INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Check for updates

OPEN ACCESS

Received: 03/08/2026

Accepted: 31/08/2026

Published: 09/09/2026

Citation: Bruntha S, Dr. M. Chidambaram (2026) Fast Severity Ordered Boundary Residual Learning for Multi-Class Knee Osteoarthritis Grading from X-Ray Images. Indian Journal of Science and Technology 19(32): 2305-2315. https://doi.org/10.17485/IJST/v19i32.1137

* **Corresponding author**.

brunthaloga@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2026 Bruntha & Dr. M. Chidambaram. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Fast Severity Ordered Boundary Residual Learning for Multi-Class Knee Osteoarthritis Grading from X-Ray Images

S. Bruntha[1]*, **Dr. M. Chidambaram**[2]

1 Research Scholar, Department of Computer Science, Rajah Serfoji Government College (Autonomous), Thanjavur-613005, Affiliated to Bharathidasan University, Tiruchirappalli - 620 024

2 Research Supervisor & Associate Professor, Department of Computer Science, Rajah Serfoji Government College (Autonomous), Thanjavur-613005, Affiliated to Bharathidasan University, Tiruchirappalli - 620 024

Abstract

Objectives: To develop a model for the classification of knee OsteoArthritis (OA) on radiographs, which is computationally efficient and severity ordered. **Method:** This study focuses on class imbalance, adjacent-grade confusion and high computational cost of multi-model hybrid systems. The proposed SOBR-KneeNet (Severity Ordered Boundary Residual Learning) includes a pretrained ConvNeXt-Tiny encoder and three heads for supervised learning: classification, cumulative-ordinal and residual-estimation. Severity structure in training is provided by auxiliary ordinal and residual objectives and the final inference distribution through validation calibration. In experiments, 8,260 Osteoarthritis Initiative knee radiographs are used, where the 5,778, 826 and 1,656 are used as training, validation and test partitions, respectively. The training protocol consists of the AdamW optimization, progressive backbone unfreezing, mixed precision, cosine learning-rate decay and horizontal-flip test-time augmentation. **Findings:** The test accuracy of SOBR-KneeNet was 68.90%, balanced accuracy was 71.20%, macro F1-score was 70.38%, weighted F1-score was 68.86%, grade mean absolute error was 0.3904, quadratic weighted kappa was 0.8450 and within one grade accuracy was 95.59%. The accuracy was 2.11 % higher than existing study, 2.59% higher than DenseNet-121 and 3.54 % higher than ResNet-50 V2. The percentage with grade-4 recall was 90.20% and all errors were within two grades of the correct level. **Novelty:** Severity information is fed into optimization directly, without the need for a transformer branch, multi-model ensemble or iterative post-processing, through fixed cumulative boundaries. Validation-selected inference retained the calibrated 5-class head, with the ordinal and residual branches serving as low-cost training regularizers. The design offers one-backbone solution that has 28.63 million parameters and a fast epoch execution time.

Keywords: knee osteoarthritis; Kellgren-Lawrence grading; ordinal learning; ConvNeXt; CLAHE; multi-task learning

1 Introduction

Knee osteoarthritis (KOA) severity assessment from radiographs remains an important task for treatment planning and disease monitoring. However, automated Kellgren-Lawrence (KL) grading of radiographs provides a repeatable alternative to manual assessment, but there are challenges with reliable discrimination between adjacent severity grades. The deep learning-based radiographic analysis has shown significant potential for assessing KOA [1]. A Systematic Review and Meta analysis revealed that advanced KL grades showed better diagnostic performance than early grades and this was consistently reported to be difficult around subtle structural changes [2]. Reader-level agreement has also been shown to be good for external-dataset evaluation, supporting automated grading, as well as sensitivity to dataset variation [3]. Hence, the impact of clinically meaningful KOA classification is not limited to simply the amount of correct classifications, but must also consider ordinal consistency, class-specific behavior, generalization and computational feasibility.

One of the most important strategies for radiographic KOA analysis is transfer learning, which can help to reduce training requirements and maintain discriminative image features. The results of the KOA classification are encouraging with transfer based feature engineering, deep neural networks, advanced detection architectures and pretrained convolutional models [4–8]. Kinger [9] proved the feasibility of deep architectures for automatic severity grading and Rani et al. [10] presented the severity grading of radiographic KOA using CNN. Together, these results support the use of convolutional feature representations, yet the typical categorical training uses KL grades as unordered classes due to the ordered clinical progression from Grade 0 to Grade 4 across the grades.

In recent years, attention mechanisms, multi-scale processing, ensembles and deeper feature integration have been among the research interests in architecture. Representation learning over multiple scales has achieved higher representation quality for severity-sensitive tasks [11] and multi-network ensembles have shown robust classification performance on OAI-based radiographs [12]. Further examples of the usefulness of advanced feature extraction techniques are found in inception-based transfer learning [13], in comparative deep architectures [14] and in multi-scale attention frameworks [15]. Multi-network ensembles, transformer branches, multi-resolution processing and stacked attention modules, however, increment the complexity of the model and computational demands. Thus there is a practical need for a single-backbone architecture that can maintain severity ordering without relying on multiple independently trained networks.

