MEKATRONIKA
e-ISSN: 2637-0883
VOL. 4, ISSUE 1, 104 – 109
DOI: https://doi.org/10.15282/mekatronika.v4i1.8676

ORIGINAL ARTICLE

Automated Detection of Knee Cartilage Region in X-ray Image

Teo Jia Chern[1], Ismail Mohd Khairuddin[1,*], Mohd Azraai Mohd Razman[1], Anwar P.P Abdul Majeed[1], and Wan Hasbullah Mohd Isa[1]

[1]Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.

ABSTRACT – The prevalence of a symptomatic knee or osteoarthritis (OA) is approximately 9.6% in men and 18.0% in women over 60 years of age according to the OARSI 2016 report. Using early on-stage clinical qualitative assessments through means of X-ray scans, the cartilage health and degradation of an individual can be monitored through cartilage shape and surface over time. In this paper, we implement the application of transfer learning models such as InceptionV3, Xception and DenseNet201 for feature extraction of a rebalanced 1,000 knee X-ray images taken from Osteoarthritis Initiative (OAI) dataset with 5 classes graded 0–4 according to Kellgren-Lawrence grading split into a 70/15/15 training/validation/testing split. The features extracted are subsequently fed into machine learning classifiers, namely support vector machine (SVM). An average multiclass accuracy of 71.33% was achieved for hyperparameter fine-tuned DenseNet201-SVM model.

ARTICLE HISTORY
Received: 16th May 2022
Revised: 5th June 2022
Accepted: 28th June 2022

KEYWORDS
Knee osteoarthritis
Transfer learning
Deep learning
Machine learning
Classification
Osteoarthritis Initiative

INTRODUCTION

The knee appears to be the largest joint in the human body, serving as a critical site for movement between the thigh and lower leg to establish movement and locomotion (Stoker, 1980). Regardless of the complexity of the knee, it is prone to infections and injuries due to vigorous physical activity or physical stress in collisions or accidents. Research suggests that monitoring cartilage thickness, volume, and surface over time can be used to assess a person's cartilage degradation [1].

Currently, methods of quantifying the progress of degeneration are limited and include the classification of knee osteoarthritis (OA) severity grading health based on Kellgren-Lawrence grading (KLG). Despite medical imaging modalities such as MRI, optical coherence tomography, and ultrasound for OA diagnosis, radiography (X-ray) has been traditionally preferred and remains the main accessible tool or "gold standard" for preliminary knee OA diagnosis [2].

Machine learning has exploded in prominence in medical applications in recent years, transforming the way large amounts of medical data are processed and analysed. Generally, machine learning denotes a set of mathematical algorithms that "teach" or "train" a machine to determine the correlation between an input and output data without explicit instructions. Deep learning, a branch of machine learning that focuses on analysing images and knowledge extraction from high volumes of data, including medical scans, has been used by radiologists and orthopaedic surgeons to provide automatic interpretations of medical pictures, improving diagnosis accuracy and speed [3].

The primary objective of this research is to evaluate the efficacy of transfer learning models for the feature extraction of knee OA severity grading based on KLG. Further subobjectives of this study is to develop an accurate multi-class classification framework based on machine learning to classify knee OA severity in X-ray images.

The remainder of this paper is organized as follows: Related literature reviews and summations are discussed in Section 2, which highlights on the model's performance, improvement techniques, parameters, and limitations. In Section 3, the methodology and workflow are discussed in detail, which outlines the framework of the pipelines used, applied techniques and performance metrics selection and evaluation. The results of the methodology are then presented and discussed in Section 4 through the application of performance metrics. Finally, the conclusion of the thesis and the best pipeline deployment would be discussed in Section 5.

RELATED WORK

This part provides a brief overview of the literature review regarding in which all related articles and journals that meet the research objectives are examined and reviewed. Literature review is done on frameworks and models used for the classification of the knee OA severity grading and related literature (classification/segmentation of anterior cruciate ligament (ACL), femoral and patellar joints, etc).

General classification of knee OA severity grading based on Kellgren-Lawrence scores is carried out on the OAI dataset. Deep learning methods for the classification of knee KL grade include the development of custom convolutional neural networks (CNN) with the distinction of modified final layers in addition to softmax non-linear functions for probabilitistic representation of five KLG scores [4].

