



(CS 6956)

Project Check In

Uncertainty Quantification in Quality Control of LGE-MRI Images

K M Arefeen Sultan

u1419693@utah.edu

Nicholas Jacobs

nicholas.jacobs@utah.edu

Md Rakibul Haque

rakibul.haque@utah.edu

Md Hasibul Husain Hisham

hasibul.hisham@utah.edu

March 21, 2025

Literature Review

0.1 Ordinal Regression

Traditional multi-category cross-entropy loss fails to capture the inherent order in the target labels. In ordinal tasks, the difference between adjacent labels (e.g., 1 vs 2) is not equivalent to the difference between non-adjacent labels (e.g., 1 vs 5), making conventional classification methods inappropriate.

Li et al. [6] were among the first to propose extending binary classification methods to solve ordinal regression by transforming each ordinal label into a binary vector representing whether the label exceeds a specific threshold. While effective, this method faces rank inconsistency issues, where the predicted probabilities do not respect the inherent ordering in the target space.

Cao et al. [2] proposed the CORAL (Rank-consistent Ordinal Regression for Neural Networks) method, which introduces a weight-sharing constraint in the output layer of the neural network to enforce rank consistency. While CORAL improved performance, it limited the model's expressiveness because of this constraint. To overcome this limitation, Shi et al. [7] proposed CORN (Conditional Ordinal Regression for Neural Networks), which eliminates the weight-sharing constraint. CORN achieves rank consistency by using conditional training sets and applying the chain rule for conditional probability distributions.

Another work on improving ordinal regression tasks was made by Zhang et al. [8], who introduced an ordinal entropy regularizer. This regularizer is designed to encourage higher entropy in the feature representations while maintaining the ordinal relationships in the target space. They demonstrated that classification-based methods (such as those using cross-entropy loss) outperformed regression-based methods (such as those using mean squared error loss) because they promote a higher marginal entropy of the learned representations, which helps improve model performance in tasks like depth estimation and crowd counting.

We are planning to explore these loss functions for our project.

0.2 Aleatoric & Epistemic Uncertainty

Uncertainty quantification in deep learning models has gained significant attention, especially in classification tasks where the model's predictions can have substantial real-world implications. In ordinal classification, where the target classes are discrete but ordered (e.g., 1, 2, 3, 4, 5), understanding the uncertainty of predictions becomes essential for tasks like medical diagnosis, sentiment analysis, and many others. In these tasks, the predicted classes need to respect the ordinal relationships between the classes, such as $1 < 2 < 3 < 4 < 5$, while also capturing the uncertainty that may arise from the model's predictions.

The uncertainty in deep learning models is typically divided into two main types: - Aleatoric uncertainty: This arises from the inherent noise in the data, such as measurement errors or variability in the data itself. - Epistemic uncertainty: This is related to the model's uncertainty due to limited knowledge, which can be reduced as more data is provided or the model is improved.

0.2.1 Total Uncertainty Decomposition for Ordinal Classification

In the context of ordinal classification, where the output consists of ordered categories, the total uncertainty of the model's predictions can be decomposed into two components: aleatoric and epistemic uncertainty. The total predictive uncertainty $H(\mathbb{E}[p(y_*|f(x_*;\theta))])$ is given by[1]:

$$H(\mathbb{E}[p(y_*|f(x_*;\theta))]) = \mathbb{E}[H(p(y_*|f(x_*;\theta)))] + I(y_*; \Theta)$$

Where: $H(p(y_*|f(x_*;\theta)))$ represents the aleatoric uncertainty, which is related to the inherent noise or variability in the data. $I(y_*; \Theta)$ is the epistemic uncertainty, which reflects the uncertainty due to the model's parameters and is associated with the lack of knowledge or limitations in the model.

The first term, $\mathbb{E}[H(p(y_*|f(x_*;\theta)))]$, quantifies the aleatoric uncertainty (data-related noise) in the model's predictions. The second term, $I(y_*; \Theta)$, captures the epistemic uncertainty (model-related uncertainty) due to the model's parameters.

0.2.2 Aleatoric Uncertainty in Ordinal Classification

In ordinal classification, aleatoric uncertainty arises due to the inherent noise in the data [4], [3]. For example, this could be due to mislabeled data points or data ambiguity, such as the difficulty in distinguishing between closely related ordinal classes (e.g., 3 and 4). Since ordinal regression is treated as a classification task, aleatoric uncertainty is often captured by the entropy of the model's predicted class probabilities.

For our ordinal classification task with classes $\{1, 2, 3, 4, 5\}$, the model outputs probabilities $p(y_i|x)$ for each class. The entropy of these probabilities measures the uncertainty in the model's classification decision, where higher entropy indicates greater uncertainty due to data noise or ambiguity.

The entropy of the predicted probability distribution for class y_i is given by:

$$H(p(y_*|f(x_*;\theta))) = - \sum_{i=1}^k p(y_i|x) \log p(y_i|x)$$

Here, $p(y_i|x)$ is the probability of predicting class y_i given input x , and k is the number of classes.

0.2.3 Epistemic Uncertainty in Ordinal Classification

Epistemic uncertainty reflects the uncertainty in the model due to incomplete knowledge or insufficient data, and it can be reduced by gathering more data or improving the model.

