# A Hybrid Contrastive Ordinal Regression Method for Advancing Disease Severity Assessment in Imbalanced Medical Datasets

Afsah Saleem[1,2], Joshua R. Lewis[1,2], and Syed Zulqarnain Gilani[1,2,3]

[1] Centre for AI & ML, School of Science, Edith Cowan University, Joondalup, Australia
[2] Nutrition and Health Innovation Research Institute, Edith Cowan University, Joondalup, Australia
[3] Computer Science and Software Engineering, The University of Western Australia, Perth, Australia
afsah.saleem@ecu.edu.au

**Abstract.** Accurate disease grading is critical for early diagnosis and effective treatment planning. However, class imbalance and subtle inter-class variations in real-world disease grading datasets make it challenging for traditional classification models to differentiate between neighboring disease stages and preserve ordinal label relationships. Existing approaches emphasize inter-class ordinal relationships but fail to distinguish closely related categories effectively. To address these limitations, we consider disease grading as an ordinal regression problem and adopt a supervised contrastive learning approach to design a hybrid supervised contrastive ordinal learning framework. Our framework consists of three basic modules: 1) prototype-based contrastive ordinal learning, 2) weighted sample-based contrastive learning and 3) disease stage grading using regression. To deal with class imbalance while enhancing intra-class consistency and inter-class separation, we design a distance-based prototype contrastive ordinal loss, which pushes the samples closer to their class centers while maintaining their ordinality. This approach captures subtle differences within closely related disease stages and results in a separable ordinal latent space. Additionally, a per-sample class weighting strategy is integrated into weighted supervised contrastive ordinal learning to prevent class collapse, ensuring balanced gradient contributions and robust inter-class separation. Our approach effectively captures both large-scale and fine-grained variations, enabling precise ordinal classification for disease grading. We validate the framework on diabetic retinopathy and breast cancer datasets, demonstrating its adaptability across medical conditions and potential to enhance diagnostic accuracy in medical imaging applications.

**Keywords:** Ordinal learning · Supervised contrastive learning · Prototype contrastive learning · Class imbalance · Inter-class variations · Intra-class variations

## 1 Introduction

Automated disease grading systems play an important role in accurately classifying disease severity into distinct stages, enabling early intervention to reduce long-term complications and deaths [15,12,10]. However, developing such systems for real-world medical datasets presents significant challenges due to the inherent complexities of the data. These datasets often have long-tailed class distributions, with underrepresented severe or rare cases, making it challenging for models to learn from minority samples. Additionally, many medical grading problems have an ordinal structure, where disease severity progresses in a sequential manner. The subtle differences between the neighboring stages make the automated disease grading even more challenging [14]. The conventional classification losses focused on enhancing the separability fail to preserve ordinal relationships [20]. Conversely, regression-based approaches rely on continuous labels but struggle to handle inter-class variations effectively [20].

Recent advances in deep neural networks have improved ordinal regression by preserving the hierarchy of labels while mapping features to ordered labels [5,11]. A common approach is to treat ordinal regression as a multi-binary classification problem, summing binary decisions on whether output exceeds predefined ranks [13,3]. Some methods model probabilistic distributions within CNN logits or ordinal labels [3,12], while some explore continuous-space mappings for better predictions [17,5]. Hybrid models that combine CNNs with random forests explore ordinal learning through ensemble-based feature partitioning and distribution modelling [21,10]. Despite their effectiveness, these methods struggle with long-tailed medical datasets, failing to capture subtle variations and improve minority-class performance.

In this work, we aim to capture subtle variations while enhancing inter-class variations and intra-class diversity in imbalanced medical datasets to improve diagnostic accuracy, particularly for underrepresented and minority classes. To address these challenges, we adopt supervised contrastive learning (SupCon) [8] and introduce a novel hybrid contrastive ordinal framework for disease grading. While SupCon has been explored for medical image classification [6,19], it cannot preserve ordinal relationships in grading tasks. To address this, Dai et al. [4] introduced Adaptive Contrastive Loss (AdaCon), incorporating an adaptive margin based on the Empirical Cumulative Distribution Function (ECDF) [16]. Although effective for large datasets, AdaCon struggles with long-tailed imbalanced datasets [4]. Supervised Contrastive Ordinal Loss (SCOL) [14] extends SupCon with a metric learning approach based on Euclidean distances, improving the separation between classes and the compactness between classes by adjusting the distances based on labels. However, its performance degrades under severe class imbalance, highlighting the need for further improvements.

