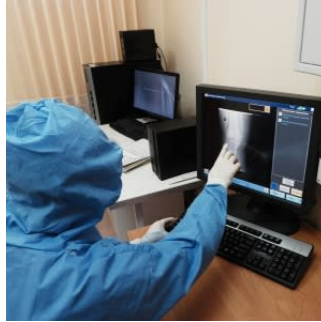


The Need for Reproducible Scoring Systems to Enhance Radiomics Research Quality



The aim of this recent study was to assess both the intra- and inter-rater reliability of the total radiomics quality score (RQS) and to examine the reproducibility of individual RQS items' score in a comprehensive multi-reader investigation.

Nine raters with diverse backgrounds were randomly assigned to three groups, each group reflecting varying levels of proficiency in utilising the RQS. 33 original research papers on radiomics were assessed by raters from groups 1 and 2. Among these 33 papers, 17 were subjected to a second evaluation, with a one-month interval, by raters belonging to group 3.

Despite its widespread adoption, the findings indicated that the RQS tool is not always easy to grasp and adopt, and its outcomes may not be consistently reproducible in many cases.

Based on study's findings, it is evident that there is room for improvement to establish a user-friendly scoring framework that can be readily employed by authors, reviewers, and editors to assess the quality of radiomics studies.

The study revealed an unexpected outcome related to the training session conducted before the application of the RQS. The raters of group 1 revealed poor inter-rater reliability despite receiving training, while group 2 exhibited moderate inter-rater reliability even though they had not received any prior instructions. Additionally, the impact of greater experience was evident in the intra-rater reliability analysis, as the more experienced raters demonstrated perfect intra-rater reliability results. By contrast, the less experienced raters achieved only moderate reliability in this aspect.

The raters noted that the RQS instructions were often not self-explanatory, and this meant additional time for interpretation of the RQS items was required, as well as to assign a score.

An increasing number of studies have used deep learning for radiomics analysis. However, it's worth noting that the current RQS tool mainly focuses on hand-crafted radiomics, lacking specific items that address the methodological challenges related to deep learning approaches in radiomics. As a result, robust designed deep learning studies may receive lower RQS total scores as they fail to address questions relevant to deep learning methodology.

Overall, reproducibility of both the total RQS and the scores of individual RQS items has been found to be low. There is a need for the development of a more robust and reproducible assessment method to effectively gauge the quality of radiomics research.

Source: [European Radiology](#)

Image Credit: [iStock](#)

Published on : Mon, 9 Oct 2023