option and further cropped into 300×300 pixel dimensioned images called tiles. These tiles then undergo sorting, where they are manually selected or discarded based on the information that a tile possesses (which includes the presence of chondrocytes, any fissures, background or artifacts, etc.). If the entire tile is covered with the tissue section, it is selected. Similarly, any tile displaying only the slide background or incomplete image, or any kind of histological artifact including folded, damaged, and torn tissues with dark spots in between was considered faulty and thus, was discarded.

The dimension of the tiles (300×300 pixels) was determined based on the locality of OA pathology. Chondrocyte size was a crucial factor in determining the tile dimension. As can be seen from Supplementary Figure S1, four different types of chondrocytes were observed in the Safranin-O-stained cartilage tissue sections. They were categorized based on their size. The chondrocytes size ranged from 12 to 87 μm when they were checked for its size using ImageJ. Given the case, 300×300 pixels was found to be most suitable tile dimension where different stages of chondrocyte hypertrophy could be visualized. The selected tiles were graded by an experienced user (by the grading methodology mentioned in Grading Methodology section).

Once labeled, these tiles were used for training the model. A ratio of 80:20 was used for splitting the images for training: testing, while ensuring that there is a reasonable distribution of images from each group for the model to predict decent enough. Various DL models were tried, and the best performing model with an acceptable or significantly high accuracy was selected, which would be further used for inferencing on new data for testing.

Grading methodology

Three simple scoring parameters were considered for grading here, which includes—staining intensity, cell distribution and arrangement, and cell morphology. In this modified and simplified scoring system (Readers are directed to refer the Section 3 in Supplementary Data for understanding the need and considerations for developing a modified scoring system), each parameter was considered for a scale of 0–3 (Table 1), and the sum of all these three parameters would result in a scoring system that ranges between 0 and 9. Each tile would receive a score between 0

TABLE 1. SCORING CATEGORIES BASED ON STAINING INTENSITY, CELL DISTRIBUTION AND ARRANGEMENT, AND CELL MORPHOLOGY

Scoring parameters	*Score*	*Definition*
Staining intensity	0	Dark staining
	1	Decent even staining
	2	Uneven and weak traces of staining
	3	Almost unstained
Cell distribution and arrangement	0	Cells spaced individually
	1	Two cells starting to cluster
	2	Three or more cells clustering together
	3	Multicellular chondrocyte clusters
Cell morphology	0	Healthy flat/round cells
	1	Mixed spindle and round cells
	2	Condensed, necrotic/pyknotic cells
	3	Hypertrophic cells

Table 2. Grade and Score Category Relation for Assessing the Severity of Osteoarthritis

Grade	*Score category*	*Description*
0	0	Satisfactory joint function
1	1–3	Mild to moderate OA
2	4–6	Moderate to severe OA
3	7–9	Severe OA

OA, osteoarthritis.

and 9, which can be further grouped into a subcategory to indicate the grade. This 0–9 scale score can be divided into four grades, indicating the severity of the disease (Table 2).

It is to be noted that higher the grade, higher is the severity of OA progression (i.e., the biomechanical properties of the cartilage deteriorate with increasing grades). ''0'' being the lowest grade indicates that the cartilage is healthy, and ''4'' being the highest grade indicates that the cartilage is highly osteoarthritic. Figures 2 and 3 show the representative images for each of the individual parameters selected for grading, and each score under our four-class grading system, respectively.

Dataset preparation

The 300 × 300 pixel tiles were split into 80% and 20% as training and validation datasets, respectively. The 80:20 split for training and testing was carried out keeping into account the images coming from the same patient. It was made sure that tiles from patients in the training category were not present in the testing category as this would have led to data leakage. Care was taken while grouping the tiles for training and testing from individual patients to prevent such target leakage, and in turn the overfitting. Detailed information on the sample dataset taken for model training and testing has been given in Supplementary Table S2. The number of tiles taken into consideration (after data augmentation) from each of the four classes as per the labels given by user 1 has been outlined.

Since the number of images obtained from microscopy was still less for a DL convolutional neural network to train on, data augmentation was carried out to considerably increase the training dataset size. Various techniques were used for data augmentation including flipping the images, rotating them by 90° and 180°, and blurring them using Gaussian spread. This helped in increasing the variability of the training dataset size by 400%. This expansion of the dataset acts as a regularizer and helped to reduce the overfitting by making the set of possible datapoints more comprehensive.[21]

DL model development

State-of-the-art convolutional neural networks were used as the base networks for transfer learning. These networks were pretrained to classify images into thousand object categories. The fully connected classification layers of each pretrained model selected were reshaped to predict the four different classes of cartilage degradation (Grade 0–3) in our dataset. More specifically, ResNet50,[22] DenseNet121,[23] VGG16,[24] and InceptionV3[25] were the four different models used.

Python version—Python 3.7.14 and the DL framework, TensorFlow 2.8.2, were used to train and test the models. Other Python libraries used were os, time, Matplotlib, the Python Imaging Library (PIL), NumPy, and Scikit-learn. Free version of Google colab having Intel(R) Xeon(R) CPU @ 2.20 GHz with system memory of 12 GB and Tesla T4 GPU of 16 GB memory was used.

Model training

Using Sci-kit learn package's GridSearchCV module, multiple models with different hyper-parameters, such as optimizer, learning rate, rho value, beta_1, and beta_2, were trained. The optimum value for each of these was found as follows.

1. Root Mean Squared Propagation (RMSProp) with a learning rate of 0.001 and a rho value of 0.9.
2. Adam optimizer with learning rate = 0.001, beta_1 = 0.9, and beta_2 = 0.999.

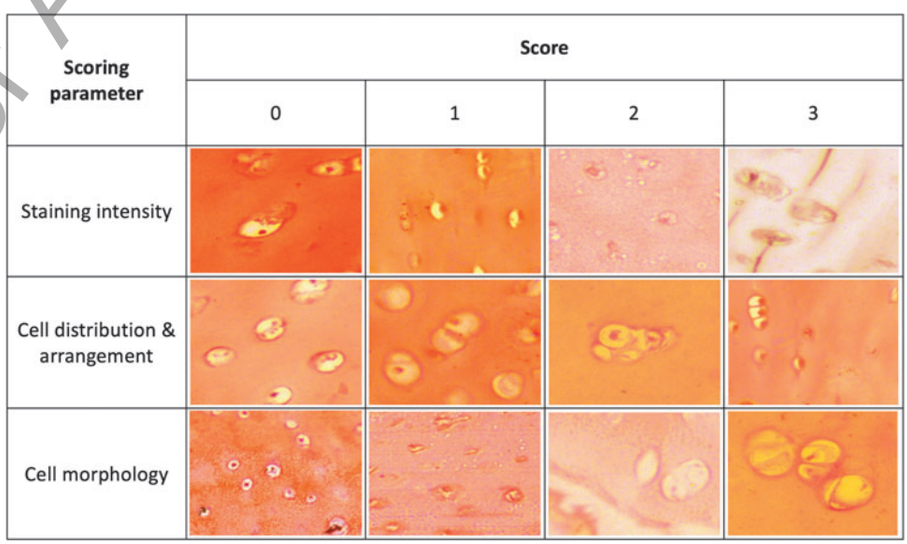


FIG. 2. Representative images for each of the individual parameters selected for grading. (Refer Table 2 for description regarding the histological scores given for all the three individual parameters on a scale of 0–3. Note that a score of 0 indicates healthy condition and with increasing scores 1, 2, and 3, the condition worsens).

