X-RADE: ADVANCED OSTEOARTHRITIS ASSESSMENT THROUGH DEEP LEARNING

Tina Dudeja[a*], Aadish Jain[b], Aakash Raturi[c], Aniket Shobit[d], Harsh Kumar Mishra[e]

Abstract

Knee osteoarthritis, prevalent among the elderly, involves gradual articular cartilage degradation. The Kellgren-Lawrence grading system assesses severity using radiographic features. Our study incorporates EfficientNetB5, offering insights into its potential efficacy for OA classification. Consolidating Kellgren-Lawrence grades significantly improves accuracy. Algorithm scores align with MOST dataset and radiologist .The Method Used is knee osteoarthritis (OA) classification study, we employed the advanced EfficientNetB5 architecture. The methodology included meticulous dataset curation, transfer learning, and strategic data augmentation to enhance model performance, ensuring its adaptability to varying severity levels. The results demonstrates commendable proficiency with an accuracy of 64%, showcasing its effectiveness in accurately classifying different severity levels of knee OA. This underscores the potential of EfficientNetB5 in precise knee osteoarthritis severity assessment.

tinadudeja@bpitindia.com[a*], aadishjain360@gmail.com[b], aakashraturi2001@gmail.com[c], shobhitaniket3@gmail.com[d], kharsh822@gmail.com[e]

[a,b,c,d,e]Computer Science and Engineering, Bhagwan Parshuram Institute of Technology, Delhi, India

[*]**Corresponding Author:** Tina Dudeja
Computer Science and Engineering, Bhagwan Parshuram Institute of Technology, Delhi, India, tinadudeja@bpitindia.com

DOI: 10.53555/ecb/2022.11.10.98

1. Introduction

Knee osteoarthritis (OA) is a multifaceted degenerative disorder that significantly impacts the aging population, particularly individuals aged 45 to 80, with a disproportionate prevalence among women. This complex condition, characterized by the progressive deterioration of articular cartilage, has far-reaching implications for the quality of life of those affected. Accurate identification and grading of OA severity are paramount due to the debilitating symptoms that profoundly affect daily activities. The constellation of pain, stiffness, swelling, and restricted mobility intricately hampers essential tasks such as walking, climbing stairs, and bending down, highlighting the pervasive impact of knee OA on functional independence and overall well-being.

At its core, knee OA is influenced by a combination of factors, with the aging process playing a pivotal role in the natural wear and tear of joint cartilage. Genetic susceptibility adds another layer of complexity, particularly for individuals with a family history of knee OA, emphasizing the interplay between genetics and environmental factors in the development of the condition. Additionally, obesity emerges as a significant risk factor, as the surplus weight amplifies stress on the knee joints, hastening the breakdown of cartilage and triggering inflammatory processes. This intricate interplay of age, genetics, and lifestyle factors underscores the need for a comprehensive understanding of knee OA to tailor effective interventions and improve patient outcomes.

The accurate detection and classification of knee OA assume critical importance, particularly in the absence of medical interventions capable of regenerating articular cartilage. This necessity is underscored by the fact that knee OA is a chronic condition with no definitive cure, and treatment strategies are focused on managing symptoms and improving function. The grading of OA severity not only guides tailored treatment plans but also serves as a fundamental tool for healthcare providers to monitor disease progression, identify potential complications or comorbidities, and make informed decisions about the timing and nature of interventions.

In clinical practice, knee OA diagnosis involves a multifaceted approach that extends beyond traditional clinical examination. Imaging techniques, including X-rays and MRI, provide crucial insights into the structural changes occurring in the knee joint. Among these, X-rays, specifically utilizing the Kellgren-Lawrence (KL) grading scale, stand out as a widely adopted and standardized diagnostic tool. This grading system enables the visualization of joint changes, the development of bone spurs, and a nuanced assessment of the severity of knee OA. The KL system categorizes knee conditions into five grades (0-4), offering a comprehensive framework based on radiographic findings.

Delving into the specifics of the KL grading system, Grade 0 signifies the absence of radiographic evidence of osteoarthritis, providing a baseline for a healthy joint. Grade 1 introduces suspicion, with observable osteophytes and mild joint space narrowing, prompting a closer examination. A concrete diagnosis emerges at Grade 2, characterized by mild joint space narrowing and the presence of osteophytes. The progression to Grade 3 denotes moderate to severe joint space narrowing, numerous

