Radiology report terminology to characterise reports in Southern Africa

Zola Mahlaza[1][0000−0001−9829−1480], Ernest Obbie Zulu[2][0009−0005−3260−7717], and Lighton Phiri[3][0000−0003−3582−9866]

[1] Department of Computer Science, University of Cape Town, South Africa
[2] University Teaching Hospitals - Adult Hospital, Zambia
[3] Department of Library and Information Science, University of Zambia, Zambia
zmahlaza@cs.uct.ac.za, obbiernest@gmail.com, lighton.phiri@unza.zm

Abstract. There is a shortage of trained radiologists in Southern African countries such as Zambia. The shortage calls for the use of artificial intelligence to bolster the efforts of the few practising radiologists to improve efficiency. Such AI-guided tools require knowledge on how to author good quality reports. Since there is no normative standard for Zambian reports, metadata is required to annotate existing reports to determine characteristics of good reports. As there are no Zambian guidelines for the information to be included in reports, we analyse papers, international guidelines, published structured reports, and existing structured reporting templates to create contemporary and international radiology report terminology, as a first step towards metadata. We identified 3199 terms from vetted templates published by the Radiological Society of North America's RadReport Template Library[4]. We also augmented them with 323 terms extracted from published papers (71% were manually annotated with SNOMED codes for quality assurance).

Keywords: Metadata · Artificial Intelligence · Radiology reports.

1 Introduction

There is a critical shortage of radiologists in Zambia. Specifically, there are less than 15 radiologists in the public sector [2]. As such, there is a need to investigate the extent to which computational solutions can improve radiology workflows. Globally, a number of researchers have identified opportunities for the use of Artificial Intelligence (AI) driven information systems (e.g., [21]). Such efforts have largely focused on the tasks that require specialised training (e.g., lesion detection), and less attention has been paid to tasks such as annotating reports using metadata, to determine the characteristics of useful reports, and generating useful reports.

Useful radiology reports need to include a lot of information. For instance, they may include the patient's clinical history, description of the imaging devices, associated imaging parameters, comparison to previous observations, etc [8].

[4] https://radreport.org/

However, there is no agreed upon metadata standard designed with Southern Africa in mind and in the case of Zambia, there are also no guidelines that specify what information ought to be included in radiology reports. Consequently, if one intended to deploy AI solutions for authoring useful radiology reports then they would be hampered by the lack of guidelines/standards.

In this paper, we aim to answer the following question thus creating terminology that can be used to determine how radiology reports ought to be written: what categories of information ought to be included in radiology reports?

To answer the question, we extracted information from templates that are published via the RadReport Template Library. We also gathered information from international guidelines and practising experts who published their work. Specifically, we analysed published radiology reporting guidelines — this was done by tabulating the recommendations by radiology bodies in North America, the European Union, and United Kingdom and searching for publications that build on those recommendations using popular academic search engines. We used the guidelines to create a list of radiology terms that ought to be included in a report. We do not limit ourselves to a single regional body (e.g., European Society of Radiology) for the requirements so as to determine a truly international standard for radiology metadata. We have been able to collect 3522 terms covering different types of radiology reports. These terms represent an international collection that can be used by Southern African radiologists to design archetypes for structured report templates. The resulting dataset is released at `https://zenodo.org/record/8377480`.

The rest of the paper is structured such that Section 2 will introduce existing work on radiology metadata and report generation/authoring, and demonstrate that there is no work that aims to characterise radiology reports, especially for Zambia, Sections 3 will discuss the method used to create a list of the information units, Section 4 will present the dataset, Section 5 demonstrates the utility of the dataset via a use case, and Section 6 concludes.

2 Related work

The Digital Imaging and Communications in Medicine[5] international standard is the closest thing to a radiology metadata and terminology standard. The twenty-three part standard includes Information Object Definitions (IODs) that can be used to specify information about a patient, machines used, planned procedure(s), etc. One can use numerous attributes in IODs to specify metadata. For instance, when describing the patient, the attributes[6] *Patient's Name*, *Patient's Sex*, *Issuer of Patient ID* are three examples of the many available attributes. Nonetheless, the standard does not provide an exhaustive list of attributes that be used to form a metadata standard for the information that ought to be included in reports. Park et al. [17] have attempted to address this

[5] `https://www.dicomstandard.org/`

[6] `https://dicom.nema.org/medical/dicom/current/output/chtml/part04/sect_Q.4.3.html`, Accessed: 11 July 2023

issue by extending them using terminology extracted from RadLex[7]. In order to support Southern African countries as well, there is need to extend the work beyond terminology developed within the auspices of the Radiology Society of North America.

There is a long history of medical document generation in Natural Language Generation (NLG), a subfield of AI focused on generating text from non-linguistic inputs [4]. Since such work needs to determine the information that ought to be included in the generated reports, we now turn to the field to determine whether terminology has been created as part of such efforts. Most existing NLG work either explores the use of traditional techniques to generate documents meant to assist human in decision making, planning, etc. (e.g., health information management and dissemination [6,10,11,7]) or generating reports while identifying ailments from scans (e.g., transformer-based models trained on the the MIMIC-CXR dataset for purposes of generating clinical correct reports [9]). In this branch of research, Byamugisha's [3] work is the most relevant to the Zambian use case since they focus on producing text for an African audience. However, the work focuses on producing prescriptions; hence it is not useful for radiology reports.

