ANATOMICAL ENTROPY IN CANCER REGISTRIES: NOS CODING IN PHARYNGEAL ONCOLOGY AND IMPLICATIONS FOR ARTIFICIAL INTELLIGENCE

Abarana Atchuthan*

PGIM, University of Colombo, Sri Lanka

Abstract

Artificial intelligence (AI) increasingly depends on population-scale health data, yet the usefulness of such systems is constrained by the specificity of the information available for learning. Cancer registries frequently use "Not Otherwise Specified" (NOS) categories when precise anatomical subsites cannot be assigned. We propose anatomical entropy as a conceptual framework describing loss or unresolved uncertainty in anatomical specificity as cancer information enters structured datasets. Population-based pharyngeal cancer registry data from Sri Lanka were examined for the available years between 2015 and 2022. Cases classified within anatomical NOS categories were identified and expressed as a proportion of annual pharyngeal cancer registrations. Annual proportions and their range were summarized. Data for 2020 were unavailable for the present analysis. This was a registry data-quality analysis; no machine-learning model was developed or evaluated. The relevance to AI was therefore examined as an implication of reduced input-data granularity rather than as an empirically measured change in algorithmic performance. Anatomical NOS classifications accounted for 42.3%–49.4% of pharyngeal cancer registrations across the analyzed years, indicating that approximately two in five to one in two registered cases had limited subsite specificity. We distinguish intrinsic anatomical entropy, where advanced, overlapping, multifocal or otherwise indeterminate disease prevents confident localization, from system-induced anatomical entropy, where more specific information may exist in clinical, imaging, endoscopic, operative or pathological records but is not preserved in the registry. A substantial and persistent burden of anatomical non-specificity exists within the analyzed pharyngeal cancer registry data. Although this study did not test an AI model, reduced anatomical granularity could plausibly limit future algorithms seeking to learn subsite-specific epidemiological, biological, treatment or outcome patterns. Anatomical entropy therefore represents a testable data-quality hypothesis for AI-ready cancer registries, rather than demonstrated algorithmic performance loss. Practical responses include auditing the reasons for NOS assignment, strengthening standardized anatomical documentation and registrar training, linking registry records with imaging, pathology and endoscopy data, and evaluating human-verified automated information extraction. Future studies should directly test whether progressive anatomical information loss affects machine-learning discrimination, calibration and subgroup performance.

Keywords: anatomical entropy; cancer registry; pharyngeal cancer; Not Otherwise Specified; data quality;