To address the above challenge, we design a hybrid contrastive ordinal framework to address class imbalance and ordinal consistency in disease grading by proposing Prototype-based Contrastive Ordinal Loss (PCOL) and weighted Supervised Contrastive Ordinal Loss ($SCOL_w$). PCOL enhances global structure learning by leveraging class prototypes as representative embeddings, aligning

individual samples with their respective prototypes. This promotes intra-class compactness, inter-class separability, and ordinal consistency through a distance-aware regularization term, enabling the model to capture subtle variations in disease severity. Further to avoid class collapse and improve robustness against class imbalance, we formulate $SCOL_w$ by refining SCOL using weighted class averaging, where each sample is weighted based on the inverse frequency of its class. $SCOL_w$ preserves intra-class compactness while learning fine-grained inter-class variations through bounded ordinal penalties, preventing the over-penalization of underrepresented samples and stabilizing feature representations. Our framework effectively balances local pairwise alignment and global class-level structure by integrating sample-to-prototype contrastive learning via PCOL and sample-to-sample contrastive learning through $SCOL_w$.

We evaluate our framework on two public datasets: Diabetic Retinopathy (DR) fundus photographs [7] and Breast Ultrasound Images (BUSI) [1], both widely used for disease grading in medical imaging [12,10,9]. The ablation study results validate the effectiveness of each component of the proposed framework. The overall performance of the proposed framework demonstrates its effectiveness in enhancing disease grading accuracy compared to state-of-the-art methods [12,10,9]. The code is publicly available at [2].

Our contributions are summarized as follows: *(i)* We design a novel loss function, called Prototype-based Contrastive Ordinal Loss (PCOL), that aligns samples with class prototypes to enforce intra-class compactness, inter-class separability, and ordinal consistency. A distance-aware regularization term preserves ordinal relationships, enabling the model to capture subtle variations in neighboring classes. *(ii)* We propose a hybrid contrastive ordinal framework integrating PCOL with weighted SCOL ($SCOL_w$) to enhance ordinal representation learning. By refining SCOL with per-sample class weighting ($w_i$), $SCOL_w$ prevents over-penalization, ensuring robust feature representation and better generalization in long-tailed distributions. Combining sample-to-prototype (PCOL) and sample-to-sample ($SCOL_w$) contrastive learning, our approach effectively addresses class imbalance while maintaining ordinality. *(iii)* Our approach achieves SOTA results on two public medical datasets, demonstrating robustness, generalizability and effectiveness in disease grading under severe class imbalance.

## 2 Methodology

### 2.1 Preliminaries:

The generic feature map of an input image $X$ extracted from the CNN backbone model is represented as $\psi = f_{\theta_b}(X)$, where $\psi \in \mathbb{R}^{h_0 \times w_0 \times c}$ and $h_0, w_0$, and $c$ denote the height, width, and number of channels, respectively. We aim to find an optimal mapping of $\psi$ using supervised contrastive ordinal learning methods to improve the disease severity grading. Fig. 1 shows the workflow of our proposed framework. It comprises of three modules: i) prototype-based contrastive ordinal learning, ii) weighted supervised contrastive ordinal learning, and iii) regression for predicting disease stage.