The mutual information between the model's predictions and its parameters, $I(y_*; \Theta)$, is used to represent epistemic uncertainty mathematically[5]:

$$I(y_*; \Theta) = H(p(y_*)) - \mathbb{E}_\theta[H(p(y_*|x, \theta))]$$

Here, $\mathbb{E}_\theta[H(p(y_*|x, \theta))]$ represents the expected entropy over different parameter samples

(drawn from the posterior), capturing the model’s uncertainty due to limited knowledge or variability in the model’s parameters.

Data Preprocessing

In this study, we will employ a dataset comprising 424 LGE MRI scans, each labeled for the purpose of Quality Assessment (QA) and has Left Atrium segmentations. Expert reviewers rated these scans on a scale from 1 to 5. These 424 scans have a class imbalance problem because most scans are in the 2 to 4 range.

Our dataset will be splitted into training and testing sets following an 80:20 split. Subsequently, the training set was further partitioned into a secondary training set and a validation set, adhering to the same 80:20 ratio. We will first obtain the region of interest by using the LA volume masks to crop it. This will be our input to the model.

Model & Loss Function

Our approach to uncertainty quantification in LGE-MRI quality assessment builds upon deep learning frameworks while incorporating uncertainty estimation techniques. The core of our model architecture processes 3D MRI scans through a ResNet3D encoder followed by a linear feature extractor and a classifier that outputs quality scores ranging from 1-5, where lower scores (1, 2) indicate non-diagnostic images and higher scores (3, 4, 5) represent diagnostic-quality images.

Currently, we have implemented a baseline model that resizes 3D MRI scans to fit GPU memory constraints and uses cross-entropy loss for multi-class training. We are extending this model in several key directions:

- **Input Representation:** Rather than processing entire 3D volumes, we will extract cropped windows containing only regions of interest (ROI) from the LGE-MRI scans. This targeted approach will allow the model to focus on the most relevant cardiac structures while reducing computational requirements.
- **Multi-Rater Learning:** Our dataset contains 424 LGE-MRI scans (3D), each independently rated by Three experts on a scale of 1-5. These different ratings by different experts reveal inter-rater variability of the scans. To address this inherent uncertainty, we will implement a multi-branch model architecture where each branch is trained to predict the ratings of a specific expert. By comparing predictions across branches, we can identify images with high diagnostic uncertainty that may require additional review in clinical settings.
- **Loss Function Design:** We are implementing a composite loss function combining:
 - Cross-Entropy (CE) loss for classification accuracy across the five quality classes
 - Binary Cross-Entropy (BCE) for the diagnostic/non-diagnostic distinction

The final loss function will be: $\lambda_1 \cdot \text{CE} + \lambda_2 \cdot \text{BCE}$, where λ values are weighting coefficients that will be tuned during experimentation. We are also planning to include an ordinal regression loss, as mentioned in the literature review.

Moreover, based on the model outputs from Three expert ratings, we will be implementing aleatoric and epistemic uncertainty methods, as mentioned in the literature review.

Future Direction & Timeline

Building on our current progress, we have revised our tentative project timeline to focus on implementing and evaluating uncertainty quantification on the quality assessment:

March 22-31, 2025: Complete Model Implementation and Baseline Experiments

- Complete ROI extraction and implement composite loss function
- Establish baseline model with proper evaluation metrics

April 1-15, 2025: Uncertainty Quantification

- Implement Uncertainty Quantification methods
- Design visualization tools for uncertainty estimates

April 16-24, 2025: Experimental Evaluation and Final Reporting

- Compare uncertainty quantification approaches
- Analyze results and prepare final report (structured like a conference paper)

References

- [1] Neil Band et al. *Benchmarking Bayesian Deep Learning on Diabetic Retinopathy Detection Tasks*. 2022. arXiv: 2211.12717 [stat.ML]. URL: <https://arxiv.org/abs/2211.12717>.
- [2] Wenzhi Cao, Vahid Mirjalili, and Sebastian Raschka. "Rank consistent ordinal regression for neural networks with application to age estimation". In: *Pattern Recognition Letters* 140 (2020), pp. 325–331.
- [3] Gustavo Carneiro et al. "Deep learning uncertainty and confidence calibration for the five-class polyp classification from colonoscopy". In: *Medical image analysis* 62 (2020), p. 101653.
- [4] Marc Combalia et al. "Uncertainty estimation in deep neural networks for dermoscopic image classification". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020, pp. 744–745.
- [5] Stefan Depeweg et al. "Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning". In: *International conference on machine learning*. PMLR. 2018, pp. 1184–1193.

- [6] Ling Li and Hsuan-Tien Lin. “Ordinal regression by extended binary classification”. In: *Advances in neural information processing systems* 19 (2006).
- [7] Xintong Shi, Wenzhi Cao, and Sebastian Raschka. “Deep neural networks for rank-consistent ordinal regression based on conditional probabilities”. In: *Pattern Analysis and Applications* 26.3 (June 2023), pp. 941–955. ISSN: 1433-755X. DOI: 10.1007/s10044-023-01181-9. URL: <http://dx.doi.org/10.1007/s10044-023-01181-9>.
- [8] Shihao Zhang et al. *Improving Deep Regression with Ordinal Entropy*. 2023. arXiv: 2301.08915 [cs.CV]. URL: <https://arxiv.org/abs/2301.08915>.