*CORRESPONDING AUTHOR | Ismail Mohd Khairuddin | ismailkhai@ump.edu.my

Further exploration of deep learning methods include customized one-stage detection architecture YOLOv2 with a novel ordinal loss as a replacement for cross-entropy loss in fine tuning of various CNN architectures such as VGG16, InceptionV3, variants of DenseNet and ResNet [5].

The application of reappplying pre-trained CNNs based on similarity to the target domain and adapting to a new repurposed tasks, allowing for rapid progress when modelling the second task represents the essence of transfer learning.

The implementation of transfer learning with pretrained ImageNet weights and retraining its last softmax layer using the target domain dataset was done by [6] with input images scaled to 224x224 or 299x299, depending on the architecture (VGG, ResNet, InceptionV3, DenseNet) used.

The implementation of transfer learning models for feature extraction and subsequent machine learning classifiers for knee OA severity classification was first attempted by [7], where features from the convolutional, pooling and fully connected layers are extracted by using pre-trained networks such as VGG16, VGG-M-128 and BVLC CaffeNet, and subsequently, fed into trained linear SVMs for classification. Formulating the classification of KL grades as a regression, a mean squared loss was used to fine-tune BVLC CaffeNet for knee KL grade.

METHODOLOGY

Process Flow

The study is seperated in 3 phases. In phase 1, there is the research phase, in which all related articles and journals that meet the research objectives are examined and reviewed. A literature review is done on frameworks and models used for the classification of the knee OA severity grading and related literature (classification/segmentation of anterior cruciate ligament (ACL), femoral and patellar joints, etc).

Data collection employs the use of popular datasets used in previous literatures, sourced from reputable repositories, such as GitHub, Kaggle, Mendeley. Data pre-processing enquires the use of general dataset preparation techniques, including image resizing, orientation fixing, noise removal, normalization (colour normalization and N4 bias field correction) and data augmentation (for smaller datasets, pixel shifting).

In phase 2, there will be focus on the development and evaluation of popular transfer learning architectures based on previous literature review such as InceptionV3, DenseNet201 and Xception for feature extraction, along with machine learning classifiers such as support vector machine (SVM), random forests (RF) and logistic regression (LR) for the classification of OA severity. The best models from preliminary benchmarking for each classifier would be subjected to hyperparameter optimization to increase the accuracy of each model.

In phase 3, models generated in the previous phase 2 would be evaluated and compared with each other in performance metrics in terms of accuracy, a widely accepted metric in the classification of knee OA severity.

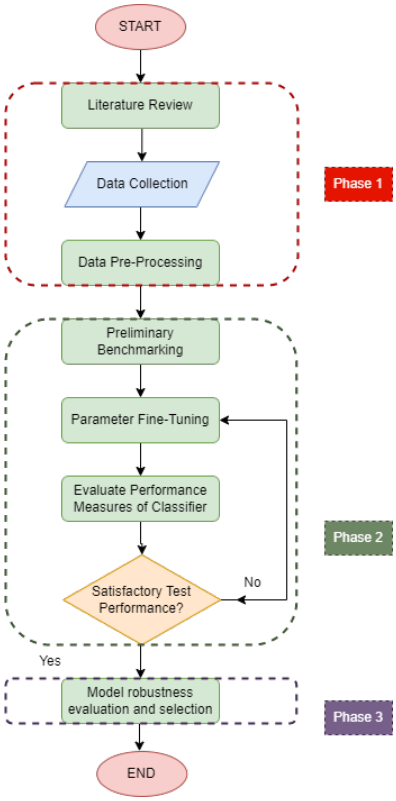

Figure 1. Process Flow Chart.

Dataset

The dataset of knee X-ray images was retrieved from the public open-source dataset, Osteoarthritis Initiative (OAI), a multi-centre, ten-year observational study of men and women, sponsored

by the National Institutes of Health (part of the Department of Health and Human Services). The goals of the OAI dataset were to provide resources to enable better understanding of prevention and treatment of knee osteoarthritis, one of the most common causes of disability in adults.

Using a modified, reorganized version of the OAI dataset [5], our dataset consists of 1,000 knee X-ray images with 5 classes of knee osteoarthritis severity grading (Kellgren-Lawrence grading of 0 – 4) divided into 200 images per class to reduce the influence of other classes weights due to a lack of grade 4 images [6].

Images are resized to fit the input of different transfer learning models used in this study such as 299x299 for InceptionV3 and Xception and 224x224 for DenseNet201. A training/validation/test split of 70/15/15 was carried on the images.