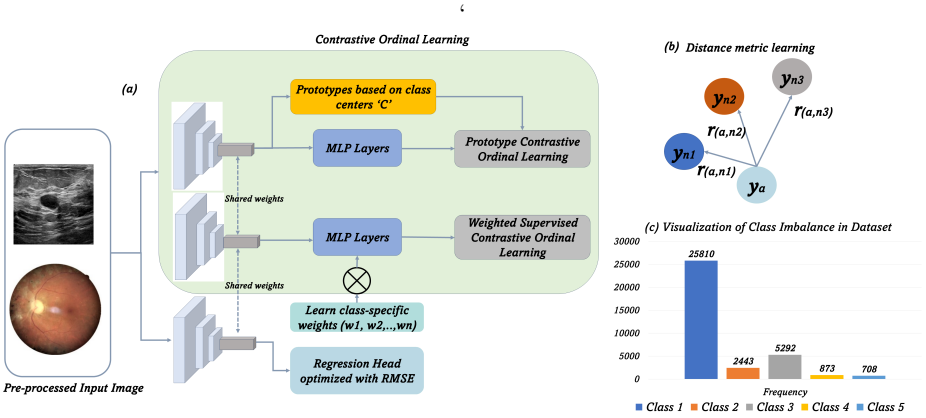


**Fig. 1.** (a) An overview of the proposed Hybrid Supervised Contrastive Ordinal Learning framework. During training, all three heads are jointly optimized using a single-stage contrastive learning approach. In the testing phase, only the regression head is used for inference. (b) depiction of distance metric learning components, (c) Class imbalance in the dataset, emphasizing the long-tailed distribution of disease severity.

### 2.2 Prototype-based Contrastive Ordinal Loss (PCOL)

To improve the performance of supervised contrastive ordinal learning frameworks [8,4,14] on imbalanced disease grading datasets with minimal inter-class variation and high intra-class diversity, we design Prototype-based Contrastive Ordinal Loss ($PCOL$). This loss function leverages class centers as prototypes for each label and integrates an ordinal distance-aware regularizer to enforce inter-class separability while maintaining ordinal consistency in the latent space.

Within a mini-batch $I$, consider $C = \{C_1, C_2, ...C_l\}$ is the set of class centers for $l$ classes. $C_i$ is computed during training as the mean of feature embeddings for all samples with label $y_i$ in the current mini-batch. Consider $c_p$ and $c_n$ are the prototypes of positive and negative classes with labels $y_p$ and $y_n$, respectively and $f_a$ are the feature embeddings of anchor sample $a$. $N(i)$ is a set of all the negative prototypes associated with labels different from the label of the anchor $a$ and $r_{a,n}$ is the Euclidean distance between the label of anchor sample and the negative prototype. Our proposed Prototype-based Ordinal Loss ($PCOL$) loss enforces the anchor embedding $z_a$ to be closely aligned with the positive class center $c_p$ while being pushed away from the negative class centers $c_n$ according to the ordinal distance term $r_{a,n}$. PCOL can be formulated as:

$$\mathcal{L}_{PCOL} = -\sum_{i \in \mathcal{I}} \log \frac{\exp\left(f_a^T \cdot c_p/\tau\right)}{\sum_{c_n \in N(i)} \exp\left((f_a^T \cdot c_n + r_{a,n})/\tau\right)} \tag{1}$$

The inclusion of $r_{a,n}$ in $PCOL$ ensures that misalignment penalties are proportional to ordinal distances, preserving ordinal relationships in the latent space. This formulation enhances intra-class compactness while maintaining inter-class distances aligned with ordinal label relationships.

### 2.3 Weighted Supervised Contrastive Ordinal Loss

We propose weighted Supervised Contrastive Ordinal Loss ($SCOL_w$) by incorporating per-sample class weighting $w_i$ with supervised contrastive ordinal loss [14] to enhance inter-class variations and intra-class diversity. In $SCOL_w$, weights are dynamically assigned to each sample in the training batch based on the inverse frequency of its class in the dataset. This loss enforces embeddings with similar labels to align closely while progressively increasing separation based on ordinal differences between their labels.

To define $SCOL_w$, consider $f_a$, $f_p$ and $f_n$ represents the feature embeddings of the anchor $a$, positive $p$ and negative $n$ samples, respectively. P(i) and N(i) are the sets of all the positive and negative samples, respectively. $\omega_i$ is per-sample class weighting and $r_{a,n}$ is the Euclidean distance between the anchor label $y_a$ and the negative label $y_n$, then $L_{SCOL_w}$ is given as:

$$L_{SCOL_w} = \sum_{i \in B} \frac{-w_i}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(f_a^T \cdot f_p / \tau\right)}{\sum_{n \in N(i)} \exp\left((f_a^T \cdot f_n + r_{a,n})/\tau\right)} \tag{2}$$