The use of ordinal learning as the base for KL grading is appropriate since adjacent grade errors are of different clinical meaning than large grade displacement errors. Ordinal loss formulations have shown the benefits of severity-driven optimization [16] and automated DNN grading has also shown the feasibility of structured radiographic classification [17]. Stable deep representation optimization by residual learning [18] and evidence of severity-related structural information by OAI-based radiographic predictor modelling [19] justify deep representation. Longitudinal gait-based prediction studies are also indicative of more general interest in structured KOA severity estimation [20]. In recent years, the research of KOA grading has been extended to the field of transfer learning, deep network optimization, multimodal transformer and traditional-deep feature fusion [21–26]. However, there is still a practical research gap in the use of a computationally compact single-backbone radiographic grading with explicit supervision to determine exact class identity, cumulative severity boundaries and grade-distance error. To fill this gap, SOBR-KneeNet proposes a shared ConvNeXt-Tiny encoder with categorical, cumulative-ordinal and residual-grade supervision. The main contribution is the severity-ordered representation learning in the optimization without the need for fusion of multiple transformer models, multiple model inference, or iterative post-processing; and the evaluation using accuracy, balanced accuracy, macro F1 score, grade mean absolute error, quadratic weighted kappa and within-one-grade accuracy.

2 Methodology

SOBR-KneeNet develops established methodological principles into a single severity-aware training framework rather than modifying a single predecessor architecture. Transfer-learning based KOA classification serves as the foundation for pretrained classification of radiographic features [4,9] and ordinal severity modelling follows the principle of explicitly penalizing grade displacement instead of assuming a uniform penalty for each class error [16]. In addition to stable representation optimization and continuous grade-error modelling, another basis of residual learning is provided [18]. Attention-based, multi-scale and ensemble architectures [11,12,15] provide important performance benchmarks, but impose extra branches, resolutions, attention modules and/or independently optimized networks. Instead, SOBR-KneeNet simultaneously shares a ConvNeXt-Tiny representation to the categorical goal, cumulative-boundary goal and residual-grade goal.

2.1 Dataset and severity labels

The knee osteoarthritis severity dataset consists of 8,260 fixed-flexion knee radiographs from the Osteoarthritis Initiative (OAI) that are publicly available. Predefined partitions: There are 5,778 training images, 826 validation images and 1,656 test images. Training-class counts show a 13.21:1 ratio between grade 0 and grade 4. The training distribution and split totals are summarized in Table 1. No test sample was used in optimization or calibration.

Table 1. Dataset composition across KL grades

KL grade	Clinical interpretation	Training images	Training share (%)
0	No radiographic OA	2,286	39.57
1	Doubtful OA	1,046	18.10
2	Mild OA	1,516	26.24
3	Moderate OA	757	13.10
4	Severe OA	173	2.99
Split totals	Train / validation / test	5,778 / 826 / 1,656	Dataset total: 8,260

2.2 Radiograph preprocessing and augmentation

All the X-rays are processed through grayscale-preserving contrast-limited adaptive histogram equalization (CLAHE), resized to 256 × 256 pixels and then normalized with pretrained-backbone statistics. CLAHE reduces intensity saturation all over the image while enhancing the contrast in the local regions showing the boundaries between the joint spaces. Training augmentation consists of horizontal reflection and limited geometric or photo-metrical perturbation. The validation and test transformations are still deterministic, except for horizontal-flip test-time augmentation for final inference. Effective-number weighting mitigates the problem of class imbalance without creating a static synthetic data set of 10K per grade.

$$w_k = \min\left(w_m ax, [(1-\beta)/(1-\beta^{n_k})]^p / Z\right) \tag{1}$$

In Equation (1), n_k represents the number of samples in a certain grade k, β is 0.999, p is 0.25, $w_{\max}$ is 1.8 and Z is a normalization of the class weights. The formula penalizes over-representation of minorities while giving more weight to less-represented grades.

2.3 SOBR-KneeNet architecture

The proposed architecture is shown in Figure 1. A CLAHE enhanced radiograph is fed into a ConvNeXt-Tiny model that was pretrained on ImageNet-22K and fine-tuned on ImageNet-1K. The encoder is composed of a convolutional stem and four hierarchical stages, each of which features depthwise convolution, layer normalization, pointwise expansion, GELU activation and pointwise projection. Global average pooling reduces the last feature map to a shared embedding. The shared embedding is communicated to three lightweight output heads: a five-logit categorical head, a four-logit cumulative ordinal head and a scalar residual-grade head. It has 28,628,106 trainable parameters.

2.4 Categorical, boundary and residual supervision

Each categorical head performs softmax normalization of its logits z_k to obtain class probabilities:

$$p_k = \exp(z_k)/\Sigma_{j=0}^{4}\exp(z_j) \tag{2}$$

Direct 5-class supervision is provided by weighted cross-entropy with label smoothing parameter ε = 0.02. For soft ordinal targets, 35% of target mass is assigned based on Gaussian distribution with σ = 0.65; the neighborhood of the reference grade is ordered. Equation (3) shows the categorical objective.

$$L_{cls} = -\Sigma_{k=0}^{4} w_k q_k \log(p_k) \tag{3}$$

The four cumulative targets are used to indicate if the severity is above each boundary $b \in 0,1,2,3$. Maximizes the ordinal logits o_b:

$$L_{ord} = -\Sigma_{b=0}^{3}[t_b log\sigma(o_b) + (1-t_b)\log(1-\sigma(o_b))] \tag{4}$$

