



Rightsizing the Concept of Work Relative Value Units (wRVU) through Incorporation of an Automated Cognitive Burden Metric: A Feasibility Study

Adam E. Flanders, MD, CIIP, FSIM, Thomas Jefferson University; Steven A. Rothenberg, MD; Abin Bandyopadhyay, MD; Paras Lakhani, MD; Christopher G. Roth, MD; Ayis Pyrros, MD; Avishkar Sharma, MD, CIIP; Adam Zoga, MD MBA; Elias Melhem, MD

Introduction/Background

Radiology reimbursement assigns uniform value to all interpretations of the same exam code (wRVUs), ignoring the additional cognitive burden imposed by patient and exam complexity and management. Our hypothesis is that by selecting for a correct set of reporting characteristics, a cognitive load score can be created that serves as a proxy to better measure radiologist effort to produce a given report. The purpose of this study was to determine if an LLM could serve as a proxy evaluation for case complexity.

Methods/Intervention

Using a large language model (GPT-5.4 mini), we analyzed 100 random reports for five high-volume studies: CT Chest, CT Head, CT Abdomen/Pelvis, MRI Brain, and MRI Knee (n=500 total). LLM prompt was designed to focus on various aspects of report complexity including: trainee involvement, structured template use, supplemental clinical information, bundled exams, patient age, impression complexity, positive findings, organ systems, measured findings, comparisons, explicit communication, post-operative state, post-processing and follow-up recommendations. Each report was also scored on six ordinal dimensions (0–5): findings burden, measurements recorded, impression severity, clinical complexity, report structure and organ systems. A composite complexity score (0–100) was derived. Coefficient of variation (CV%) derived from STDEV quantified intra-exam score dispersion independent of report length. 300 reports were randomly sampled and rated by two board certified radiologists for exam/report complexity and correlated with the LLM results using Pearson coefficient.

Results/Outcome

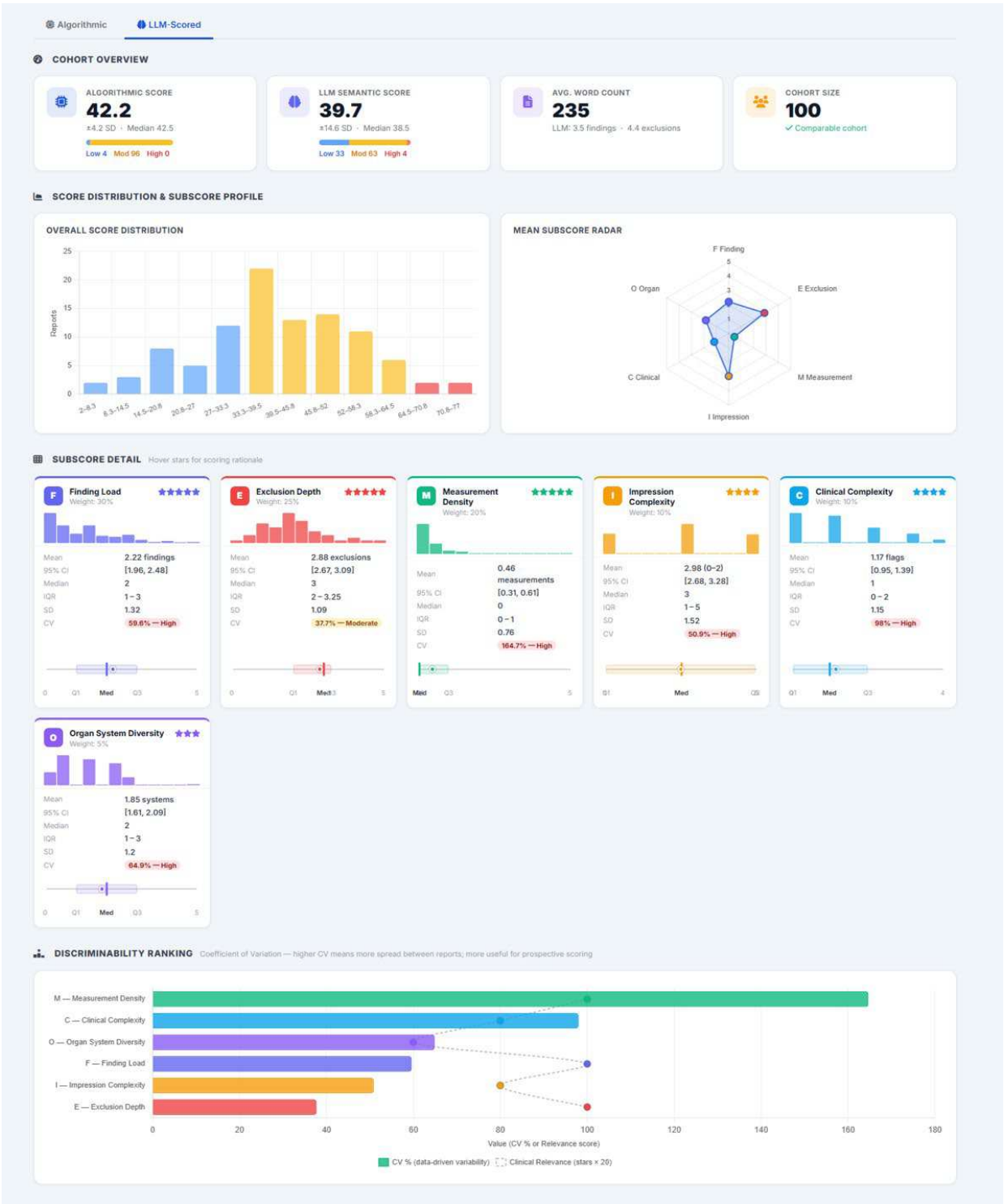
Composite complexity scores varied substantially within every exam type: CT Abd/Pelvis (mean 56.6, CV 30%), CT Chest (54.1, CV 23%), MRI Knee (55.0, CV 21%), MRI Brain (39.7, CV 37%), CT Head (33.2, CV 40%). There was moderate agreement between human expert ratings and the LLM ratings (Pearson 0.37-0.51). Finding counts ranged 2-fold to 10-fold within single CPT codes. Word count correlated poorly with complexity, confirming that verbosity is not a proxy for cognitive effort.