Feature Extraction

Transfer learning represents the machine-learning method of reapplying pre-trained models based on similarity to the target domain, and adapting to a new repurposed task, allowing for rapid progress when modelling the second task. Features extracted via transfer learning models are often rich and essential features that can be subsequently fed into image classification models [12].

InceptionV3 is CNN for image analysis and object detection. Being the third edition of Google's Inception CNN, it comes with a multi-level feature extractor, computing 1x1, 3x3, 5x5 convolutions within same module, concatenate results into a single output, originally introduced during the ImageNet Recognition Challenge.

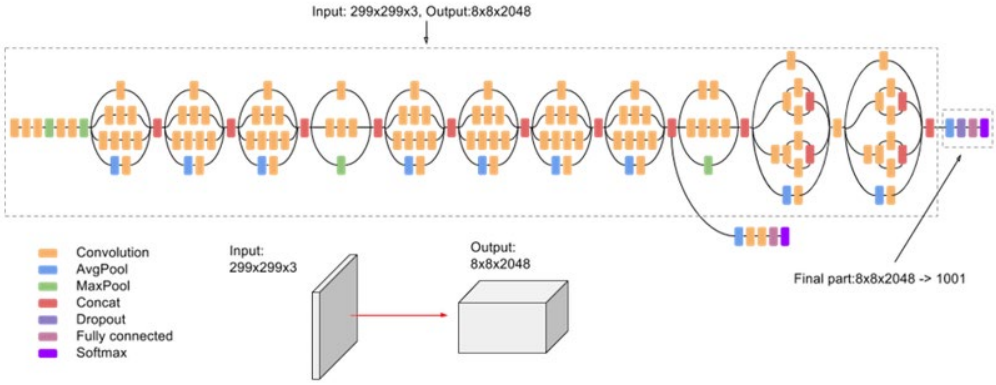

Figure 2. InceptionV3 model architecture.

Xception is a deep CNN that is 71 layers deep and involves Depthwise Seperable Convolutions. Being a more "extreme" version of an Inception module (outperforming it on ImageNet dataset), it takes into consideration depth, to capture cross-channel correlation. Unlike Inception's 1x1 convolutions used for compression of original input followed by filters on each depth space, Xception uses filters on each depth map prior to compression of input space using 1x1 convolution, applying it across the depth.

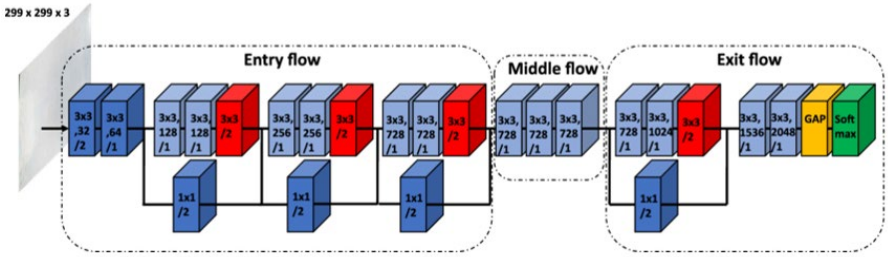

Figure 3. Xception model architecture.

DenseNet201 is a dense CNN that is 201 layers deep which connects each layer to every other layer in a feed-forward fashion, with each layer obtaining additional inputs from all preceding layers and passing it onto subsequent layers. This allows for diversified features and maintains low complexity features.

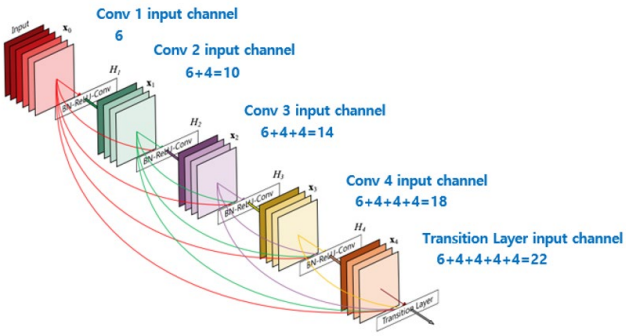

Figure 4. DenseNet201 model architecture.

Feature Extraction

For feature classification, the implementation of ML methods such as SVM, LR and RF are used. Predictive performance of simpler ML techniques allows for efficient low size feature spaces and good generalization in a significant number of studies [8].