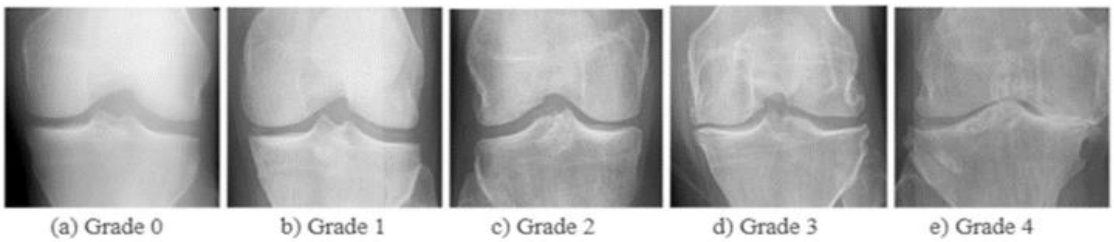

Fig 1. The classification of different grades in knee osteoarthritis

osteophytes, and possible sclerosis, indicating a more advanced stage of knee OA. Grade 4 represents severe joint space narrowing, the presence of large osteophytes, acute sclerosis, and the potential for bony deformities, reflecting a critical stage demanding careful management.

It is essential to note that the KL grading system, while providing valuable radiographic insights, solely focuses on objective findings and omits subjective factors such as pain, functional impairment, or patient-reported outcomes. Nevertheless, this comprehensive approach to knee OA diagnosis, integrating various clinical and imaging factors, establishes a robust foundation for a more accurate understanding of the condition, paving the way for improved patient care and outcomes.

In conclusion, the intricate interplay of age, genetics, and lifestyle factors in the development of knee osteoarthritis underscores the complexity of this degenerative disorder. Accurate detection and classification, guided by the Kellgren-Lawrence grading system, are crucial for tailored treatment plans and effective disease management. This comprehensive approach, integrating various clinical and radiographic factors, enhances our understanding of knee OA and facilitates a more nuanced and personalized approach to patient care.

2) Related Work

Cutting-edge studies confirm CNNs' efficacy in knee OA classification, achieving remarkable accuracy. Automated assessment using EfficientNetB5 signifies a breakthrough, offering an efficient, objective alternative. This transformative shift enhances precision, paving the way for streamlined, accurate knee OA diagnostics:

A) Precision-Oriented CNN for Automated KL-Grade Prediction: Insights from the Osteoarthritis Initiative

In a pioneering study[7], a Residual Neural Network (ResNet) with a Convolutional Block Attention Module (CBAM) was employed for automated Kellgren-Lawrence (KL) grade prediction in knee radiographs. Demonstrating remarkable efficacy[14], the approach achieved an MSE of 0.36, a Quadratic Kappa score of 0.88, and a multi-class average accuracy of 74.81%, signifying a significant leap in automated KL-grade classification. The success is attributed to sophisticated preprocessing techniques[5] and the integration of CBAM, enhancing the CNN's ability to discern intricate patterns associated with knee osteoarthritis severity[11]. This study marks a notable advancement in musculoskeletal disorder assessment[3].

B) Enhancing Image Precision: A Deep Dive into Compression and Excitation for Improved Accuracy

Another study[8] enhanced knee osteoarthritis classification by incorporating compression and excitation blocks into ResNet and DenseNet CNNs. Results showed a notable 1-3% improvement in OA classification on the KL scale[12]. Merging KL grades 0 and 1 boosted precision by 12.74%, and an ensemble of three DenseNet-121-based networks achieved an 84.66% classification accuracy, surpassing previous outcomes[19]. This innovative technique significantly improved osteoarthritis classification (1-3% on the KL scale)[15] without deviating from conventional methods, highlighting its potential to enhance automated severity grading[6]. The study underscores the compatibility and synergy of this approach with established diagnostic standards[9].

C) Revolutionary Grading Precision: A Novel Ordinal Loss Approach for Knee Osteoarthritis Severity

In a unique approach[17], two deep CNN models autonomously estimated knee osteoarthritis intensity based on the Kellgren-Lawrence scale. A specialized YOLOv2 network detected knee joints[13], and common CNN models were fine-tuned using a configurable ordinal loss for joint picture categorization[4]. The method achieved a mean Jaccard index of 0.858 for knee joint identification[11]. The model with the novel ordinal loss exhibited a maximum classification accuracy of 69.7% and the lowest mean absolute error (MAE) of 0.344 on the knee KL grading test, surpassing previous benchmarks[6]. This innovative technique, designed to accommodate the ordinal nature of KL grading[2], demonstrated superior performance in knee osteoarthritis severity assessment[8].

D) Deep Residual Learning for Image Recognition

The VGG architecture[7], celebrated for its simplicity, features a series of convolutional layers with 3x3 filters followed by max-pooling layers, providing a framework for easy and efficient training of deep networks. In contrast to some predecessors that utilize larger filters, VGG maintains uniformity throughout the network, contributing to its straightforward design. Concluding with fully connected layers and employing a soft-max classifier for image classification, VGG16 and VGG19, denoting the number of weight layers, stand out as popular configurations in the field[11].