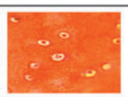
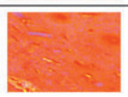
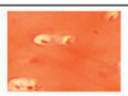
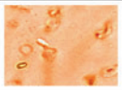
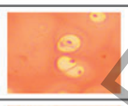
Grade	Score	Individual score for the parameters	Representative tile
0	0	0+0+0	
1	1	0+0+1	
	2	1+1+0	
	3	1+2+0	
2	4	2+0+2	
	5	3+0+2	
	6	2+2+2	
3	7	1+3+3	
	8	3+2+3	
	9	3+3+3	

FIG. 3. Representative images for each score under our four-class grading system. The tiles were graded based on the three parameter-staining intensity (0–3), cell distribution and assembly (0–3), and cell morphology (0–3). Sum of the scores from all these three parameters gives the overall score on a scale of 0–9. The tiles here are grouped based on the final grades (0–3). For instance, 1 + 2 + 0 indicates that the tile has received a score of 1 for staining intensity, 2 for cell distribution and arrangement, and 0 cell morphology. The overall score therefore becomes 3, and it gets covered under grade 1 (Mild to moderate OA). OA, osteoarthritis.

In conjunction with these optimizers, the Categorical Cross-Entropy loss function was used because there are multiple categories for the model to predict. The training was carried out with different batch size values, the values ϵ {32, 64, 128} were tried in particular. Among the 50 epochs, the hyper-parameters at the particular epoch at which the model achieved the best validation accuracy was saved and used.

Statistical analysis

Various statistical measures were used to determine the performances of different models created. Fivefold validation[26] was used to determine the best performing model, with all the statistical analyses being performed in python. For assessing the four-class classification of the final grades, that is, Class 0, 1–3, 4–6, and 7–9 (Grade 0, 1, 2, and 3, respectively), a linear-weighted kappa statistic was used. Cohen's kappa with linear weights has been used to assess the performance of the image classification model, by evaluating the agreement between two users. Scikit-learn library in python[27] has been used for all the statistics and calculations, including kappa statistics,[28] confusion matrices,[29] receiver operating characteristic (ROC),[30] precision–recall curves,[31] and the area under those curves (AUC).

Model decision visualization

Gradient-weighted Class Activation Mapping (Grad-CAM) has been used to visualize the predictions made by our models. A single image can have multiple possible concepts that can pose difficulties in prediction and in that case, Grad-CAM can give better quality visual explanations without any demand for retraining or architectural changes.[32] Weights from the final convolution layer of our DL models were used to create a rough map of the regions that the model detected as important for decision making. This map was then superimposed on the original input image for easier understanding of the important regions that the model considers before classification. Code adapted from keras DL Application Programming Interface written in python, 2017 (https://keras.io/examples/vision/grad_cam/) has been used here for generating the class activation heat map for the image classification model.

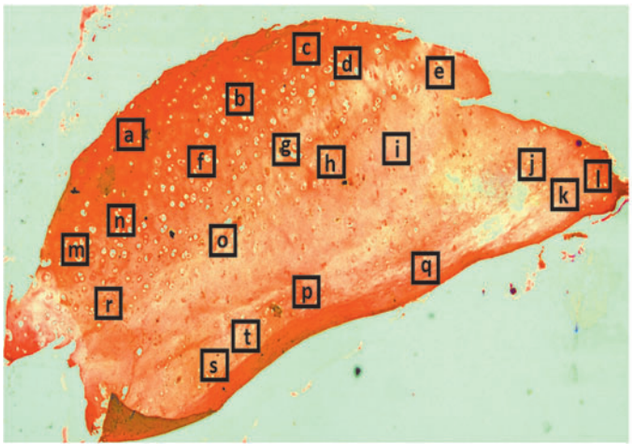
Calculation of overall score and the final grade

Once the model has been trained based on the labels given, it will present the score for the new test data on testing. We get the overall score by taking average of the scores given by the model for all the tiles taken from a single patient sample. For obtaining the grade from the overall score to a reasonably accurate value, the decimal valued overall score is approximated to a whole numbered grade. Figure 4 provides an overall idea of how several sample tiles from a single patient were graded by the model, averaged out, and approximated to get the overall grade of OA severity. It is to be noted here that the 80:20 splitting ratio for training:testing was carried out at the level of patients. More specifically, 80% of the patients were kept in the training category, whereas 20% of them were kept in the testing category. Now that the grades are given by the model, it needs to be checked for its reliability and reproducibility for proving that these DL models can be considered as a potential alternative to the manual grading systems.

Results

DL model comparison

Four different models, namely ResNet50, DenseNet121, VGG16, and InceptionV3 were trained and tested for their performances. It is to be noted that with the various models trained and tested, all the layers of pretrained models were frozen in all the experiments. Table 3 details the various metrics such as the accuracy, ROC-AUC Score, and Cohen-kappa score achieved by all the four models ResNet50,



Tile	Grade	Tile	Grade
a	2	k	1
b	3	l	0
c	2	m	3
d	3	n	3
e	2	o	2
f	3	p	1
g	3	q	2
h	2	r	2
i	0	s	1
j	1	t	1
Overall grade (model prediction)		2 (Moderate to severe OA)	
Grade given by pathologist		2 (Moderate to severe OA)	

FIG. 4. An example showing how individual tiles from a whole histological image are graded to present the final grade of OA severity. In this case, 20 tiles from different parts of the specimen image obtained from a single patient is taken into consideration. Grades of the individual tiles have been averaged out to obtain the overall grade, which in turn indicates the extent of OA severity.

DenseNet121, VGG16, and InceptionV3. The models were trained for 50 epochs, and saved once the training was complete for future inferencing.

Cochran's Q test[33] has been used to compare the multiple classifiers quantitatively. It is a statistical method that aims to determine whether there are significant differences in the classification accuracies of the four different models. The test is conducted with a predetermined significance level, denoted as ''α,'' which is set to 0.1 in this case. This means that we are willing to accept a 10% chance of making a Type I error (incorrectly rejecting the null hypothesis when it's true). After conducting Cochran's Q test, the obtained p-value is 0.099. The p-value represents the probability of observing the results (differences in classification accuracies) by chance alone. When the p-value is less than the chosen significance level ''α,'' it suggests that the results are unlikely to have occurred randomly.

In our case, the fact that the p-value (0.099) is slightly smaller than the significance level (0.1) indicates that there is a <10% chance of obtaining the observed differences in classification accuracies by random chance. Consequently, we reject the null hypothesis. Rejecting the null hypothesis implies that there is a statistically significant difference between the classification accuracies of the four models we tested. In other words, we have evidence to conclude that these models are not equally effective in their classification tasks, and there are meaningful distinctions in their performance.