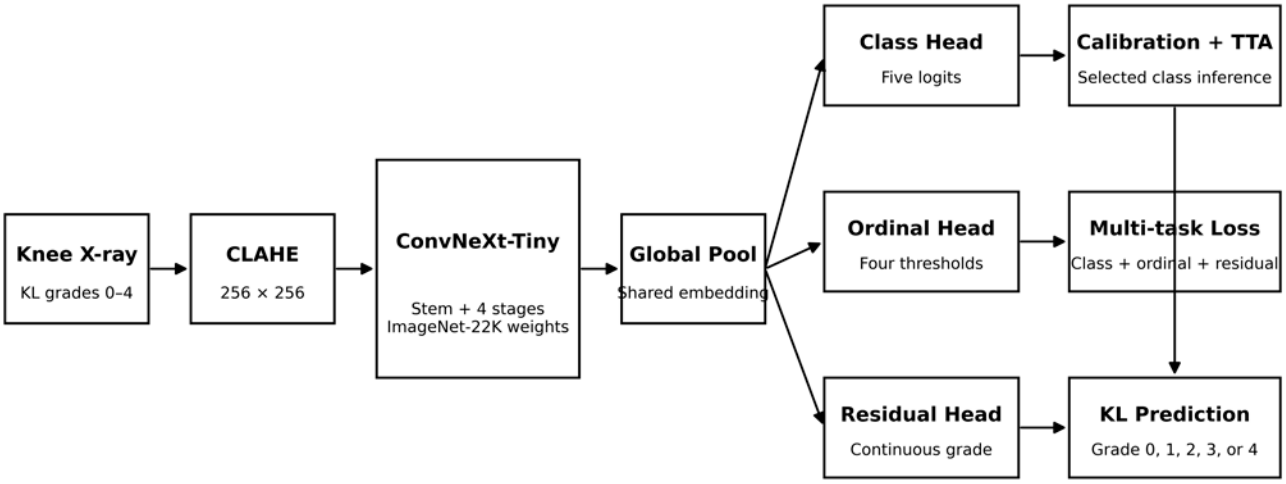


Fig 1. Black-and-white architecture of the proposed SOBR-KneeNet

If it is a continuous expected grade, then there is a residual target. A robust absolute deviation restricts the scalar regression estimate r to the integer KL grade y:

$$L_{reg} = |r - y| \tag{5}$$

Monotonicity regularization discourages violations in the cumulative probabilities as the probability that the true probability exceeds a higher boundary should not be greater than the probability that the true probability exceeds a lower boundary:

$$L_{mono} = \Sigma_{b=0}^{2} \max\left[0, \sigma\left(o_{b+1}\right) - \sigma\left(o_b\right)\right] \tag{6}$$

The final training goal is a combination of the four parts, together holding fixed coefficients:

$$L = 1.00L_{cls} + 0.30L_{ord} + 0.05L_{reg} + 0.01L_{mono} \tag{7}$$

Fixed coefficients ensure fast computation and reproducibility. Convolutional feature extraction is shared by all objectives while the ordinal and residual branches are just small linear projections.

2.5 Calibration and Inference

Validation calibration involves assessing a class-ordinal mixing coefficient α, a class temperature T_c and an ordinal temperature T_o. The grade probabilities derived from categorical probabilities are a convex combination. A horizontal-flip test time augmentation is used for averaging predictions with the original and reflected image. The validation search selected $\alpha = 1.0, T_c = 0.75\ and T_o = 0.75$. So, the final test prediction is done using calibrated categorical probabilities and the boundary and residual branches are used as auxiliary training regularizers.

$$p_{final} = \alpha p_{cls}\left(T_c\right) + (1-\alpha) p_{ord}\left(T_o\right); = argmax_k p_{final}, k \tag{8}$$

An unsupported inference-fusion claim is avoided by Equation (8). In the reported experiment, α = 1.0 makes Equation (8) equivalent to calibrated class-head inference. Shared backbone representations are still influenced by auxiliary losses in optimization.

2.6 Experimental settings and hyperparameter configuration

All experiments were run on an NVIDIA CUDA enabled GPU using PyTorch 2.12.1 and CUDA 12.6, CUdNN 9.1.0. Compared to the standard single precision, automatic mixed precision decreased memory traffic. AdamW set the learning rate for the backbone and the head separately and decayed them with cosine decay. The backbone was frozen for three epochs and then was progressively fine-tuned. Validation macro F1-score was used in early stopping (patience 15). The best checkpoint was at epoch 33 and training ended at epoch 48. Full reproducible configuration is given in Table 2.

Table 2. Experimental settings and hyperparameter configuration

Component	Configuration	Component	Configuration
Backbone	ConvNeXt-Tiny	Pretraining	IN-22K → IN-1K
Input size	256 × 256	Parameters	28,628,106
Batch size	16	Maximum epochs	60
Completed epochs	48	Best epoch	33
Optimizer	AdamW	Head learning rate	$2 \times 10^{-}$
Backbone LR	1×10^{-5}	Weight decay	$1 \times 10^{-}$
Scheduler	Cosine decay	Frozen epochs	3
Dropout	0.30	Drop-path rate	0.10
Label smoothing	0.02	Soft ordinal ratio	0.35
Ordinal σ	0.65	Class-weight β	0.999
Early-stop patience	15	Mixed precision	Enabled
Test augmentation	Horizontal flip	Random seed	42

2.7 Evaluation measures

Accuracy indicates correctness in terms of exact grade. Balanced accuracy is the average of the recall scores at five grades. Macro F1-score is equal to every grade's F1-score and weighted F1-score considers test support. The mean absolute error is an average of the ordinal distance. Quadratic weighted kappa gives more penalty to opposing views that are further apart. The proportion satisfying $|-y| \leq 1$ is called the within one grade accuracy. Combined reporting ensures that there is no reliance on a single measure that is sensitive to imbalance.