### 2.4 Hybrid Contrastive Ordinal Regression for Disease Severity Prediction

The contrastive ordinal embeddings learned through two projection heads are the passed through the regression head to predict disease severity. This final component is trained with a root-mean-squared loss ($L_{RMSE}$) for precise ordinal predictions. The projection heads and regression head are optimized jointly in a single-stage supervised contrastive learning framework, ensuring robust feature learning and improved disease grading accuracy. The total loss is defined as:

$$L_{total} = \alpha L_{PCOL} + \beta L_{SCOL_w} + L_{RMSE} \tag{3}$$

## 3 Experimental Setting

**Datasets:** We evaluate our approach on two public datasets widely used for disease grading: *(i)* Diabetic Retinopathy (DR) Fundus Photograph Dataset[7]: This dataset contains 35,126 high-resolution fundus images categorized into five DR severity levels: No DR (25,810, 74%), Mild (2,443, 7%), Moderate (5,292, 15%), Severe (873, 3%), and Proliferative DR (708, 2%). Following [18,3,12], we use 10-fold subject-independent cross-validation for evaluation. *(ii)* Breast Ultrasound Images (BUSI)[1]: BUSI includes 780 ultrasound images categorized into three classes: Normal (133, 17%), Benign (487, 56%), and Malignant (210, 26%). Following [12,9], we perform 5-fold subject-independent cross-validation.

**Evaluation Metrics:** Following state-of-the-art methods [18,3,12], we evaluate the performance of our framework in terms of Accuracy (Acc.) and Mean Absolute Error (MAE).

**Implementation details:** We use EfficientNet-V2S as the backbone encoder, resizing images to $300 \times 300$ and normalizing them to [0,1]. The projection heads in contrastive learning blocks have 1280 and 128 neurons. The feature map $\psi$ from the encoder is passed through global average pooling (grey layer) to convert into feature-embeddings, then to two separate MLP layers (blue box), each consisting of two dense layers with 1280 and 128 neurons. We use class-stratified batch sampling to ensure the stability of class prototypes even in small batch sizes. A one-stage contrastive learning framework is trained for 75 epochs with a batch size of 24 and an initial learning rate of $1 \times 10^{-3}$, optimized based on validation loss. To prevent overfitting, the learning rate is reduced by a factor of 0.2 after 5 epochs of no improvement, and early stopping is applied with a patience of 13 epochs.

## 4 Results and Discussion

### 4.1 Comparison with State-of-the-Art Ordinal Learning Methods:

Our proposed framework achieves state-of-the-art performance on the BUSI dataset, with 91.0% accuracy and a 0.10 reduction in MAE, outperforming CORE [9] by 8.8% in accuracy and 8% in MAE, demonstrating its effectiveness in capturing discriminative features and minimizing prediction errors.

On the DR dataset, our model achieves 83.8% accuracy and the lowest MAE (0.21), effectively reducing prediction errors across disease stages. While trailing Ord2Seq [18] by 0.7% in overall accuracy, it significantly improves classification for minority classes (2, 3 and 5) (see Fig.1) and performs comparably for class 4, highlighting its reliability and consistency. Notably, our model (G5) (see Fig. 2), further reduces misclassification rates in these classes, achieving a higher proportion of correct classifications reducing both adjacent and non-adjacent errors where previous methods struggled. These results highlight its effectiveness in ensuring ordinal consistency, class separability, and robustness in long-tailed, imbalanced medical datasets.

**Table 1.** Performance comparison with state-of-the-art ordinal learning methods using BUSI and DR datasets. The best results are in Bold, and the second best are underlined.

| Method | CNN Model | BUSI | | DR | |
|---|---|---|---|---|---|
| | | Acc. (↑) | MAE(↓) | Acc. (↑) | MAE (↓) |
| Poisson [3] | VGG-16 | 75.2 | 0.24 | 77.1 | 0.38 |
| SORD [5] | VGG-16 | 78.2 | 0.20 | 78.2 | 0.73 |
| POE [11] | VGG-16 | 75.0 | 0.21 | 80.5 | 0.30 |
| CORE [9] | VGG-16 | 82.2 | 0.18 | 83.3 | 0.25 |
| Ord2Seq [18] | VGG-16 | - | - | **84.0** | 0.25 |
| SCOL [14] | Eff-V2S | 85.0 | 0.17 | 82.5 | 0.32 |
| Ours | VGG-16 | <u>88.5</u> | <u>0.15</u> | 83.5 | <u>0.23</u> |
| Ours | Eff-V2S | **91.0** | **0.10** | <u>83.8</u> | **0.21** |