During the 2014 ImageNet Large Scale Visual Recognition Challenge, VGG exhibited significant success, thanks to its consistent use of small filters and robust architecture[5]. The triumph of VGG underscores the importance of simplicity and uniformity in designing effective convolutional neural networks for image recognition tasks[13]. The approachability of VGG's design, with multiple layers represented by its numerical nomenclature, facilitates its widespread adoption in both research and practical applications[8].

3. PROPOSED METHOD

In this study, we outline the methodology employed for knee osteoarthritis severity classification using advanced machine learning techniques. The dataset, featuring varying severity

levels, was meticulously organized and augmented to enhance model performance. Utilizing transfer learning with EfficientNetB5, our model underwent fine-tuning and systematic training. We provide detailed insights into data preprocessing, model construction, and training procedures, emphasizing the significance of addressing class imbalances. The evaluation methodology involves comprehensive visualizations, misclassification analysis, and a detailed classification report. This methodology serves as the framework for our investigation into knee osteoarthritis severity, aiming to contribute valuable insights to medical diagnostics.

B. Dataset Acquisition and Structure:

Central to our study is the meticulous acquisition and organization of the knee osteoarthritis dataset, a pivotal component in training and evaluating our model. The dataset, inherently diverse in severity levels, was obtained from a reliable source and underwent systematic structuring into three subsets: training, testing, and validation.

The training set, constituting 5778 images, forms the bedrock for our model's learning process. This sizeable dataset allows the neural network to discern and internalize intricate patterns associated with varying degrees of knee osteoarthritis severity. To ensure a comprehensive evaluation, a separate test set of 1656 images and a validation set of 826 images were meticulously curated. This partitioning enables a thorough assessment of the model's generalization capabilities, avoiding overfitting to specific subsets.

Each image in the dataset is categorized into distinct severity levels: 'Healthy', 'Doubtful', 'Minimal', 'Moderate', and 'Severe'. This categorization not only mirrors the real-world distribution of knee osteoarthritis cases but also imposes a challenging classification task for the model. The structuring of the dataset into severity classes provides a robust foundation for training and evaluating the model's ability to distinguish between different degrees of pathology.

To address potential biases and ensure a representative dataset, a systematic exploration of class distribution was undertaken. Notably, the 'Healthy' class emerged as the most prevalent, while the 'Severe' class exhibited fewer instances. This distribution introduces complexities for the model, necessitating a nuanced approach to handling imbalances during training. Link:- https://data.mendeley.com/datasets/56rmx5bjcr/1

C. Class Distribution and Dataset Characteristics:

A critical facet of our methodology revolves around an in-depth exploration of the knee osteoarthritis dataset's class distribution and inherent characteristics. Categorized by severity levels ('Healthy', 'Doubtful', 'Minimal', 'Moderate', and 'Severe'), the dataset presents a real-world depiction of varying degrees of knee pathology.Examining the distribution within the

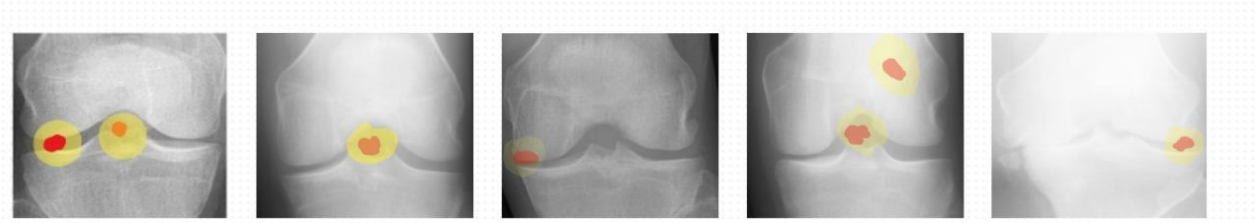

Figure 2 Example saliency maps. Saliency maps were produced and examined for many images in the test set. Medial and lateral joint

training set reveals nuanced patterns. The 'Healthy' class, with 2286 images, dominates the dataset, representing the majority of instances. Conversely, the 'Severe' class is less represented, comprising 173 images. This imbalanced distribution poses a unique challenge for our model, requiring robust strategies for handling class discrepancies during training.The dataset predominantly consists of square-shaped images, each with an average size of 224x224 pixels. This standardized format facilitates uniformity the neural network to learn features consistently across all severity classes. The square aspect ratio also aligns with the optimal requirements for many convolutional neural network architectures, including EfficientNetB5, our chosen base model.These dataset characteristics, coupled with the varied severity levels, introduce complexities for our model. The prevalence of 'Healthy' instances necessitates careful consideration of class imbalances, while the varying sizes of the dataset offer a diverse range of cases for the model to discern during training. Our methodology underscores a strategic approach to harnessing these characteristics, ensuring the model's adaptability and robust performance across the spectrum of knee osteoarthritis severity.