SVM is a machine learning algorithm that performs supervised learning for classification or regression analysis of data groups. At first approximation, SVM generates a hyperplane that separates data into different classes. By finding the distance between the line and support vectors (points closest to the line from both classes), we can compute the margin, with the goal of maximising margin through training, creating the optimum hyperplane.

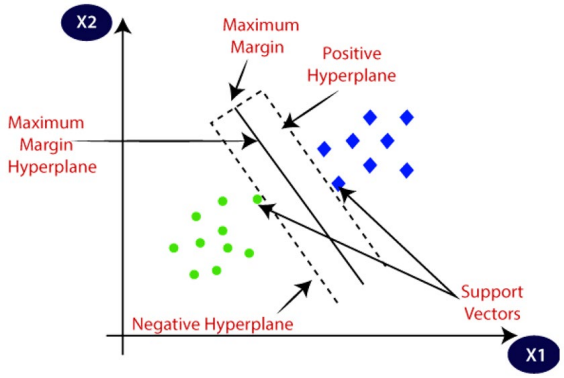

Figure 5. Support Vector Machine.

Hyperparameter Fine Tuning

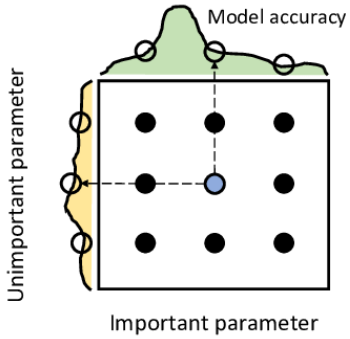

Figure 6. Grid Search Method.

Grid search is a tuning technique uses different combination of hyperparameters, calculates the performance of each and generates the optimum value of hyperparameters for the model.

Performance metrics

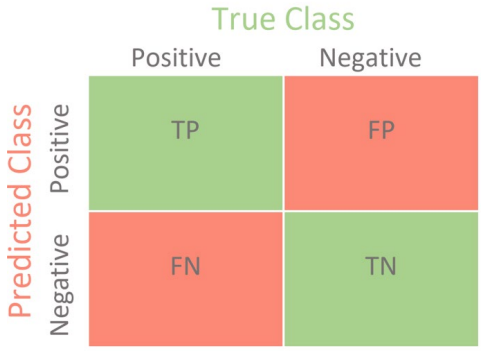

Figure 7. Confusion Matrix.

Performance metrics measured extensively in this study is accuracy as in Equation 1, calculated from the confusion matrix.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Table 1. Average 5-class OA Severity Grading Accuracy.

Author	Models	Author
(Guida et al., 2021) [6]	InceptionResNetV2	55.5%
	InceptionV3	54.0%
	DenseNet121	55.0%
(Antony et al., 2017) [7]	CNN	60.3%
(Zhang et al., 2020) [11]	ResNet-34	73.75%
	ResNet34 + CBAM	74.81%
(Chen et al., 2019) [5]	VGG19	70.4%
(Thomas et al., 2020b) [4]	CNN	64.35%
Average multiclass average		64.35%
Multiclass accuracy range		54.0% ~ 74.81%

Based on Table 1, the literature regarding 5-class OA severity grading accuracy based on OAI dataset presents an average multiclass accuracy of 64.35% with a range of 54.0% to 74.81%.

EXPERIMENTAL RESULTS

Preliminary Benchmarking

Table 2. Preliminary Benchmarking Summary.

Classification Accuracy (%)				
		Inception V3	Xception	DenseNet201
SVM	Training	94	76	84
	Validation	47	47	56
	Testing	49	47	55
	Average	63.3	56.7	64.7

Based on Table 3, the DenseNet201 model architecture has performed the best with all machine learning classifiers used. Therefore, all classifiers in DenseNet201 are subjected to hyperparameter fine tuning.

Table 3. Hyperparameter Results.

Classification Accuracy (%)			
		Before tuning	After tuning
SVM	Training	94	100
	Validation	47	59
	Testing	49	55
	Average	63.3	71.33

For SVM, parameters set for fine tuning include kernel, C, and gamma. Kernel represents the algorithm for mapping observations into a feature space, with feature range of 'linear', 'polynomial', and 'radial basis function' used. C represents the penalty, where a larger C would cause the SVM to minimize the number of misclassified examples due to high penalty, leaving smaller margin of errors, with a range of 0.01, 0.1 1, 10, and 100 used. Gamma represents the separation line, where a low gamma would cause a large similarity radius, causing more data to get grouped together. High gamma often results in overfitting due to points being close and any noise will cause the data to fall out of the class with a range of 0.01, 0.1, 1, 10, and 100 used. Best accuracy was achieved with parameters of polynomial kernel, C of 0.01, and gamma of 0.01.