3 Results and Discussion

3.1 Overall and class-wise performance

SOBR-KneeNet was able to make 1141 out of the 1656 test radiographs correct, giving an accuracy rate of 68.90%. The accuracy of the balanced accuracy was 71.20%, the macro F1-score was 70.38% and the quadratic weighted kappa was 0.8450. The results of Grade MAE: 0.3904 and within-one-grade accuracy: 95.59% indicate a good level of severity consistency. Class-wise values and aggregate values are given in Table 3. The best recall is obtained in the grade 4 (90.20%) and the highest F1-score is obtained in the grade 4 (87.62%). The strongest class is Grade 0, with an F1-score of 61.50%, while followed by Grade 2 (55.71%) and Grade 1 (36.85%) classes, which indicate that it is the hardest to distinguish from these two classes visually.

Table 3. Test performance of SOBR-KneeNet

Level	Precision (%)	Recall (%)	F1-score (%)	Support
Grade 0	79.15	79.03	79.09	639
Grade 1	36.55	37.16	36.85	296
Grade 2	69.03	65.32	67.13	447
Grade 3	78.33	84.30	81.21	223
Grade 4	85.19	90.20	87.62	51
Macro average	69.65	71.20	70.38	1,656
Weighted average	68.88	68.90	68.86	1,656
Ordinal measures	MAE = 0.3904	QWK = 0.8450	Within ± 1 = 95.59%	—

Figure 2 shows representative grade-0 predictions and an error of grade-0 to grade-1. Most of the displayed normal knees have an expected severity score which is continuous from below the first boundary to near it. The wrong sample gets score 0.88, indicating a lack of confidence on the edge of grade-0 and grade-1.

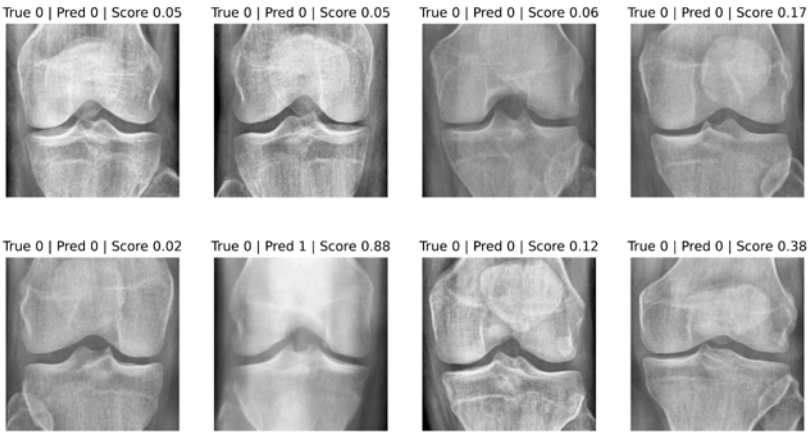


Fig 2. SOBR-KneeNet prediction with continuous severity scores

3.2 Confusion matrix and ordinal behavior

The three dimensional confusion matrix is shown in Figure 3. Diagonal counts include 505 grade-0, 110 grade-1, 292 grade-2, 188 grade-3 and 46 grade-4 radiographs. There is a total of 515 errors with adjacent grade mistakes making up 442 of those errors. There are an additional 73 errors that cover two grades and none that cover 3 or 4 grades. Grade-1 errors tend to be corrected to grade 0 (102 cases) or grade 2 (79 cases). There are only five grade-3 predictions with grade-4 confusion. The near diagonal pattern accounts for the high quadratic weighted kappa and within one grade accuracy.