TABLE 3. METRICS OF VARIOUS MODELS USED FOR GRADING

Model	*Best validation accuracy, %*	*Epochs to attain the best validation accuracy*	*Training time*	*Cohen-kappa score*
DenseNet121	84	22	09 min 55 s	0.623
ResNet50	81	45	12 min 00 s	0.668
VGG16	76	33	15 min 20 s	0.586
InceptionV3	67	34	09 min 20 s	0.430

Pretrained DenseNet121 model with ImageNet weights and our custom classification layers yielded a validation accuracy of 84% (which is the highest among all the models tried) with a validation loss of 0.5615. The model also achieved a Cohen kappa score of 0.623 with linear weights. It can be considered the best performing model since the number of epochs it took to achieve the best validation accuracy is only 22, with a training time of 9 min 55 s. Hence, DenseNet121 was chosen for analyzing and inferencing on new data. ResNet50, on the contrary, also achieved a validation accuracy of 81% with a Cohen-kappa score of 0.668, which shows that both DenseNet121 and ResNet50 can be regarded as competent among all the four models tried for achieving our objective. Figure 5 presents the training and testing logs (Accuracy and loss [vs.] number of epochs plot) for the four models tried.

Cohen's kappa statistics

Cohen's kappa gives the percent agreement between two raters for a two-user system. It can be seen from Table 3 that our Cohen's kappa result ranges from 0.430 to 0.668 for all four models tried. Among the different models tried, the highest kappa score obtained was from ResNet50, which is ~0.668. The corresponding kappa score for our best performing model DenseNet121 is 0.623, which suggests that there is substantial agreement (Refer Supplementary Section 6.1) between the grades given by user 1 and the model. So, this gives an idea that our best-fit model Densenet121 is performing decent enough to predict the labels to an acceptable level given by user 1. Figure 6A provides the linear weighted kappa statistic for our best performing model DenseNet121, to show the agreement between user 1, user 2, and the model.

Normalized confusion matrices

For defining the performance of a classification algorithm, confusion matrices are used. Confusion matrices are tables that visualize and summarize the performance of a classification algorithm by quantifying the classification

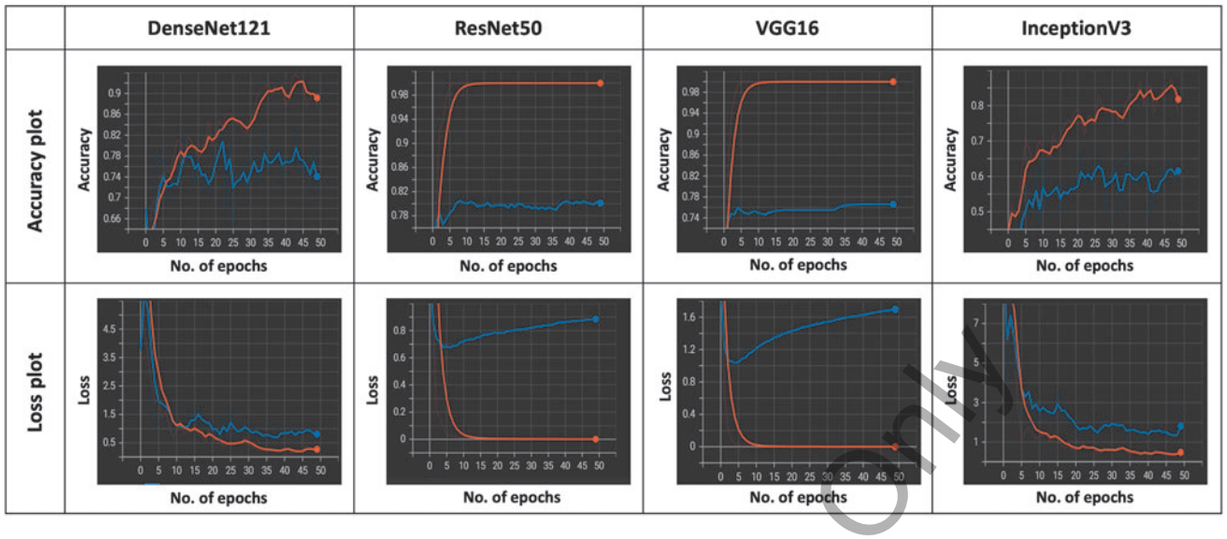


FIG. 5. Accuracy and loss plots (for training and testing) obtained for all the four models DenseNet121, ResNet50, VGG16, and InceptionV3. It is to be noted here that the orange curves represent the training logs and the blue curves represent the testing logs. All the models are trained for 50 epochs and the number of epochs to achieve the best validation accuracy with training time has already been reported in Table 3.

overlap.[29] The normalized confusion matrix was generated using predictions made by the corresponding models (Fig. 6). Figure 6B and C provides the matrices constructed between the two users, and for each user versus the model predictions, respectively. An ideal model would have non-zero values only on the diagonal and it can be seen from the matrix that our top two best performing models ResNet50 and DenseNet121 perform well to a great extent.

ROC curves

The ROC being a probability curve basically indicates the performance for classification issues at different threshold levels. Multiclass ROC curves[30] plotted between the users and the four different models (Fig. 7). AUC provided in Table 4 was considered for analyzing the discrimination capacity and the accuracy of the diagnosis.

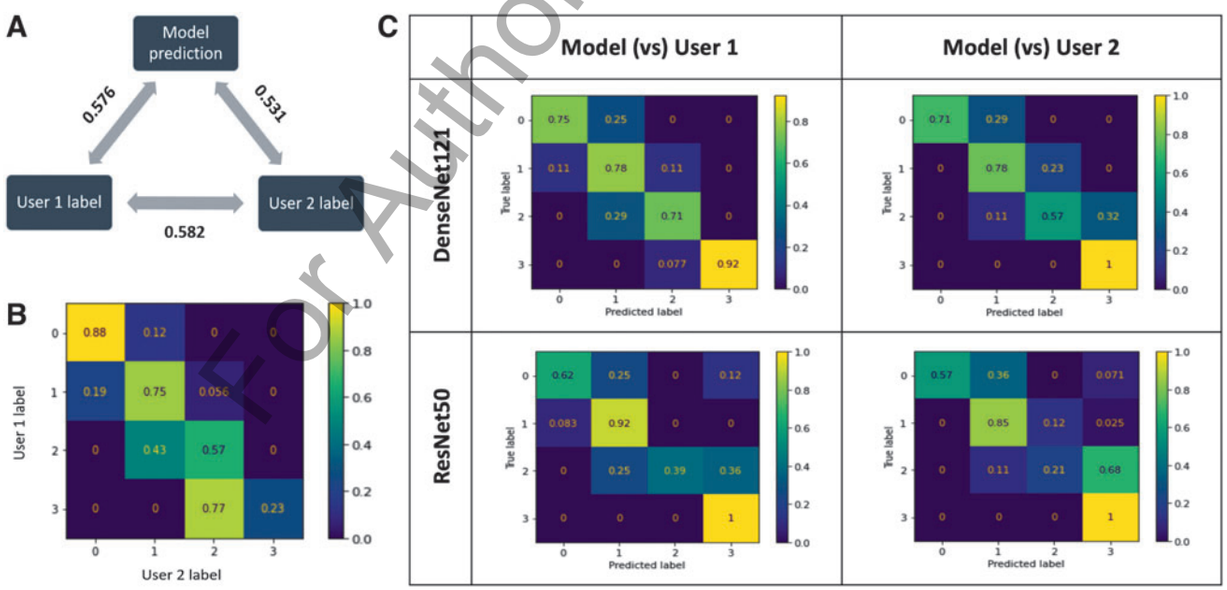


FIG. 6. Analysis of interuser agreement and reliability of the model prediction by linear weighted kappa statistic and confusion matrices. **(A)** Linear-weighted kappa statistic for the best performing model DenseNet121 indicating the agreement between user 1, user 2, and the model. **(B)** Normalized confusion matrix indicating the label predictions by user 1 and user 2. **(C)** Normalized confusion matrix between the model and both the users for the top two best performing models DenseNet121 and ResNet50, respectively (True label indicates the label given by the user, whereas the predicted label is the label given by the model).