Conclusion

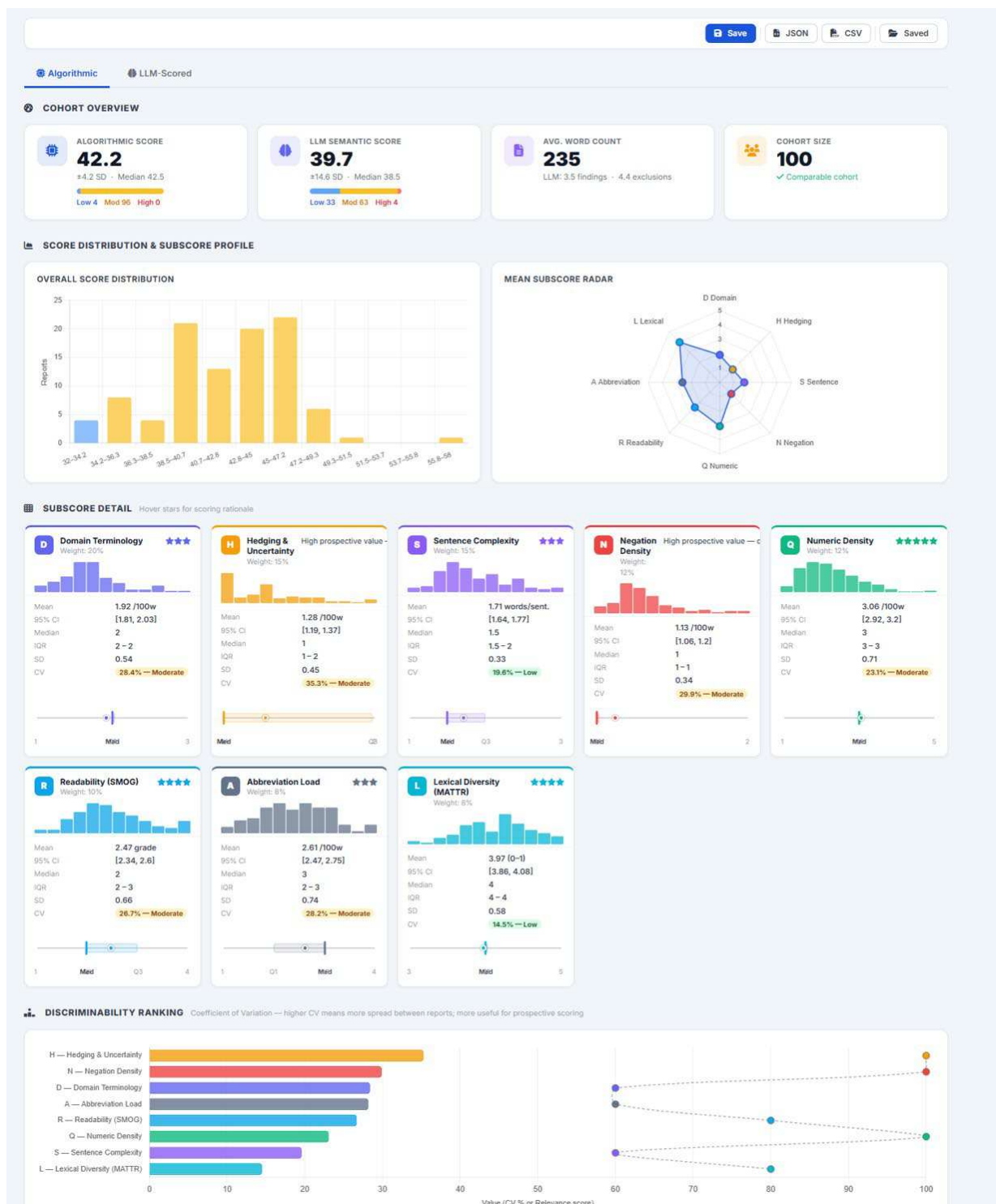
Automated methodology for internally augmenting the value of radiology reporting could provide a better solution for achieving reward based on cognitive burden and effort rather than base professional value of imaging exam.

Statement of Impact

Radiologists could be rewarded not only for a base wRVU value but also for a "virtual" RVU that reflects cognitive effort.



A complexity analysis population-level metrics dashboard evaluating a sample of 100 reports (out of 69,513 available) for MRI Brain W WO Contrast, showing LLM semantic scores alongside weighted subscore profiles for finding load, exclusion depth, and measurement density.



A complexity analysis population-level metrics dashboard evaluating a cohort of 100 reports for MRI Brain W WO Contrast, breaking down overall score distributions, mean subscore radars, and detailed linguistic metrics (including Domain Terminology, Hedging & Uncertainty, Sentence Complexity, Negation Density, Numeric Density, Readability, Abbreviation Load, and Lexical Diversity).

Keywords

LLM; RVU; reimbursement; complexity; economics