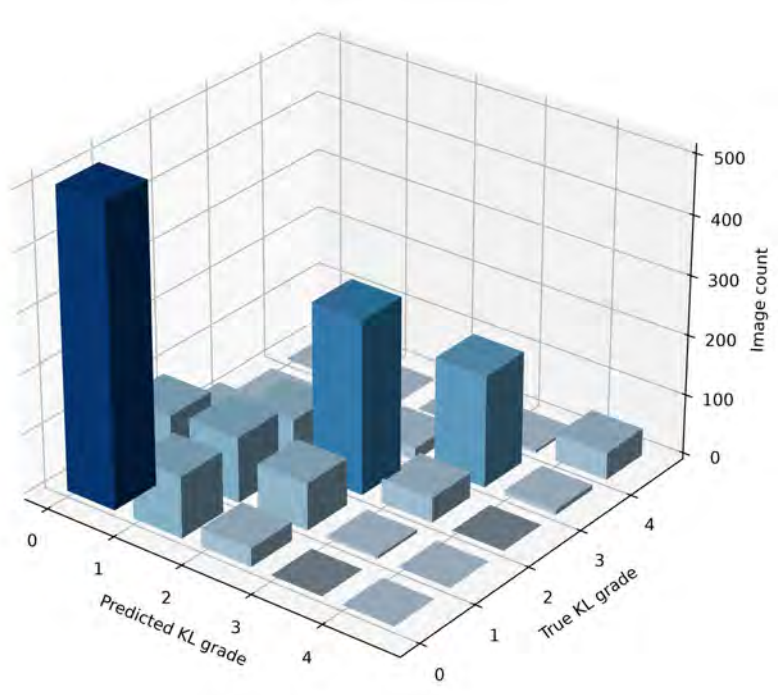


Fig 3. Three-dimensional confusion matrix for five KL grades

Figure 4 illustrates the precision, recall and F1-scores for all the classes. The surface shows good performance for grades 0, 3 and 4 and a noticeable grade-1 depression. The pattern is exacerbated by clinical ambiguity and training support that is less than grade-1. Future gains of grade 1 will be about improved local joint space representation or stronger boundary specific sampling, not on training alone.

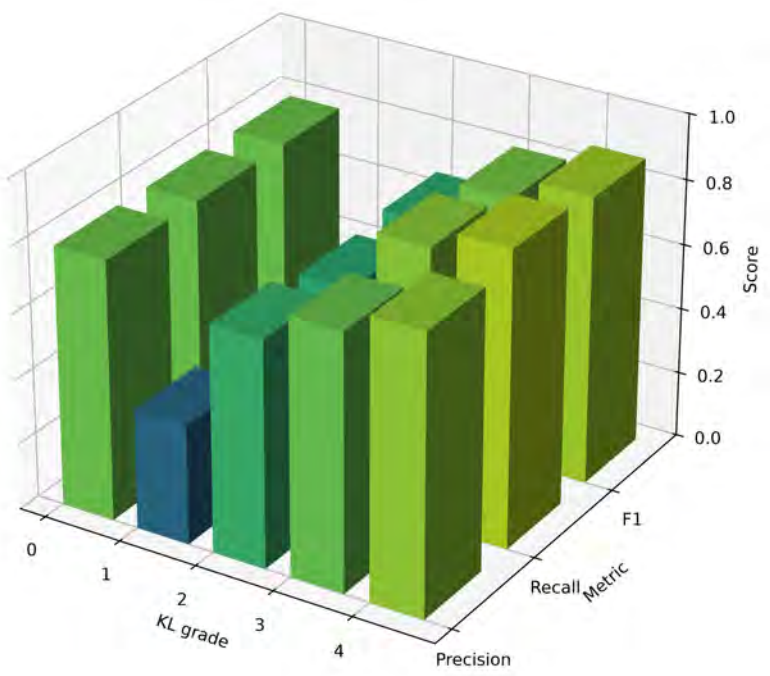


Fig 4. Three-dimensional class-wise precision, recall and F1-score

Figure 5 shows a mapping of the continuous grades to the reference KL grades. Probability density is focused in sequences of ordered diagonal regions, enabling learning of severity-aware representation. The confusion matrix is followed by wider spread of grades 1 and 2. The clear separation of normal and severe radiographic appearances is suggested by stronger terminal peaks at grades 0 and 4.

3.3 Training behavior

The loss on the training data and the macro F1 score on the validation data are plotted in Figure 6 for 48 epochs completed. Training loss declined from 1.4356 to 0.6503. The F1 score obtained by the validation macro increased sharply in the first 12 epochs and peaked at the checkpoint value of 0.6928 at epoch 33. Subsequent validation fluctuations remained small while training loss continued to decrease, indicating stable convergence after the peak validation checkpoint. To avoid selecting sub-optimal later checkpoints, early stopping was used, which kept the weights at epoch 33. In the recorded environment, the backbone unfreezing caused mean value of the epoch duration to stay around 13.3 seconds.

3.4 Comparative study

Calculations were performed to determine the difference between SOBR-KneeNet and existing models that have been evaluated on the same 8,260 image OAI derived dataset and fixed 1,656 image test split as shown in Table 4. The accuracy of SOBR-KneeNet exceeded ResNet-50 V2 by 3.54 percentage points, DenseNet-121 by 2.59 percentage points and ResNet34-ViT hybrid by 2.11 percentage points. Macro F1-score improved by 1.71 points from the ResNet34-ViT class average reported F1-scores. Grade-4 recall increased from 88.24% to 90.20%. The optimized 8-model ensemble reported by Pi et al.[12] achieved the accuracy of 76.93% and F1-score of 76.65%, outperforming the accuracy of the single model SOBR-KneeNet. Pi et al. need eight optimized independently backbones and several input resolutions. The computational design of SOBR-KneeNet is extremely simple with only 28.63 million-parameter backbone and 256-pixel input stream. A direct comparison of runtimes is not yet suitable, because of the variation in hardware, cross validation and model selection methods.