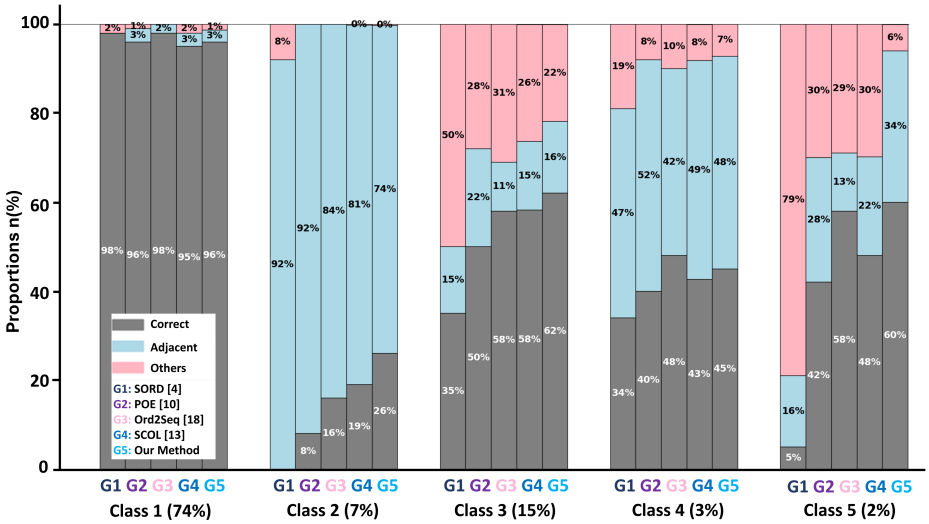


**Fig. 2.** The stacked bar plot to compare the class-wise performance of the proposed method with SOTA methods using the Diabetic Retinopathy dataset. G1: SORD [5], G2: POE [11], G3: Ord2Seq [18], G4: SCOL [14], and (G5): our proposed method. Each bar represents the proportion of predictions classified correctly (grey), misclassified into adjacent classes (blue), or assigned to non-adjacent classes (pink).

### 4.2 Comparison with SOTA Contrastive Learning Methods:

Table 2 compares our proposed method with SOTA contrastive learning approaches SupCon [8], AdaCon [4], and SCOL [14] on the BUSI dataset. Our method achieves the highest overall accuracy (91.0%) and the lowest MAE (0.10), demonstrating superior performance. It significantly improves Benign classification with 94.7% accuracy and 0.05 MAE, while also achieving the best performance in the Malignant class (85.2% accuracy, 0.15 MAE). While SCOL achieves slightly higher accuracy for Normal (88.7%), our model remains competitive (87.2%) with the lowest MAE (0.15), ensuring more stable predictions. These results confirm the effectiveness and robustness of our hybrid contrastive ordinal framework in handling class imbalance and improving disease grading accuracy.

**Table 2.** Comparison with SOTA Contrastive Learning Methods using BUSI dataset.

| **Method** | **Normal** | | **Benign** | | **Malignant** | | **Average** | |
|---|---|---|---|---|---|---|---|---|
| | Acc.(↑) | MAE(↓) | Acc.(↑) | MAE(↓) | Acc.(↑) | MAE(↓) | Acc.(↑) | MAE(↓) |
| SupCon [8] | 84.6 | 0.17 | 95.4 | 0.13 | 80.9 | 0.31 | 87.0 | 0.20 |
| AdaCon [4] | 85.0 | 0.20 | 92.0 | 0.11 | 80.0 | 0.33 | 86.0 | 0.21 |
| SCOL [14] | **88.7** | 0.17 | 93.8 | 0.13 | 79.0 | 0.31 | 89.0 | 0.17 |
| Ours | 87.2 | **0.15** | **94.7** | **0.05** | **85.2** | **0.15** | **91.0** | **0.10** |

**Table 3.** Ablation study to analyze the effect of $PCOL$ and $SCOL_w$ on each class and to compare the effect of regression vs. classification head using BUSI dataset. CE: Cross entropy, RMSE: Root mean squared error.