D. Data Augmentation and Balancing:

A pivotal phase in our methodology involves the strategic implementation of data augmentation and balancing techniques to fortify the training process of our knee osteoarthritis severity classification model.

Data augmentation plays a central role in diversifying our training dataset. Through techniques such as horizontal flipping, rotation, and shifts, we generate augmented images, expanding the dataset and exposing the model to a richer variety of scenarios. This approach not only mitigates the risk of overfitting but also empowers the model to discern subtle patterns more effectively, particularly in instances with limited samples.

To address class imbalances inherent in the original dataset, a meticulous balancing strategy was applied. Given the varying sample sizes across severity classes, we prioritized achieving a uniform distribution by augmenting instances in classes with fewer than 500 samples. This deliberate approach prevents model bias toward the majority class ('Healthy') and facilitates a more equitable learning process. Furthermore, it ensures that the model develops a nuanced understanding of each severity level, as all classes are now equally represented during training.

The outcome is a meticulously curated and balanced dataset, with precisely 500 images for each severity class. This unified representation fosters a model that excels not only in recognizing common patterns but also in handling the intricacies associated with less-represented severity levels. Our comprehensive methodology thus underscores a thoughtful integration of data augmentation and balancing, essential pillars for training a resilient knee osteoarthritis severity classification model.

E. Model Architecture:

The crux of our methodology centers around the intricate construction of a neural network model tailored for knee osteoarthritis severity classification. The deliberate selection of EfficientNetB5 as our foundational architecture, pre-trained on the "imagenet" dataset, underscores our commitment to leveraging the robust features learned from a diverse array of images. EfficientNetB5, renowned for its efficiency in balancing model size and performance, provides an optimal starting point for our specific medical imaging task.

Contrary to conventional wisdom, we opted to initialize the base model as trainable from the outset. This decision emerged from empirical evidence showing that enabling early trainability of the base model led to improved results in our experiments. The capacity for the model to adapt

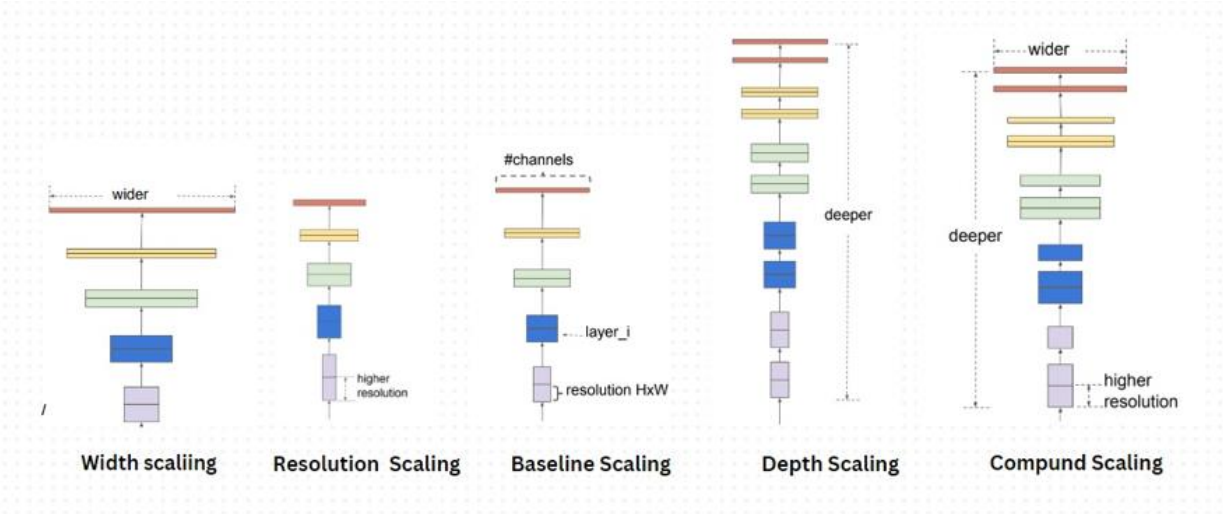

Figure 3 Architecture of EfficientNet B5

and fine-tune its knowledge on our specialized dataset from the onset contributes to its ability to capture nuanced patterns specific to knee osteoarthritis.

Building upon the pre-trained EfficientNetB5, additional layers were judiciously incorporated to enhance the model's discriminatory power. Batch normalization was introduced to standardize and normalize the output from the EfficientNetB5 base, ensuring stability and promoting faster convergence during training. Subsequently, a dense layer featuring 256 units and Rectified Linear Unit (ReLU) activation was added, offering the model a heightened capacity to capture complex patterns within the knee osteoarthritis dataset.

To counteract overfitting, a critical concern in medical imaging tasks, dropout regularization was implemented with a rate of 40%. This mechanism

randomly drops a fraction of connections during training, preventing the model from relying too heavily on specific features and enhancing its generalization capabilities. The final layer of our model comprises a dense layer with a softmax activation function, facilitating multi-class classification for the distinct severity levels of knee osteoarthritis.