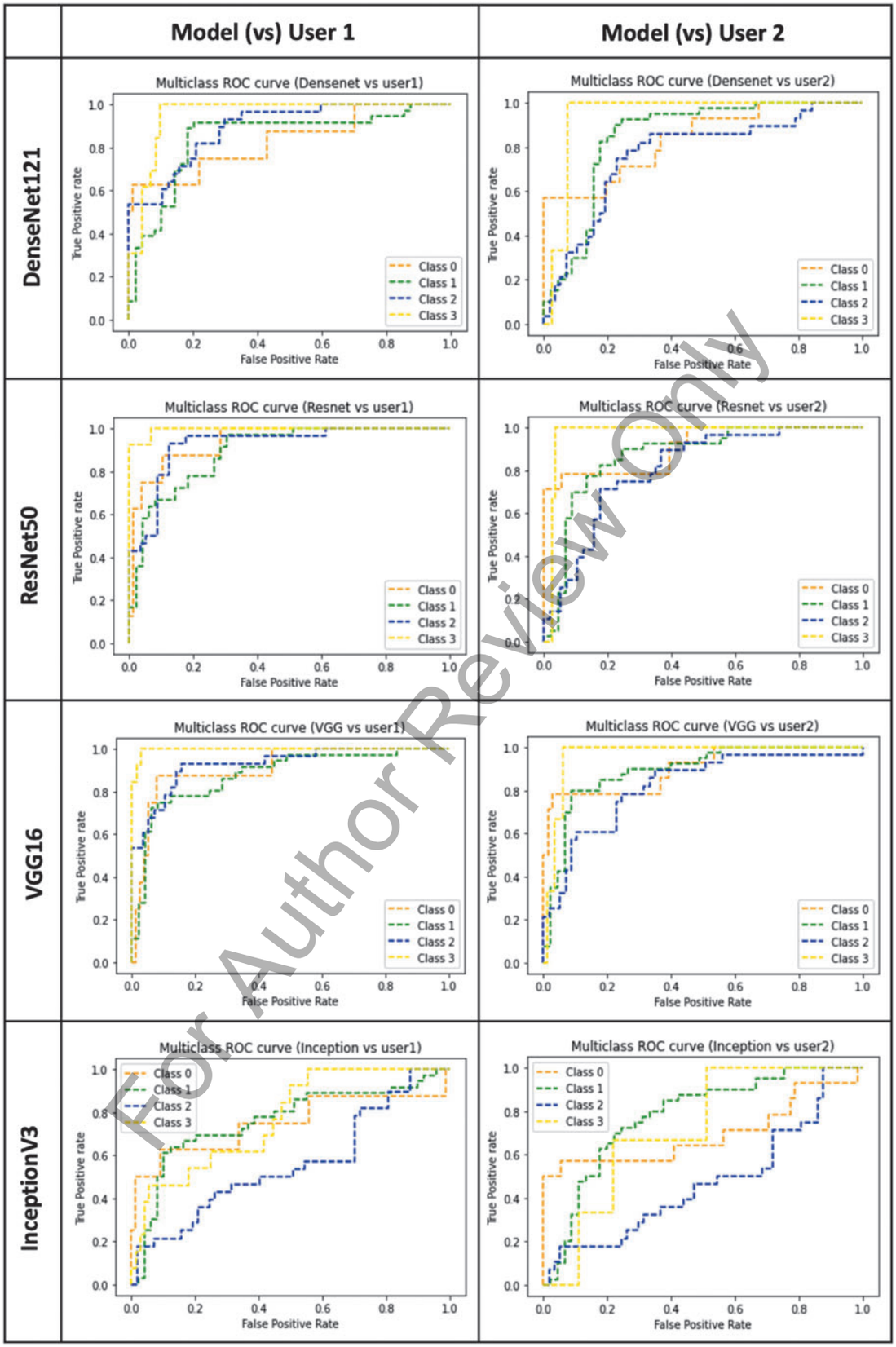


FIG. 7. True-positive and false-positive rates plotted between the users and the model (i.e., model [vs.] user 1, and model [vs.] user 2) for all four different grades of OA severity, represented by multiclass ROC curves obtained from the four different models DenseNet121, ResNet50, VGG16 and InceptionV3.

Table 4. Receiver Operating Characteristic–Area Under the Receiver Operating Characteristic Curve Values for the Top Two Best Performing Models with 95% Confidence Interval

	DenseNet121		*ResNet50*	
	Model vs. user 1	*Model vs. user 2*	*Model vs. user 1*	*Model vs. user 2*
Grade 0	0.94 (0.81–1.00)	0.91 (0.75–1.00)	0.83 (0.60–1.00)	0.84 (0.61–1.00)
Grade 1	0.89 (0.81–0.97)	0.87 (0.77–0.96)	0.85 (0.73–0.96)	0.85 (0.73–0.95)
Grade 2	0.93 (0.84–1.00)	0.81 (0.67–0.93)	0.89 (0.79–0.98)	0.77 (0.64–0.89)
Grade 3	0.99 (0.96–1.00)	0.97 (0.90–1.00)	0.95 (0.84–1.00)	0.94 (0.82–1.00)

Our AUC data show that DenseNet121 has an outstanding and excellent discriminatory ability with reference to the labels given by user 1 and user 2, respectively (refer Supplementary Section 6.2). These can be regarded as good scores, which evinces that our model can substantially differentiate across the four different grades of cartilage degradation in our OA samples.

Precision–recall curves

The tradeoff between precision and recall for various thresholds is depicted by the precision–recall curve.[31] High precision is correlated with a low false-positive rate, whereas big recall is correlated with a low false-negative rate. A high AUC denotes both high recall and high precision, indicating a better model. For all four models ResNet50, DenseNet121, VGG16, and InceptionV3, multiclass precision–recall curves have been plotted between both the users and the model (Fig. 8).

The AUC for precision–recall curves is given in Table 5 for the top two best performing models. From the curves, it can be seen that ResNet50 and DenseNet121 have performed well in terms of getting low false-positive and false-negative rates for almost all the four classes, and the model (vs.) user 1 curves appear to be having better precision and recall than the model (vs.) user 2 curves.