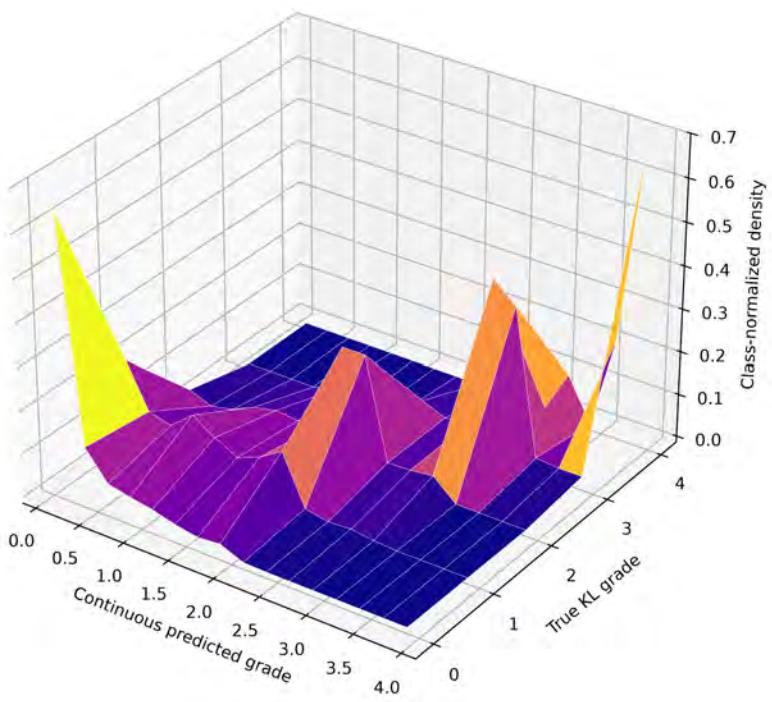


Fig 5. Ordinal prediction surface across continuous and reference grades

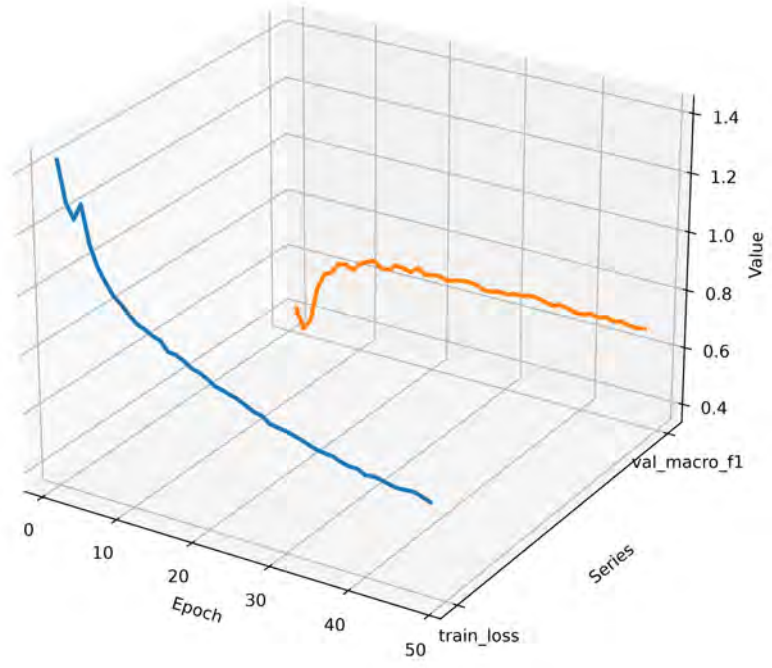


Fig 6. Three-dimensional training-loss and validation macro-F1 trajectory

Table 4. Comparative performance against existing KOA grading models

Study/model	Architecture	Accuracy (%)	F1-score (%)	Computational form
Baseline	ResNet-50 V2	65.36	—	Single CNN
	DenseNet-121	66.31	—	Single CNN
	ResNet34-ViT	66.79	68.67*	CNN-transformer hybrid
Pi et al.[12]	Eight-model ensemble	76.93	76.65	Eight optimized CNNs
Proposed SOBR-KneeNet	ConvNeXt-Tiny + auxiliary heads	68.90	70.38	Single CNN backbone

*Macro F1-score calculated as the unweighted mean of five class F1-scores reported for ResNet34-ViT. Comparisons were on reported values – Pi et al.[12] used five-fold model development and architecture-specific image sizes.

3.5 Discussion

SOBR-KneeNet achieved 68.90% exact-grade accuracy, 71.20% balanced accuracy, 70.38% macro F1-score, 0.3904 grade MAE, 0.8450 quadratic weighted kappa and 95.59% within-one-grade accuracy. The combination of categorical supervision, cumulative severity boundaries, residual-grade estimation and monotonicity regularization provides a plausible explanation for the strong ordinal consistency. Under the fixed comparison setting, DenseNet-121 and ResNet-50 V2 achieved accuracy improvements of 2.59 and 2.11 percentage points, respectively, when compared with ResNet34-ViT and an improvement of 3.54 percentage points. These gains are for severity-aware optimization, rather than inference-time fusion. For SOBR-KneeNet, Pi et al.[12] reported higher accuracy of 76.93% with an optimized eight network ensemble, which is an important accuracy-complexity trade-off rather than the absolute state-of-the-art superiority. The single backbone design with 28.63 million parameters is then a computationally simpler model for the ordered five-class grading.