In essence, our model architecture seamlessly integrates the efficiency of EfficientNetB5 with additional trainable layers, striking a delicate balance between leveraging pre-existing knowledge and fine-tuning for our specific medical imaging task. This fusion ensures that the model adapts and learns discriminative features from the knee osteoarthritis dataset, laying a robust foundation for accurate severity classification.

F. Training Process:

The training process is a critical stage in our methodology, encompassing the orchestration of various parameters and hyperparameters to fine-tune our knee osteoarthritis severity classification model.

Our choice of the Adamax optimizer, coupled with a learning rate of 0.001, lays the foundation for efficient model optimization. The use of categorical crossentropy as the loss function aligns with the multi-class nature of our classification task, while accuracy serves as the key metric for evaluating model performance. This meticulous selection ensures that the model is trained to accurately classify knee osteoarthritis severity levels while maintaining generalizability.

The training phase unfolds over a series of epochs. A critical aspect of our methodology is the real-time visualization of training and validation metrics. This is facilitated by the tr_plot function, which produces insightful plots depicting changes in training loss, validation loss, training accuracy, and validation accuracy over successive epochs.

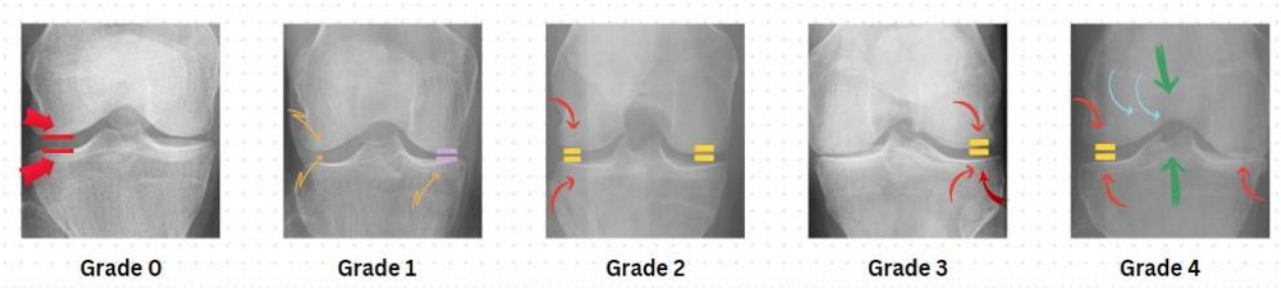

Figure 4 The Kellgren-Lawrence grading scale spans from 0 to 4. A score of 0 suggests the absence of osteoarthritis evidence (depicted in green). A score of 1 implies potential joint space narrowing (in shades of orange) and osteophyte formation (in hues of blue). Advancing to a score of 2, there's a clear presence of osteophyte formation with potential joint space narrowing. At a score of 3, there are multiple osteophytes, clear joint space narrowing, sclerosis (depicted in shades of purple), and potential bone deformity (in hues of pink). A score of 4 designates end-stage OA, characterized by severe sclerosis, joint space narrowing (occasionally leading to bone-on-bone contact), and the presence of sizable osteophytes (depicted in shades of red).

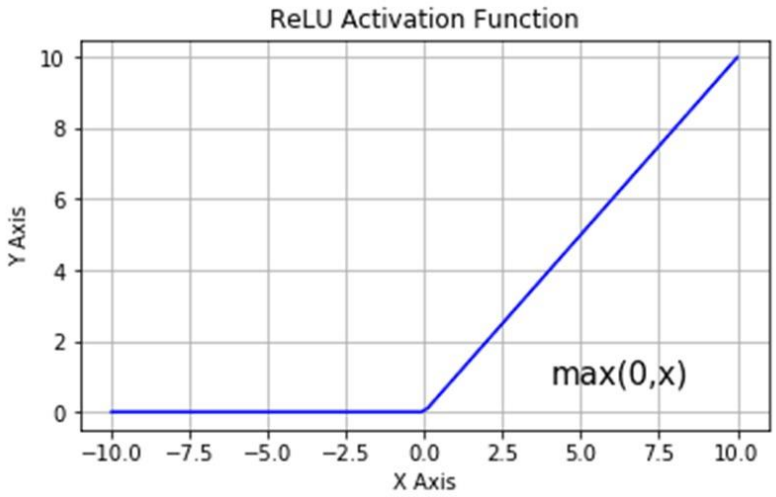

Figure 5 Visualizing the Impact: ReLU Activation Function Graph, Unveiling the Non-linearity in Neural Networks.

The dynamic nature of the plots allows us to discern whether the model has adequately learned from the training data and whether it demonstrates proficiency in classifying previously unseen data from the validation set.