Grad-CAM visualization

Grad-CAM visualization was investigated to know the critical regions for grade prediction by the model. Considering DenseNet121 as the best performing model, Grad-CAM visualization was run using keras. The grade labels given by user 1, user 2, and the model, along with the class activated map that the model considers for grade prediction (Fig. 9). It is important to note the parts of every image that was considered important for decision making by the model. Because the final grade is averaged out from different tiles of a single patient, the grades can be regarded reliable. This class activated map is also useful for investigating the disagreement between the users and the model as a result of incorrect model prediction. Different cases of disagreement from all three ends (user 1, user 2 and model) have also been presented for the reader's reference.

Discussion

In this study, an attempt has been made to investigate the feasibility of automating the grading of human OA cartilage histology sections using DL. Although DL has already been used for classifying, analyzing, and grading several biomedical images including that of the tumor,[34,35] chondrogenesis of a tissue engineered cartilage,[16] and OA with microcomputed tomography[36] and other radiographic techniques,[37–42] automated grading of human OA histological images has not been reported yet. A DL model that grades the severity and extent of OA, based on three simple parameters (staining intensity, cell distribution and arrangement, and cell morphology) has been reported for the first time. Transfer learning[43] technique was used on several pretrained state-of-the-art convolutional neural networks along with new trainable fully connected linear classification layers to achieve our objective.

Several studies have demonstrated the implementation of machine learning in diagnosis and prognosis of OA. Previous machine learning–oriented research works on X-ray and MRI-based localization (detection and segmentation) of knee joints, classification and prediction of OA severity, and progression have considered several radiological OA features such as joint space narrowing (JSN), formation of osteophyte, cyst, or geode, coronal tibiofemoral subluxation, and subchondral sclerosis.[44–47] Although all these works on OA classification focus on extracting meaningful, relevant, and informative features from the radiographs, chances of misinterpretation in its classification with knee X-ray data are pretty high, as geometric distortions can be found in the cartilage region.[48–51] Majority of these studies have developed algorithms to read the images based on the Kellgren and Lawrence grading system that examines the formation of osteophytes, JSN, subchondral sclerosis, and bone end deformity.[52]

Nevertheless, drawbacks such as reliability issues in detecting lateral compartment OA changes, inter- and intraobserver reliability issues, and overemphasizing of osteophytes significance than JSN make it an objectionable choice for grading the severity and progression of OA.[53] This is where our work stands novel where histology images of human cartilage samples were considered, which is an absolute fundamental technique and gold standard practice to characterize human knee OA. This comprehensive automated technique of OA grading would be significantly beneficial as an assistive research tool in distinguishing and evaluating the characteristic features of OA and its subsets. The grades presented by our best performing model DenseNet121 have been validated by first and second medical experts. The collective grades obtained from the histological tiles were found to be in accordance with the opinion provided by pathologists and orthopedic surgeons.

Heterogeneity within the full sample volume of a human OA sample makes it difficult to choose the complete parameters for grading using Osteoarthritis Research Society International or Mankin scoring systems. Hence, the parameters for grading

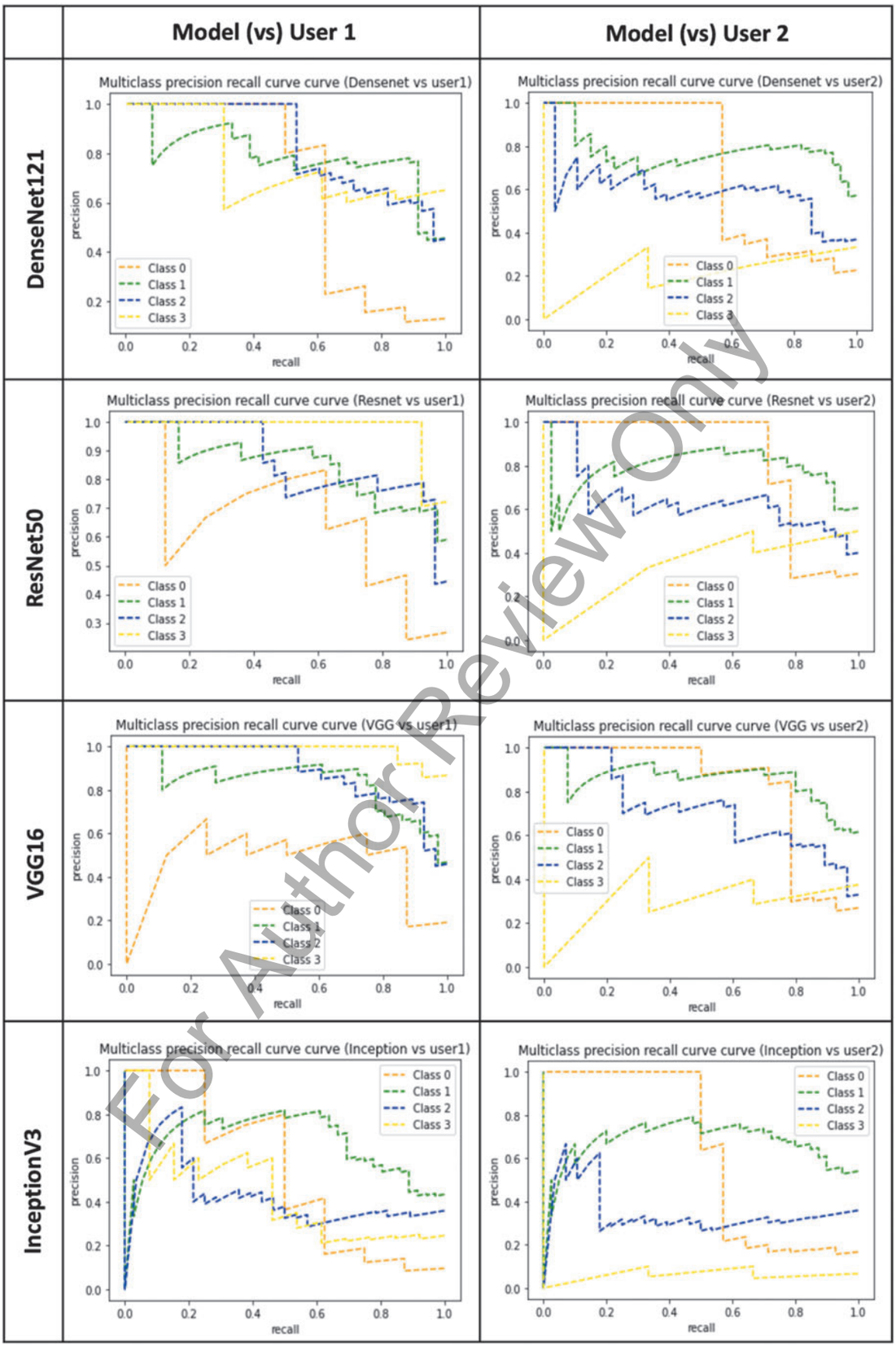


FIG. 8. Precision and recall plotted between the users and the model (i.e., model [vs.] user 1, and model [vs.] user 2) for all four different grades of OA severity, represented by multiclass precision recall curves obtained from the four different models DenseNet121, ResNet50, VGG16, and InceptionV3.