The main constraint is still Grade 1 (F1-score 36.85%). This is consistent with the overall literature not an isolated failure. Pooled sensitivities have been reported as 64.00% for KL Grade 1 and 90.32% for KL Grade 4, making the early detection of the subtle radiographic changes significantly more challenging[2]. Huang et al.[23] also found that Grade 1 is the most challenging class in an externally validated multimodal system with significantly more complex models and clinical covariates. The SOBR-KneeNet confusion pattern then still shows the same pattern of errors around the same grade levels, resulting in clinically consistent difficulty as expected for KL grades. Avoid claims of superior early stage detection, better boundary-focused sampling, better localized joint space representation and higher resolution processing of regions of interest are more appropriate path forward.

In recent literature, much higher reported performance under different experimental conditions is found. Bukhari and Altaf[22] achieved 97% accuracy using transfer learning with EfficientNetB3 and Yilmaz and Solak[25] achieved 85.39% accuracy using both traditional and deep features with joint fusion. By combining radiographs, clinical covariates, cross-modal attention and uncertainty-aware gating, Huang et al.[23] reached a balanced accuracy of 87.6% and a QWK of 0.900 on an independent, external cohort. It is inappropriate to compare the results of such studies directly, since the evaluation partitions, modalities, preprocessing techniques, validation procedures and computational requirements are all different. SOBR-KneeNet therefore contributes primarily through compact severity-structured optimization: one 256 × 256 input stream, one ConvNeXt-Tiny backbone, lightweight auxiliary heads during training and calibrated categorical inference. Some practical use can be screening support with radiographs, severity pre-grading, assisting second readers and prioritizing cases for specialist review. To widely use this in clinical practice, independent multicenter validation, prospective clinical evaluation, probability-calibration analysis and targeted improvement of the discrimination for Grade 1 are required.

4 Conclusion

SOBR-KneeNet addresses the objective of computationally efficient, severity-ordered five-class knee osteoarthritis grading from radiographs. The main novelty is in the severity-aware optimization by a single ConvNeXt-Tiny backbone shared by the categorical, cumulative-boundary and residual-grade objectives. The auxiliary severity branches are not used as multi-model ensembles or transformer-fusion architectures: They are lightweight training regularizers and the final inference is done using calibrated categorical probabilities. This design is able to maintain explicit KL order information without bringing the complexity of multi-network inference.

The experimental evaluation results are 68.90% accuracy, 71.20% balanced accuracy, 70.38% macro F1 score, 0.3904 grade MAE, 0.8450 quadratic weighted kappa and 95.59% within-one-grade accuracy. The reported comparison setting shows the benefit of severity-structured training with reported accuracy improvements of 2.11%, 2.59% and 3.54% over the selected

baseline networks ResNet34-ViT, DenseNet-121 and ResNet-50 V2. The results showed that grade 4 recall was 90.20% while grade 1 was the most restricting with a F1 score of 36.85%.

Practical applications may involve automated pre-grading of radiographs, second reader support and prioritizing worklists and making decisions based on the severity of the images in clinical settings without a large number of specialists. SOBR-KneeNet should be used as an assistive grading system and not as a stand-alone system for diagnosis. Independent multicenter validation, higher resolution joint space regions, boundary-focused sampling Grade 1, assessment of calibration and prospective clinical workflow evaluation should be the focus of future research. Such extensions can provide generalizability, reliability, clinical usefulness beyond the current fixed radiographic benchmark.

5 Acknowledgements

NIL

Author details: S. Bruntha is a Research Scholar; Dr. M. Chidambaram is a Research Supervisor & Associate Professor.

Data Availability: The knee osteoarthritis severity dataset is publicly available through Kaggle: https://www.kaggle.com/datasets/shashwatwork/knee-osteoarthritis-dataset-with-severity.