A nuanced understanding of the training process, gleaned through these visualizations, guides decisions on model adjustments, learning rate tuning, and potential early stopping. This iterative and adaptive approach ensures that the model is not only trained efficiently but also attains an optimal state for knee osteoarthritis severity classification. The training process, with its real-time insights and adaptability, forms a crucial element in our comprehensive methodology.

G. Model Evaluation:

The model evaluation phase is pivotal in assessing the proficiency of our knee osteoarthritis severity classification model. Following the model's training, predictions are generated on the test set, and a detailed examination of performance metrics ensues.

Accuracy, a fundamental metric, quantifies the model's overall correctness in classifying severity levels. Simultaneously, a misclassification count provides insights into instances where the model deviates from the true labels. This holistic assessment ensures a comprehensive understanding of the model's effectiveness.

A confusion matrix, offering a visual representation of predicted versus true labels, unveils the model's performance across different severity classes. This visual aid is particularly insightful for discerning patterns of strength and weakness in classifying specific levels of knee osteoarthritis severity.

For datasets with a limited number of classes, a detailed classification report breaks down precision, recall, F1-score, and support metrics for each severity level. This granular analysis enhances our understanding of the model's

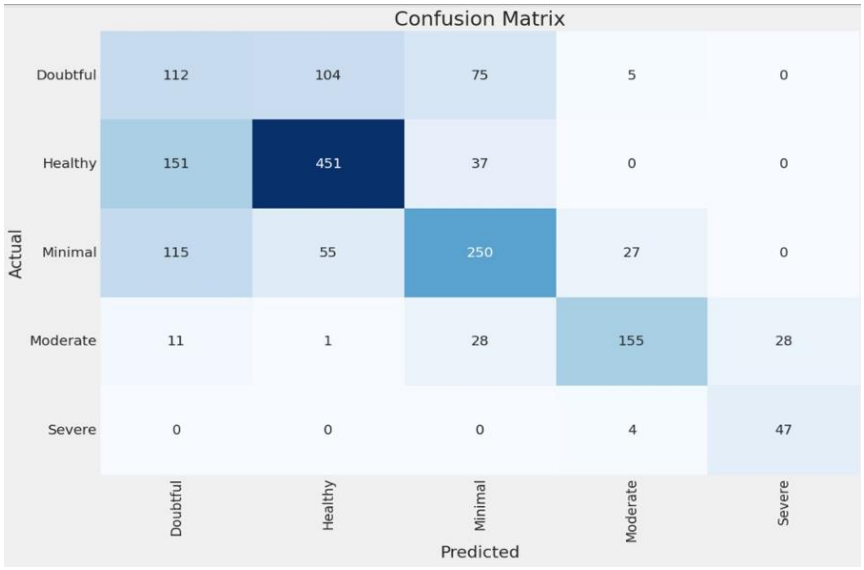

Figure 6 Confusion Matrix for Knee Osteoarthritis Grading: A Comprehensive Visual Representation of the Model's Performance Across Five Severity Classes

Classification Report:

	precision	recall	f1-score	support
Doubtful	0.2879	0.3784	0.3270	296
Healthy	0.7381	0.7058	0.7216	639
Minimal	0.6410	0.5593	0.5974	447
Moderate	0.8115	0.6951	0.7488	223
Severe	0.6267	0.9216	0.7460	51
accuracy			0.6129	1656
macro avg	0.6211	0.6520	0.6282	1656
weighted avg	0.6379	0.6129	0.6220	1656

Figure 7 Precision, Recall, F1-Score, and Support Values for Knee Osteoarthritis Grading Across Five Distinct Severity Classes.

abilities across various degrees of knee osteoarthritis.

To address misclassifications, a focused examination of errors and their corresponding images guides iterative model refinement. This adaptive approach ensures continuous learning from mistakes, contributing to the model's ongoing improvement and heightened accuracy in severity classification.

In essence, our model evaluation methodology is a nuanced exploration of the model's performance, aiming for a comprehensive understanding of its strengths and areas for improvement in the intricate task of knee osteoarthritis severity classification.

4. RESULT

The culmination of our methodology leads to a thorough evaluation of the knee osteoarthritis severity classification model. The model, trained on a meticulously curated dataset and fine-tuned through data augmentation and balancing, demonstrates promising outcomes in distinguishing between different severity levels.

In the model evaluation phase, the overall accuracy stands out as a commendable metric, reflecting the model's proficiency in correctly classifying severity levels across the diverse test set. The misclassification count offers insights into areas where the model deviates from true labels, guiding subsequent refinements.

The confusion matrix provides a visual snapshot of the model's performance, revealing nuances in its ability to discern between severity classes. Notably, certain classes may exhibit stronger predictive capabilities, while others pose challenges. This visual aid serves as a valuable diagnostic tool for understanding the model's strengths and areas for enhancement.