Table 5. Precision-Recall Area Under the Receiver Operating Characteristic Curve Values for the Top Two Best Performing Models With 95% Confidence Interval

	DenseNet121		*ResNet50*	
	Model vs. user 1	*Model vs. user 2*	*Model vs. user 1*	*Model vs. user 2*
Grade 0	0.65 (0.40–0.90)	0.83 (0.62–1.00)	0.67 (0.50–0.83)	0.70 (0.44–0.96)
Grade 1	0.85 (0.75–0.95)	0.79 (0.68–0.89)	0.79 (0.66–0.92)	0.78 (0.64–0.90)
Grade 2	0.86 (0.76–0.97)	0.64 (0.50–0.78)	0.83 (0.72–0.95)	0.59 (0.44–0.74)
Grade 3	0.98 (0.92–1.00)	0.34 (0.19–0.49)	0.75 (0.54–0.96)	0.22 (0.07–0.36)

have been selected in such a way that the model could visualize those parameters that are readable and visible. A modified and simplified version of the Histological and Histochemical Grading System/Mankin scoring system has been used here in this work to achieve the same, which enables the model to read the parameters within a tile hassle-free and predict the grades accordingly. Reproducibility of the scoring and grading is no more an issue here, because this DL model in itself makes it possible by eliminating bias unlike any other manual grading methodologies.

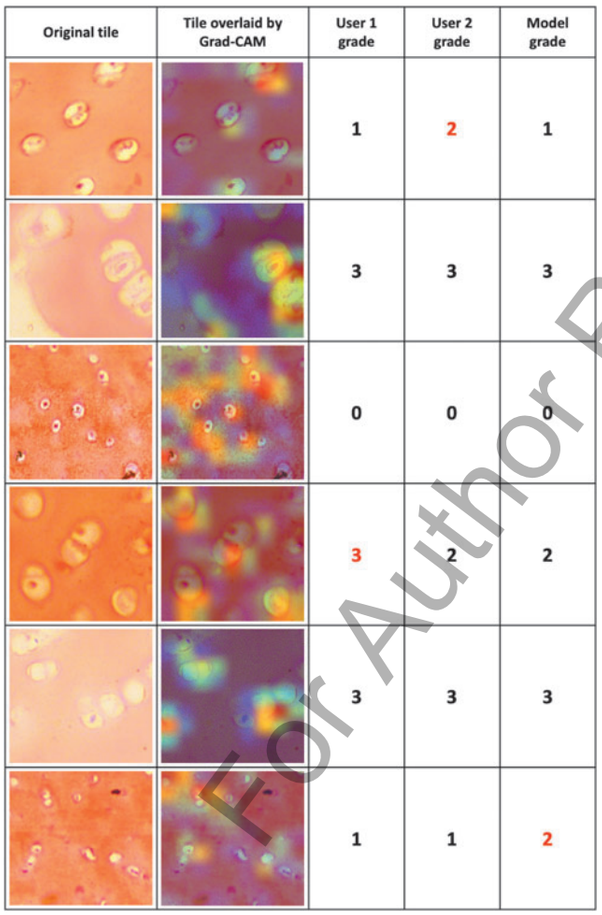


FIG. 9. Comparison between the grade labels given by user 1, user 2, and the model (here, best performing model DenseNet121 has been considered) by superimposing the original input image with the weights from the final convolution layer of the model. This rough map generated by Grad-CAM gives an idea about the regions within the tile, which the model considers to be important for predicting the grade. Grades displayed here in red are the ones that differ from the other two grades. Grad-CAM, gradient-weighted class activation mapping.

Simple and readable parameters have been chosen to avoid the complexity of classification, which can further be modified by including several other intricate details. For instance, the collagen fibril arrangement and its diameter (which will give an idea of the degree of protease-mediated collagen fibril degradation[54,55]) by picrosirius red staining, and the degree of matrix calcification by alizarin red staining can be investigated further. Thus, by expanding the parameters for scoring, we are providing additional input for the model to read few extra features from the image, which might be beneficial in improving the reliability of the score and as well as the precision-recall of the model to perform better in terms of its prediction.

The indispensable role of chondrocyte hypertrophy in OA poses difficulties in the development of a phenotypically stable cartilage.[56] Furthermore, as a symptom of putative cartilage regeneration, there is intense deposition of glycosaminoglycan content in pericellular region of clustered cells, that change during the disease progression.[57] This staining intensity indicating the newly synthesized proteoglycan can also be included as a grading parameter. However, it could not be appended in this study because of the wide variation in the tissue quality of highly osteoarthritic cartilage sections within the same image. This stands as one of the limitations, as the model might mistake it for normal glycosaminoglycan content in the extracellular matrix.

Any artifacts (including staining and processing artifacts in the slides) would influence the results of the histology assessment. Staining variability, if at all any in the images of the histological tissue sections could directly affect the standardization of the results, as it is a direct parameter in analyzing the severity. Hence as an additional endorsement, the whole histological images of the samples (instead of cropped tiles with a small dimension) with both Safranin-O staining and H&E staining were also analyzed, graded, and evaluated manually by both users 1 and 2. By manual grading system, both the users graded the samples by considering the layer-specific characteristics of healthy and OA cartilages. It was concluded from this observation that the grades obtained from manual grading system matches well with the automated DL model grades, which draws evidence that the modified grading system that was designed for grading OA severity by DL models stands felicitous in achieving our objective.

Although our model performs decent with an acceptable level of precision and recall, there is still scope for improvement. We can further reduce the annotation efforts and

increase the accuracy by using semi-supervised learning in which the model can train on data, predicting with confidence above a threshold. One of the major reasons for user disagreement in manual grading systems is the biased human visualization. This disagreement can be eliminated only if the users look at the tiles in the same way (by treating all the regions in a tile equally by giving precisely the same weightage), which is almost practically impossible. But this is not the case with the DL models. Recent work on quality control standardization for grading the chondrogenesis of a tissue engineered cartilage using DL[16] has suggested that the user disagreement could be because of varying quality of tissues within the same image, which stands true in our case as well.

We have tested only four different DL architectures here in this study. There are several other DL architectures available, which might display better results. With steady advancements in the concept of DL, it is believed that various other architectures would come into play, manifesting better ways to learn the grading parameters more effectively than the existing ones. For our best performing model DenseNet121, a schematic describing the pretrained and additional trainable layers has been given (refer Supplementary Fig. S2). Nevertheless, the aim of this study was to investigate the feasibility of automating the laborious process of OA grading with its histological images using a modified grading system, and not solely uncovering the superior model or architecture. We have thus shown here that a DL model could be trained and tested with a limited dataset of cartilage histopathology images for assessing the disease severity.

Being a distinct advancement in the field of AI and regenerative medicine, this particular work on automating the grading of human OA cartilage by DL reduces the bias and disagreement that prevails in the manual grading process and hence, this would be beneficial in sufficing the researchers on developing strategies to mitigate OA. Although histological examination is conducted only in cases with doubtful pathology, it is of great significance for research purposes. We are proposing the direct clinical utility of this work by developing automated DL models for OA grading of biopsies. With its innate faster readouts, reliability, agreement, and reproducibility to a significantly greater extent, this automated grading by DL classification models proves to be a way better and superior option to explore than the manual grading process for determining the degree of cartilage degradation, and thereby it is believed to be very propitious in analyzing the severity and extent of OA progression.