| **Method** | | | | | **Normal** | | **Benign** | | **Malignant** | | **Average** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $PCOL$ | $SCOL_w$ | $w_i$ | RMSE | CE | Acc. (↑) | MAE (↓) | Acc. (↑) | MAE (↓) | Acc. (↑) | MAE (↓) | Acc. (↑) | MAE (↓) |
| - | ✓ | - | ✓ | - | 87.0 | 0.21 | 90.4 | 0.18 | 77.5 | 0.28 | 85.5 | 0.22 |
| - | ✓ | ✓ | ✓ | - | **89.5** | 0.16 | 92.4 | 0.14 | 78.1 | 0.30 | 88.0 | 0.13 |
| ✓ | - | - | ✓ | - | 88.0 | 0.16 | 94.0 | 0.11 | 80.0 | 0.33 | 89.0 | 0.11 |
| ✓ | ✓ | ✓ | - | ✓ | 85.7 | 0.19 | 89.5 | 0.12 | **86.6** | **0.11** | 88.0 | 0.13 |
| ✓ | ✓ | ✓ | ✓ | - | 87.2 | **0.15** | **94.7** | **0.05** | **85.2** | **0.15** | **91.0** | **0.10** |

### 4.3 Ablation Study

**Effectiveness of PCOL and $SCOL_w$:** Table 3 evaluates the impact of PCOL, $SCOL_w$, and per-sample weighting ($w_i$) on the BUSI dataset. $SCOL_w$ alone achieves 85.5% accuracy, 0.22 MAE, while adding $w_i$ improves the performance to 88.0% accuracy and 0.13 MAE, highlighting the benefits of class weighting. PCOL alone further enhances accuracy to 89.0% with 0.11 MAE, demonstrating its role in improving feature representation. The overall framework achieves the best performance (91.0% accuracy, 0.10 MAE), reinforcing intra-class compactness, inter-class separability, and robustness in disease grading.

**Regression vs. Classification Head** Table 3 also compares the effectiveness of Root Mean Squared Error (RMSE) loss and Cross-Entropy (CE) loss. RMSE outperforms CE, achieving 91.0% accuracy, 0.10 MAE versus 88.0% accuracy, 0.13 MAE, indicating improved stability. It significantly improves Benign (94.7% accuracy, 0.05 MAE) and Normal (87.2% accuracy, 0.15 MAE) classification. For Malignant, both losses yield 0.15 MAE, but CE achieves slightly higher accuracy (86.6% vs. 85.2%). These results confirm RMSE's superiority in ordinal prediction, reducing misclassification errors and improving model reliability.

Fig. 3 represents the t-sne plot for Diabetic Retinopathy dataset to visualize learned feature embeddings using different contrastive ordinal loss functions. The spiral-like layout reflects the preserved ordinal structure. It can be seen in the Fig3(a&b) that the $(SCOL_w)+RMSE$ yields loosely clustered embeddings with moderate inter-class separation where $PCOL+RMSE$ produces more compact clusters around class-specific prototypes. The combined SCOL + PCOL along with RMSE loss achieves the most structured and well-separated representation in Fig3(c), balancing local discrimination and global ordinal alignment.

## 5 Conclusion

We introduced a novel hybrid contrastive ordinal learning framework to assess disease severity in long-tailed medical imaging datasets. Our findings demonstrate the strength of the prototype-based and instance-based contrastive ordinal learning components, which work together to enhance inter-class separability

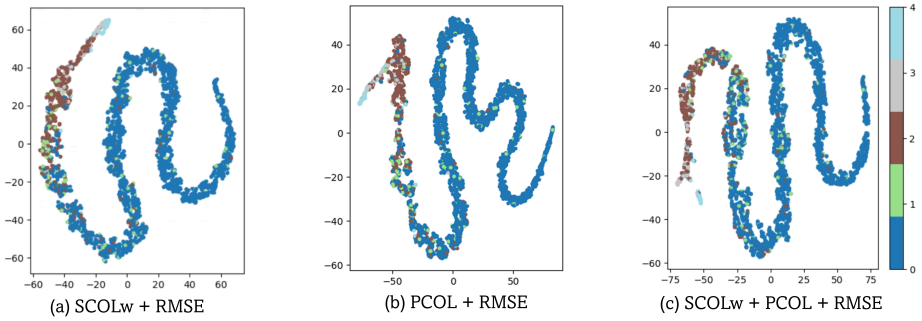


**Fig. 3.** t-sne plot for Diabetic Retinopathy dataset to visualize learned feature embeddings using different contrastive ordinal loss functions. Colourbar represents ordinal class label (0–4), and the spiral-like layout reflects the preserved ordinal structure.

and improve intra-class consistency. Ablation studies further validated that the integration of these components leads to superior performance and increased robustness, particularly in handling minority class predictions and reducing errors in various stages of the disease. Our framework sets a new benchmark for medical image-based disease grading by effectively tackling class imbalance, achieving SOTA performance with greater reliability. Future work will focus on extending this approach to more complex datasets and exploring its clinical deployment.