CONCLUSION

In conclusion, DenseNet201-SVM model performed the best, achieving a multiclass average accuracy of 71.33%, all surpassing the multiclass average accuracy based on literature review.

Future recommendation following this study include hybrid class balancing to utilize simulated annealing algorithms for under-sampling and machine learning applications for mitigation of misclassification due to imbalanced datasets [9].

Furthermore, X-rays are unable to show certain structural phenotypes of OA and accurately monitor progress of OA unlike MRI [10]. Recommendations include examining both MRIs and X-ray images [6].

REFERENCES

[1] Choi, J.-A., & Gold, G. E. (2011). MR imaging of articular cartilage physiology. Magnetic Resonance Imaging Clinics of North America, 19(2), 249–282. https://doi.org/10.1016/j.mric.2011.02.010.

[2] Shamir, L., Ling, S. M., Scott, W. W., Bos, A., Orlov, N., MacUra, T. J., Eckley, D. M., Ferrucci, L., & Goldberg, I. G. (2009). Knee X-ray image analysis method for automated detection of osteoarthritis. IEEE Transactions on Biomedical Engineering, 56(2), 407–415. https://doi.org/10.1109/TBME.2008.2006025.

[3] Borjali, A., Chen, A. F., Muratoglu, O. K., Morid, M. A., & Varadarajan, K. M. (2020). Deep Learning in Orthopedics: How Do We Build Trust in the Machine? Healthcare Transformation. https://doi.org/10.1089/heat.2019.0006.

[4] Thomas, K. A., Kidziński, Ł., Halilaj, E., Fleming, S. L., Venkataraman, G. R., Oei, E. H. G., Gold, G. E., & Delp, S. L. (2020b). Automated Classification of Radiographic Knee Osteoarthritis Severity Using Deep Neural Networks. Radiology: Artificial Intelligence, 2(2), e190065. https://doi.org/10.1148/ryai.2020190065.

[5] Chen, P., Gao, L., Shi, X., Allen, K., & Yang, L. (2019). Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Computerized Medical Imaging and Graphics, 75, 84–92. https://doi.org/10.1016/j.compmedimag.2019.06.002.

[6] Guida, C., Zhang, M., & Shan, J. (2021). Knee osteoarthritis classification using 3D CNN and MRI. Applied Sciences (Switzerland), 11(11). https://doi.org/10.3390/app11115196.

[7] Antony, J., McGuinness, K., Moran, K., & O'Connor, N. E. (2017). Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10358 LNAI, 376–390. https://doi.org/10.1007/978-3-319-62416-7_27.

[8] Kokkotis, C., Moustakidis, S., Papageorgiou, E., Giakas, G., & Tsaopoulos, D. E. (2020). Machine learning in knee osteoarthritis: A review. Osteoarthritis and Cartilage Open, 2(3), 100069. https://doi.org/10.1016/j.ocarto.2020.100069

[9] Desuky, A. S., & Hussain, S. (2021). An Improved Hybrid Approach for Handling Class Imbalance Problem. Arabian Journal for Science and Engineering, 46(4), 3853–3864. https://doi.org/10.1007/s13369-021-05347-7.

[10] Roemer, F. W., Kwoh, C. K., Hayashi, D., Felson, D. T., & Guermazi, A. (2018). The role of radiography and MRI for eligibility assessment in DMOAD trials of knee OA. In Nature Reviews Rheumatology (Vol. 14, Issue 6, pp. 372–380). Nature Publishing Group. https://doi.org/10.1038/s41584-018-0010-z.

[11] Zhang, B., Tan, J., Cho, K., Chang, G., & Deniz, C. M. (2020). Attention-based CNN for KL Grade Classification: Data from the Osteoarthritis Initiative. Proceedings - International Symposium on Biomedical Imaging, 2020-April, 731–735. https://doi.org/10.1109/ISBI45749.2020.9098456.

[12] J. Z. Lee and A. P. P. Abdul Majeed, "Classification Of Skin Cancer By Means Of Transfer Learning Models", MEKATRONIKA, vol. 3, no. 2, pp. 77–81, Dec. 2021.