Conclusion

Given that knee arthroplasty is the only partially effective treatment for OA, researchers these days have shifted their interest toward finding a better alternative. Several experimental studies target the development of disease models, where standard histopathological assessment systems are required for distinguishing the subsets of OA and defining the end points for clinical trials. Moreover, evaluation of characteristic features and disease markers of OA, and evaluation of animal OA models call for reliable grading methodologies. There has to be a standardized method for grading with good reproducibility and reliability, which manual grading fails to achieve as it is subjective to human bias. In this study, attempts have been directed toward developing an automated grading system, which can outperform the existing manual grading systems.

The current article deals with the automation of human OA specimen grading using DL, which involves developing a scoring and grading system, and a DL model that can read the parameters of the designed grading system from the cartilage histology images accordingly. With a highly heterogeneous dataset (refer Supplementary Table S2) including both HC and OAC imaged by two different microscopes, it was possible to train the model with a wide spectrum of images resulting in better performance of the classification algorithm. One of the major benefits of this study is that the extent and severity of cartilage degradation is presented directly by the developed model without any manual intervention, which stands invincible in terms of reliability, interuser agreement, and reproducibility.

Various metrics such as the accuracy, loss, kappa score, ROC, false-positive and false-negative rates, with corresponding AUC's obtained here suggest that this DL-based automated grading system tends to provide desirable and promising outcome. These latest developments in the field of AI would implement the transition from manual to automated grading, which would be swift and precise. This concise, unbiased, and unambiguous grade prediction by DL models thus will aid in delivering improved and effective detection and diagnosis of OA by acting as a standardized method to practice for OA grading.

Authors' Contributions

L.N.: Collection and assembly of data, Analysis and interpretation of the data, Drafting of the article; A.K.: Analysis and interpretation of the data, Statistical expertise; A.S.: Analysis and interpretation of the data, Statistical expertise; R.V.: Provision of patient samples, Validation of histological grading, Final approval of the article; S.G.: Conception and design, Critical revision of the article for important intellectual content, Final approval of the article.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

This study is funded by internal funding of Indian Institute of Technology, Delhi. Grad-CAM visualizations were developed using the code adapted from keras deep learning API written in python, 2017 (https://keras.io/examples/vision/grad_cam/).

Supplementary Material

Supplementary Data

References

1. Cui A, Li H, Wang D, et al. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. EClinicalMedicine 2020;29–30: 100587; doi: 10.1016/j.eclinm.2020.100587

2. Bhattacharjee M, Coburn J, Centola M, et al. Tissue engineering strategies to study cartilage development, degeneration and regeneration. Adv Drug Deliv Rev 2015;84: 107–122; doi: 10.1016/j.addr.2014.08.010
3. Davis S, Roldo M, Blunn G, et al. Influence of the mechanical environment on the regeneration of osteochondral defects. Front Bioeng Biotechnol 2021;9:603408; doi: 10 .3389/fbioe.2021.603408
4. Chameettachal S, Midha S, Ghosh S. Regulation of chondrogenesis and hypertrophy in silk fibroin-gelatin-based 3D bioprinted constructs. ACS Biomater Sci Eng 2016;2(9): 1450–1463; doi: 10.1021/acsbiomaterials.6b00152
5. Chawla S, Mainardi A, Majumder N, et al. Chondrocyte hypertrophy in osteoarthritis: Mechanistic studies and models for the identification of new therapeutic strategies. Cells 2022;11(24):4034; doi: 10.3390/cells11244034
6. Ghosh S, Laha M, Mondal S, et al. In vitro model of mesenchymal condensation during chondrogenic development. Biomaterials 2009;30(33):6530–6540; doi: 10.1016/j .biomaterials.2009.08.019
7. Murab S, Chameettachal S, Bhattacharjee M, et al. Matrix-embedded cytokines to simulate osteoarthritis-like cartilage microenvironments. Tissue Eng Part A 2013;19(15–16): 1733–1753; doi: 10.1089/ten.tea.2012.0385
8. Seda Tigli R, Ghosh S, Laha MM, et al. Comparative chondrogenesis of human cell sources in 3D scaffolds. J Tissue Eng Regen Med 2009;3(5):348–360; doi: 10.1002/term.169
9. Chakraborty J, Chawla S, Ghosh S. Developmental biology-inspired tissue engineering by combining organoids and 3D bioprinting. Curr Opin Biotechnol 2022;78: 102832; doi: 10.1016/j.copbio.2022.102832
10. Litwic A, Edwards MH, Dennison EM, et al. Epidemiology and burden of osteoarthritis. Br Med Bull 2013;105(1): 185–199; doi: 10.1093/bmb/lds038
11. Roemer FW, Guermazi A, Demehri S, et al. Imaging in osteoarthritis. Osteoarthritis Cartilage 2022;30(7):913–934; doi: 10.1016/j.joca.2021.04.018
12. Bernotiene E, Bagdonas E, Kirdaite G, et al. Emerging technologies and platforms for the immunodetection of multiple biochemical markers in osteoarthritis research and therapy. Front Med 2020;7:572977; doi: 10.3389/fmed .2020.572977
13. Pauli C, Whiteside R, Heras FL, et al. Comparison of cartilage histopathology assessment systems on human knee joints at all stages of osteoarthritis development. Osteoarthritis Cartilage 2012;20(6):476–485; doi: 10.1016/j .joca.2011.12.018
14. Nicora G, Rios M, Abu-Hanna A, et al. Evaluating pointwise reliability of machine learning prediction. J Biomed Inform 2022;127:103996; doi: 10.1016/j.jbi.2022.103996
15. Pearson RG, Kurien T, Shu KSS, et al. Histopathology grading systems for characterisation of human knee osteoarthritis—Reproducibility, variability, reliability, correlation, and validity. Osteoarthritis Cartilage 2011;19(3):324–331; doi: 10.1016/j.joca.2010.12.005
16. Power L, Acevedo L, Yamashita R, et al. Deep learning enables the automation of grading histological tissue engineered cartilage images for quality control standardization. Osteoarthritis Cartilage 2021;29(3):433–443; doi: 10 .1016/j.joca.2020.12.018
17. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: Switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Health 2020;2(9):e486–e488; doi: 10.1016/S2589-7500(20)30160-6
18. Rong G, Mendez A, Bou Assi E, et al. Artificial intelligence in healthcare: Review and prediction case studies. Engineering 2020;6(3):291–301; doi: 10.1016/j.eng.2019.08.015
19. Greener JG, Kandathil SM, Moffat L, et al. A guide to machine learning for biologists. Nat Rev Mol Cell Biol 2022;23(1):40–55; doi: 10.1038/s41580-021-00407-0
20. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015;349(6245):255–260; doi: 10.1126/science.aaa8415
21. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019;6(1):60; doi: 10.1186/s40537-019-0197-0
22. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: Las Vegas, NV; 2016; pp. 770–778; doi: 10.1109/CVPR.2016.90
23. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE: Honolulu, HI; 2017; pp. 2261–2269; doi: 10.1109/CVPR.2017.243
24. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv: Computer Vision and Pattern Recognition 2015; preprint arXiv:1409.1556; doi: 10.48550/arXiv.1409.1556
25. Szegedy C, Wei Liu, Yangqing Jia, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: Boston, MA; 2015; pp. 1–9; doi: 10.1109/CVPR.2015.7298594
26. Jung Y. Multiple predicting *K*-fold cross-validation for model selection. J Nonparametric Stat 2018;30(1):197–215; doi: 10.1080/10485252.2017.1404598
27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-Learn: Machine Learning in Python. J Mach Learn Res 2011; 2825.2830; doi: 10.48550/arXiv.1201.0490
28. McHugh ML. Interrater reliability: The kappa statistic. Biochem Medica 2012;22(3):276–282; doi: 10.11613/BM .2012.031
29. Heydarian M, Doyle TE, Samavi R. MLCM: Multi-label confusion matrix. IEEE Access 2022;10:19083–19095; doi: 10.1109/ACCESS.2022.3151048
30. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Casp J Intern Med 2013;4(2):627–635; doi: 10.1097/JTO .0b013e3181ec173d
31. Beger A. Precision-recall curves. SSRN Electron J 2016; doi: 10.2139/ssrn.2765419
32. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Int J Comput Vis 2020;128(2):336–359; doi: 10.1007/s11263-019-01228-7
33. Cohen JF, Chalumeau M, Cohen R, et al. Cochran's Q test was useful to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy. J Clin Epidemiol 2015;68(3):299–306; doi: 10.1016/j.jclinepi.2014.09 .005
34. Liu Y, Gadepalli K, Norouzi M, et al. Detecting Cancer Metastases on Gigapixel Pathology Images. arXiv: Computer Vision and Pattern Recognition 2017; preprint arXiv: 1703.02442; doi: 10.48550/arXiv.1703.02442
35. Cireşan DC, Giusti A, Gambardella LM, et al. Mitosis detection in breast cancer histology images with deep neural networks. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013. Lecture Notes in Computer Science. (Mori K, Sakuma I, Sato Y,