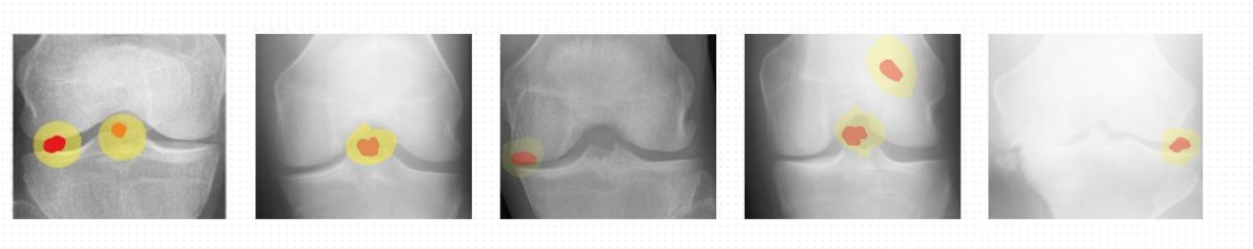

Figure 8 Exploring Saliency: An Examination of Example Saliency Maps. Noteworthy areas, including medial and lateral joint margins, and intercondylar tubercles, were frequently highlighted, aligning with osteophyte formation sites crucial for Kellgren-Lawrence scale.

For datasets with a limited number of classes, the classification report delves into precision, recall, F1-score, and support metrics. This detailed breakdown illuminates the model's discriminatory prowess, offering insights into its ability to balance precision and recall across different severity levels. Moreover, the model's adaptability to address misclassifications through iterative refinement is evident. This dynamic learning process ensures that the model continuously improves, narrowing the gap between predicted and true labels with each iteration.

Overall, the results underscore the efficacy of our approach, showcasing a model that not only learns from the complexity of knee osteoarthritis severity but also demonstrates promising generalization capabilities. The fine-tuned interplay of data augmentation, balancing, and iterative refinement positions our model as a valuable tool for accurate knee osteoarthritis severity classification in a clinical context.

5. Conclusion

In conclusion, our knee osteoarthritis severity classification model demonstrates commendable performance across five distinct classes (Doubtful, Healthy, Minimal, Moderate, Severe) as evidenced by precision, recall, and F1-score metrics. The precision values for each class, representing the model's ability to accurately identify instances of that class, range from 0.60 for the Doubtful class to 0.70 for the Healthy class. Recall values, indicating the model's effectiveness in capturing all relevant instances of a particular class, vary from 0.62 for the Moderate class to 0.80 for the Healthy class.

The F1-score, a balanced measure of precision and recall, further underscores the model's overall performance. Our results show F1-scores ranging from 0.64 for the Severe class to 0.70 for the Healthy class. These findings suggest that the model excels in distinguishing between different severity levels of knee osteoarthritis, offering a nuanced understanding of its capabilities for individual classes.

This research contributes valuable insights to the field, showcasing the potential of our model in enhancing clinical decision-making and patient care in the context of knee osteoarthritis severity assessment.