**Acknowledgement and Data Use Declaration:** SZG was partially funded by a Raine Priming Grant from the Raine Medical Research Foundation and a WANMA grant funded by the Western Australian Future Health Research and Innovation Fund. The salary of J.R.L. is supported by a National Heart Foundation Future Leader Fellowship (IDs: 102817 & 107323).

**Disclosure of Interests** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
2. AS: A-hybrid-contrastive-ordinal-regression (2025), `https://github.com/AfsahS/A-Hybrid-Contrastive-Ordinal-Regression`
3. Beckham, C., Pal, C.: Unimodal probability distributions for deep ordinal classification. In: International Conference on Machine Learning. pp. 411–419. PMLR (2017)
4. Dai, W., Li, X., Chiu, W.H.K., Kuo, M.D., Cheng, K.T.: Adaptive contrast for image regression in computer-aided disease assessment. IEEE Transactions on Medical Imaging (2021)
5. Diaz, R., Marathe, A.: Soft labels for ordinal regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4738–4747 (2019)

6. Jaiswal, A., Li, T., Zander, C., Han, Y., Rousseau, J.F., Peng, Y., Ding, Y.: Scalp-supervised contrastive learning for cardiopulmonary disease classification and localization in chest x-rays using patient metadata. In: 2021 IEEE International Conference on Data Mining (ICDM) (2021)
7. Kaggle: Diabetic retinopathy detection. `https://www.kaggle.com/c/diabetic-retinopathy-detection` (2015), `https://www.kaggle.com/c/diabetic-retinopathy-detection`, kaggle Competition
8. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in neural information processing systems (2020)
9. Lei, Y., Li, Z., Li, Y., Zhang, J., Shan, H.: Core: Learning consistent ordinal representations with convex optimization for image ordinal estimation. Pattern Recognition **156**, 110748 (2024)
10. Lei, Y., Zhu, H., Zhang, J., Shan, H.: Meta ordinal regression forest for medical image classification with ordinal labels. IEEE/CAA Journal of Automatica Sinica (2022)
11. Li, W., Huang, X., Lu, J., Feng, J., Zhou, J.: Learning probabilistic ordinal embeddings for uncertainty-aware regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13896–13905 (2021)
12. Liu, X., Zou, Y., Song, Y., Yang, C., You, J., K Vijaya Kumar, B.: Ordinal regression with neuron stick-breaking for medical diagnosis. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0 (2018)
13. Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G.: Ordinal regression with multiple output cnn for age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4920–4928 (2016)
14. Saleem, A., Ilyas, Z., Suter, D., Hassan, G.M., Reid, S., Schousboe, J.T., Prince, R., Leslie, W.D., Lewis, J.R., Gilani, S.Z.: Scol: Supervised contrastive ordinal loss for abdominal aortic calcification scoring on vertebral fracture assessment scans. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2023)
15. Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R., Sieh, W.: Deep learning to improve breast cancer detection on screening mammography. Scientific reports **9**(1), 12495 (2019)
16. Van der Vaart, A.W.: Asymptotic statistics, vol. 3. Cambridge university press (2000)
17. Vargas, V.M., Gutiérrez, P.A., Hervás-Martínez, C.: Unimodal regularisation based on beta distribution for deep ordinal regression. Pattern Recognition **122**, 108310 (2022)
18. Wang, J., Cheng, Y., Chen, J., Chen, T., Chen, D., Wu, J.: Ord2seq: Regarding ordinal regression as label sequence prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5865–5875 (2023)
19. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. Medical image analysis **81**, 102559 (2022)
20. Zhang, S., Yang, L., Mi, M.B., Zheng, X., Yao, A.: Improving deep regression with ordinal entropy. arXiv preprint arXiv:2301.08915 (2023)
21. Zhu, H., Shan, H., Zhang, Y., Che, L., Xu, X., Zhang, J., Shi, J., Wang, F.Y.: Convolutional ordinal regression forest for image ordinal estimation. IEEE transactions on neural networks and learning systems **33**(8), 4084–4095 (2021)