et al. eds.) Springer Berlin Heidelberg: Berlin, Heidelberg; 2013; pp. 411–418; doi: 10.1007/978-3-642-40763-5_51
36. Rivenson Y, Wang H, Wei Z, et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. Nat Biomed Eng 2019;3(6):466–477; doi: 10.1038/s41551-019-0362-y
37. Gaj S, Yang M, Nakamura K, et al. Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks. Magn Reson Med 2020;84(1):437–449; doi: 10.1002/mrm.28111
38. Abdullah SS, Rajasekaran MP. Automatic detection and classification of knee osteoarthritis using deep learning approach. Radiol Med (Torino) 2022;127(4):398–406; doi: 10.1007/s11547-022-01476-7
39. Chen P, Gao L, Shi X, et al. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Comput Med Imaging Graph 2019;75:84–92; doi: 10.1016/j.compmedimag.2019.06.002
40. Liu B, Luo J, Huang H. Toward automatic quantification of knee osteoarthritis severity using improved Faster R-CNN. Int J Comput Assist Radiol Surg 2020;15(3):457–466; doi: 10.1007/s11548-019-02096-9
41. Olsson S, Akbarian E, Lind A, et al. Automating classification of osteoarthritis according to Kellgren-Lawrence in the knee using deep learning in an unfiltered adult population. BMC Musculoskelet Disord 2021;22(1):844; doi: 10.1186/s12891-021-04722-7
42. Rytky SJO, Tiulpin A, Frondelius T, et al. Automating three-dimensional osteoarthritis histopathological grading of human osteochondral tissue using machine learning on contrast-enhanced micro-computed tomography. Osteoarthritis Cartilage 2020;28(8):1133–1144; doi: 10.1016/j.joca.2020.05.002
43. Mormont R, Geurts P, Maree R. Comparison of deep transfer learning strategies for digital pathology. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE: Salt Lake City, UT; 2018; pp. 2343–234309; doi: 10.1109/CVPRW.2018.00303
44. Swagerty DL, Hellinger D. Radiographic assessment of osteoarthritis. Am Fam Physician 2001;64(2):279–286.
45. Greif DN, Epstein AL, Hodgens BH, et al. Current measurement strategies of coronal tibiofemoral subluxation: A systematic review of literature. AJR Am J Roentgenol 2021;216(5):1183–1192; doi: 10.2214/AJR.20.23503
46. Khamaisy S, Zuiderbaan HA, Thein R, et al. Coronal tibiofemoral subluxation: A new measurement method. The Knee 2014;21(6):1069–1071; doi: 10.1016/j.knee.2014.07.013
47. Khamaisy S, Zuiderbaan HA, Thein R, et al. Coronal tibiofemoral subluxation in knee osteoarthritis. Skeletal Radiol 2016;45(1):57–61; doi: 10.1007/s00256-015-2244-z
48. Abedin J, Antony J, McGuinness K, et al. Predicting knee osteoarthritis severity: Comparative modeling based on patient's data and plain X-ray images. Sci Rep 2019;9(1):5761; doi: 10.1038/s41598-019-42215-9
49. Norman B, Pedoia V, Noworolski A, et al. Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs. J Digit Imaging 2019;32(3):471–477; doi: 10.1007/s10278-018-0098-3
50. Tiulpin A, Thevenot J, Rahtu E, et al. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. Sci Rep 2018;8(1):1727; doi: 10.1038/s41598-018-20132-7
51. Thomas KA, Kidziński Ł, Halilaj E, et al. Automated classification of radiographic knee osteoarthritis severity using deep neural networks. Radiol Artif Intell 2020;2(2):e190065; doi: 10.1148/ryai.2020190065
52. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16(4):494–502; doi: 10.1136/ard.16.4.494
53. Teoh YX, Lai KW, Usman J, et al. Discovering knee osteoarthritis imaging features for diagnosis and prognosis: Review of manual imaging grading and machine learning approaches. J Healthc Eng 2022;2022:4138666; doi: 10.1155/2022/4138666
54. Gottardi R, Hansen U, Raiteri R, et al. Supramolecular organization of collagen fibrils in healthy and osteoarthritic human knee and hip joint cartilage. Orgel JPRO. ed. PLoS One 2016;11(10):e0163552; doi: 10.1371/journal.pone.0163552
55. Walker GD, Fischer M, Gannon J, et al. Expression of type-X collagen in osteoarthritis. J Orthop Res 1995;13(1):4–12; doi: 10.1002/jor.1100130104
56. Majumder N, Ghosh S. Unfolding the mystery behind the onset of chondrocyte hypertrophy during chondrogenesis: Toward designing advanced permanent cartilage-mimetic biomaterials. Adv Funct Mater 2023;33(35):2300651; doi: 10.1002/adfm.202300651
57. van der Kraan PM, van den Berg WB. Chondrocyte hypertrophy and osteoarthritis: Role in initiation and progression of cartilage degeneration? Osteoarthritis Cartilage 2012;20(3):223–232; doi: 10.1016/j.joca.2011.12.003

Address correspondence to:
Sourabh Ghosh, PhD
Department of Textile and Fibre Engineering
Indian Institute of Technology Delhi
Hauz Khas
New Delhi 110016
India

E-mail: sourabh.ghosh@textile.iitd.ac.in

Received: August 3, 2023
Accepted: November 8, 2023
Online Publication Date: December 15, 2023