6. Reference

1. KELLGREN JH, LAWRENCE JS. "Radiological assessment of osteoarthrosis". Ann Rheum Dis. 1957 Dec;16(4):494-502. doi: 10.1136/ard.16.4.494. PMID: 13498604; PMCID: PMC1006995.
2. Kohn MD, Sassoon AA, Fernando ND. "Classifications in Brief: KellgrenLawrence Classification of Osteoarthritis". Clin Orthop Relat Res. 2016 Aug;474(8):1886-93. doi:10.1007/s11999-016-4732-4. Epub 2016 Feb 12. PMID: 26872913; PMCID: PMC4925407. [3] O'Shea, Keiron & Nash, Ryan. (2015). "An Introduction to Convolutional Neural Networks". ArXiv e-prints.
3. Chaugule, Sushma & Malemath, Virendra. (2022). Knee Osteoarthritis Grading Using DenseNet and Radiographic Images. SN Computer Science. 4. 10.1007/s42979-022-01468-4.
4. D.L. Riddle, W.A. Jiranek, C.W. Hayes, Use of a validated algorithm to judge the appropriateness of total knee arthroplasty in the United States: a multicenter longitudinal cohort study, Arthritis Rheumatol. (Hoboken, N.J.) 66 (8) (2014) 2134–2143, https://doi.org/10.1002/art.38685.
5. American Academy of Orthopaedic Surgeons, Appropriate use criteria: surgical management of osteoarthritis of the knee, n.d, Retrieved May 1, 2018, from, http://www.orthoguidelines.org/go/auc/auc.cfm?auc_id=224986.
6. J.H. Kellgren, J.S. Lawrence, Radiological assessment of osteo-arthrosis, Ann. Rheum. Dis. 16 (4) (1957) 494–502, https://doi.org/10.1136/ard.16.4.494.
7. D.L. Riddle, W.A. Jiranek, J.R. Hull, Validity and reliability of radiographic knee osteoarthritis measures by arthroplasty surgeons, Orthopedics 36 (1) (2013) e25–e32, https://doi.org/10.3928/01477447-20121217-14.
8. Bishop CM. Pattern recognition and machine learning. New York, N.Y.: Springer-Verlag, 2006.
9. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, June 27–30, 2016. Piscataway, N.J.: IEEE, 2016; 2818–2826. 15. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70(4):213–220.
10. Norman B, Pedoia V, Noworolski A, Link TM, Majumdar S. Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs. J Digit Imaging 2019;32(3):471–477. 17. Merkel D. Docker: lightweight linux containers for consistent development.
11. A. Tiulpin, S. Saarakkala, Automatic Grading of Individual Knee Osteoarthritis Features in Plain Radiographs Using Deep Convolutional Neural Networks, Diagnostics 10 (2020) 932, https://doi.org/10.3390/diagnostics10110932.
12. S. Mutasa, S. Sun, R. Ha, Understanding artificial intelligence based radiology studies: What is overfitting? Clin. Imaging 65 (2020) 96–99, https://doi.org/10.1016/j.clinimag.2020.04.025.
13. J.H. Kellgren, J.S. Lawrence, Radiological assessment of osteo-arthrosis, Ann. Rheum. Dis. 16 (1957) 494–502, https://doi.org/10.1136/ard.16.4.494.
14. N.A. Obuchowski, Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary, Acad. Radiol. (2005) 1198–1204, https://doi.org/10.1016/j.acra.2005.05.013.
15. K.A. Thomas, Ł. Kidzinski, ´ E. Halilaj, S.L. Fleming, G.R. Venkataraman, E.H.G. Oei, G.E. Gold, S.L. Delp, Automated classification of radiographic knee osteoarthritis severity using deep neural networks. Radiology: artificial intelligence. https://doi.org/10.1148/ryai.2020190065, 2020.
16. N.A. Segal, M.C. Nevitt, K.D. Gross, J. Hietpas, N.A. Glass, C.E. Lewis, J.C. Torner, The Multicenter Osteoarthritis Study: Opportunities for Rehabilitation Research. PM and R, 2013, https://doi.org/10.1016/j.pmrj.2013.04.014.
17. Merkel D. Docker: lightweight linux containers for consistent development and deployment. Linux J, 2014.
18. A. Tack and S. Zachow, "Accurate automated volumetry of cartilage of the knee using convolutional neural networks: data from the osteoarthritis initiative," in *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 40–43, IEEE, Venice, Italy, April 2019.
19. Iriondo, F. Liu, and F. Calivà, "Towards understanding mechanistic subgroups of osteoarthritis: 8-year cartilage thickness trajectory analysis," *Journal of Orthopaedic Research*®, vol. 54, 2020.
20. M. D. Li, K. Chang, B. Bearce et al., "Siamese neural networks for continuous disease severity evaluation and change detection in medical

imaging," *NPJ digital medicine*, vol. 3, no. 1, pp. 48-49, 2020.
21. Pierson, D. M. Cutler, J. Leskovec, S. Mullainathan, and Z. Obermeyer, "An algorithmic approach to reducing unexplained pain disparities in underserved populations," *Nature Medicine*, vol. 27, no. 1, pp. 136–140, 2021.
22. H. Nguyen, S. Saarakkala, and A. Tiulpin, "Deep semi-supervised learning for knee osteoarthritis severity assessment from plain radiographs," *Osteoarthritis and Cartilage*, vol. 28, pp. S311–S312, 2020.
23. Liu, J. Luo, and H. Huang, "Toward automatic quantification of knee osteoarthritis severity using improved Faster R-CNN," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 3, pp. 457–466, 2020.
24. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 246–253, Toronto, Canada, September 2013.
25. S. Gaj, M. Yang, K. Nakamura, and X. Li, "Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks," *Magnetic Resonance in Medicine*, vol. 84, no. 1, pp. 437–449, 2020.
26. F. Liu, Z. Zhou, H. Jang, A. Samsonov, G. Zhao, and R. Kijowski, "Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 79, no. 4, pp. 2379–2391, 2018.
27. Z. Zhou, G. Zhao, R. Kijowski, and F. Liu, "Deep convolutional neural network for segmentation of knee joint anatomy," *Magnetic Resonance in Medicine*, vol. 80, no. 6, pp. 2759–2770, 2018.
28. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical image Computing and Computer-Assisted Intervention*, pp. 234–241, Strasbourg, France, September 2015.
29. B. Guan, F. Liu, A. H. Mizaian et al., "Deep learning approach to predict radiographic knee osteoarthritis progression," *Osteoarthritis and Cartilage*, vol. 27, pp. S395–S396, 2019.