References

1) Ahmed R, Imran A. Knee osteoarthritis analysis using deep learning and XAI on X-rays. *IEEE Access*. 2024;12:1–12. https://doi.org/10.1109/ACCESS.2024.3400987.
2) Zhao H, Ou L, Zhang Z, Zhang L, Liu K, Kuang J. The value of deep learning-based X-ray techniques in detecting and classifying K-L grades of knee osteoarthritis: a systematic review and meta-analysis. *European Radiology*. 2025;35(1):327–340. https://doi.org/10.1007/s00330-024-10928-9.
3) Vaattovaara E, Panfilov E, Tiulpin A, Niinimaki T, Niinimaki J, Saarakkala S, et al. Kellgren-Lawrence grading of knee osteoarthritis using deep learning: diagnostic performance with an external dataset and comparison with four readers. *Osteoarthritis and Cartilage Open*. 2025;p. 100580. https://doi.org/10.1016/j.ocarto.2025.100580.
4) Rehman A, Alghofaily B, Saba T, Raza A, Alamri FS. Transfer learning-based smart features engineering for osteoarthritis diagnosis from knee X-ray images. *IEEE Access*. 2023;11:152040–152053. https://doi.org/10.1109/ACCESS.2023.3294542.
5) Nisha, Chhabra M. Deep learning based knee osteoarthritis grade classification. *Proceedings of INDIACom*. 2024. https://doi.org/10.23919/INDIACom61295.2024.10499112.
6) Kitukale G, Shelke NA, Agrawal R, Singh NP, Quamara S. Predicting knee osteoarthritis using deep neural network. *Proceedings of IEEE I2CT*. 2024. https://doi.org/10.1109/I2CT61223.2024.10543978.
7) Chetlapalli AK, Bhuvaneswari R. Advanced knee osteoarthritis detection using deep learning. *Proceedings of IEEE I2CT*. 2024. https://doi.org/10.1109/I2CT61223.2024.10543594.
8) Rawat A, Dhyani A, Panwar D, Bisht G, Vats S, Sharma V. Knee osteoarthritis prediction using transfer learning. *Proceedings of IEEE AIC*. 2024. https://doi.org/10.1109/AIC61668.2024.10731036.
9) Kinger S. Deep learning for automatic knee osteoarthritis severity grading and classification. *Indian Journal of Orthopaedics*. 2024;58(10):1458–1473. https://doi.org/10.1007/s43465-024-01259-4.
10) Rani S, Memoria M, Almogren A, Bharany S, Joshi K, Altameem A, et al. Deep learning to combat knee osteoarthritis and severity assessment using CNN-based classification. *BMC Musculoskeletal Disorders*. 2024;25:79. https://doi.org/10.1186/s12891-024-07942-9.
11) Jain RK, Sharma PK, Gaj S, Sur A, Ghosh P. Knee osteoarthritis severity prediction using an attentive multi-scale deep convolutional neural network. *Computers in Biology and Medicine*. 2021;137:104799. https://doi.org/10.1016/j.compbiomed.2021.104799.
12) Pi SW, Lee SH, Lee J, Kim HJ, Lee YJ. Ensemble deep-learning networks for automated osteoarthritis grading in knee X-ray images. *Scientific Reports*. 2023;13:22800. https://doi.org/10.1038/s41598-023-50210-4.
13) Sohail M, Azad MM, Kim HS. Knee osteoarthritis severity detection using deep inception transfer learning. *Computers in Biology and Medicine*. 2025;186:109641. https://doi.org/10.1016/j.compbiomed.2024.109641.
14) Sharma A, Singh J, Kumar A, Bajaj V, Gupta S. Comparative analysis of deep learning models for early detection of osteoarthritis using knee radiographs: a retrospective study. *Journal of Clinical Orthopaedics and Trauma*. 2025;p. 103212. https://doi.org/10.1016/j.jcot.2025.103212.
15) Zhang H, Kong C, Chang H, Gan Y. Automated Kellgren-Lawrence grading of knee osteoarthritis using a multi-scale attention-based deep learning framework. *Frontiers in Medicine*. 2026;13:1809768. https://doi.org/10.3389/fmed.2026.1809768.
16) Chen P, Gao L, Shi X, Allen K, Yang L. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. *Computerized Medical Imaging and Graphics*. 2019;75:84–92. https://doi.org/10.1016/j.compmedimag.2019.06.002.
17) Thomas KA, Kidziński Ł, Halilaj E, Fleming SL, Venkataraman GR, Oei EHG, et al. Automated classification of radiographic knee osteoarthritis severity using deep neural networks. *Radiology: Artificial Intelligence*. 2020;2:e190065. https://doi.org/10.1148/ryai.2020190065.
18) He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of CVPR*. 2016. https://doi.org/10.1109/CVPR.2016.90.
19) Nurmirinta TAT, Turunen MJ, Tohka J, Mononen ME, Liukkonen MK. Predicting knee osteoarthritis severity from radiographic predictors: data from the Osteoarthritis Initiative. *Annals of Biomedical Engineering*. 2025. https://doi.org/10.1007/s10439-025-03740-z.
20) Moura FA, Pelegrinelli ARM, Catelli DS, Lamontagne M, Torres RS. Knee osteoarthritis prediction from gait kinematics: exploring deep neural networks and transfer learning for time-series classification. *Journal of Biomechanics*. 2025;p. 112881. https://doi.org/10.1016/j.jbiomech.2025.112881.
21) Tariq T, Suhail Z, Nawaz Z. A review for automated classification of knee osteoarthritis using KL grading scheme for X-rays. *Biomedical Engineering Letters*. 2025;15(1):1–35. https://doi.org/10.1007/s13534-024-00437-5.
22) Bukhari SNH, Altaf S. Automated severity classification of knee osteoarthritis from radiographs using transfer learning based deep neural networks. *Journal on Artificial Intelligence*. 2026;8(1):137–152. https://doi.org/10.32604/jai.2026.077943.

23) Huang Y, Shao Z, Wang R, Qiu H. Clinically integrated multi-modal transformer framework with cross-modal gated fusion and clinical nomogram for automated Kellgren–Lawrence grading of knee osteoarthritis on X-ray images. *BMC Musculoskeletal Disorders*. 2026;27:528. https://doi.org/10.1186/s12891-026-10043-4.
24) Atik I, Polat O, Atik S, Gul E, Apalan D. Osteoarthritis severity classification in knee X-rays using optimized deep learning approaches. *Journal of Imaging Informatics in Medicine*. 2026. https://doi.org/10.1007/s10278-025-01818-7.
25) Yilmaz U, Solak FZ. A joint fusion framework integrating traditional and deep image features for improved knee osteoarthritis grading. *Concurrency and Computation: Practice and Experience*. 2026;38(2):e70546. https://doi.org/10.1002/cpe.70546.
26) Vironicka S, Sathiaseelan JGR. Classification of long-bone fractures using modified Faster RCNN for X-ray images. *Indian Journal of Science and Technology*. 2023;16(1):56–65. https://doi.org/10.17485/IJST/v16i1.1690.