In addition to the work in NLG, the work done on structured radiology reports (see [14]) is relevant. Such work has yielded templates, a number of which are findable via the Radiological Society of North America's RadReport template library, that have the potential to be useful according to some practitioners [1]. However, such efforts have made no attempt to produce a comprehensive dataset of the terms that ought to be included in reports. Instead, such research focuses on the following tasks, among many: specification of the structure of reports (e.g., [15]), perspectives on the benefits and uses of structured reports (e.g., [18]), and comparisons of structured vs. unstructured reports (e.g., [12]).

Since there is no existing work on creating a dataset of radiology terms, we now turn to discuss the method used to identify what ought to be included in radiology reports.

3 Methods and materials

In pursuit of an answer for the research question, we gather the terms from two sources: (1) templates published by Radiological Society of North America and vetted by an international panel of experts and (2) international guidelines and practising radiologists who published their work in journals, conferences, etc.

We downloaded all the historical structured reporting templates written in English from the RadReport template library[8]. We then extracted all the terms by gathering the text from the labels, sections, headers, and input tags of the HyperText Markup Language (HTML) templates using Python[9]. While the dataset can be supplemented using reports generated by Large Language Models

[7] https://www.rsna.org/practice-tools/data-tools-and-standards/radlex-radiology-lexicon

[8] https://radreport.org/

[9] https://www.python.org/

(LLMs), we opted against doing to ensure that we only consider reports authored by trained professionals.

We then cleaned the resulting units by manually removing incomplete chunks (e.g., *L/min (normal: male = 2.82-8.82 L/min; female = 2.7-6.0 l/min))*, automatically removing colons at the ends of units (e.g., *LACRIMAL GLANDS:*), and removing duplicates. We also collected papers that discuss the structure and content of useful medical reports and dictations. We found them via Google Scholar by selecting candidates based on the titles of the publications that cite the guidelines and recommendations published by the European Society of Radiology [5], Radiological Society of North America Radiology Reporting Committee [8], and The Royal College of Radiologists [20]. This produced 66 candidate publications. Since Zambia does not conduct all types of imaging examinations, we sought to limit the analysis to examination types that are within the capacity of the University Teaching hospitals in Zambia.

[EZ], a medical doctor specialising in Diagnostic Radiology , read the abstracts of the candidate papers and eliminated publications that focus on examination types that are outside the capability of the hospitals under consideration. Specifically, **[EZ]** marked each paper with *Capacity exists* or *Capacity does not exist* and indicated whether they usually conduct such examinations on a 5-point Likert scale. Out of the 66 papers, 11 described examinations for which there is no capacity, hence only making 55 eligible for further analysis.

We manually analysed the 55 papers, extracted the information that they specify ought to be in a report, created a list following an *ad hoc* methodology that is influenced by the one described in [13]:

1. Decide on a meta-characteristic: Knowledge included in radiology reports by medical professionals
2. Specify ending conditions for how to determine information used in taxonomy creation: All objects or a representative sample of objects have been examined using the relevant dimensions. No new dimensions or characteristics were added in the last iteration.
3. Create characteristics based on the meta-characteristic, determine dimensions for analysis, read a sample of the publications, and extract information.
4. If there are papers from which information has not been extracted go to Step 3.
5. Compile all the extracted information into a spreadsheet.

This process resulted in a comprehensive list of terms that ought to be included in radiology reports.

4 Radiology terminology

We obtained 266 structured radiology templates from the RadReport template library[10], encoded using HTML, that are created in English. After extracting and cleaning the resulting data, we obtained 3199 terms. They covered a range of topics and examples include the following:

[10] https://www.radreport.org/

1. *Presence of bronchial wall thickening*: Notes on whether wall thickening was detected in a scan.
2. *Severity of symptoms*: A discussion of seriousness of the symptoms experienced by the patient.

When extracting information from guidelines and publications, we used the meta-characteristic to create the characteristics:

(a) Doctors sometimes refer to a list of body parts vs. a single organ. For instance, they refer to bladder emptying and post-void residual volume simultaneously.
(b) Doctors group the terms via meaningful categories. For instance, they may need a category for information that pertains to a person, family, disease, equipment, etc.

The purpose of these characteristics is to guide the extraction and categorisation of the information. They are used to guide the identification of terms that may be obscured by grouping terminology. The characteristics were used to identify the following dimension to be used to categorise the terms extracted from the papers:

Property type: Binominal property that can take on the values 'categorisation' or 'primitive'. For instance, it can be used to mark that the information unit "Administrative information" is used to group other units while "Gender" is primitive as it takes on a single value at a time.

We read each of the 55 papers, extracted the terms, and annotated each term using the *Property type* attribute. For instance, [19] specify that structured bone marrow report must include the following information, among others: demographic data (*categorisation*) and diagnostic suspicion (*primitive*). Demographic data is a categorisation because it is used to group demographic information about a patient (e.g., gender, date of birth, age at diagnosis, etc.). Diagnostic suspicion as primitive as it lists information for which there is no value in splitting into its constituents.

We then separated the two types of terms into their respective tables after annotating all the papers. We then combined units that refer to the same element. For instance, we combine the following units *Demographic data*, *Patient identification*, *Patient demographics*, *Patient characteristics*, and *Patient details*. We also split units that were presented as one in the papers but refer to different elements. For instance, we split *Liver and spleen capsule* into *Liver capsule* and *Spleen capsule*. This process produced a total of 323 terms.

Numerous terms in the list are related to each other (e.g., a patient has an identification number), but the information is not found in the list. We enriched the list by adding SNOMED codes as a way to introduce the information and as a quality assurance measure. Specifically, two undergraduate students independently annotated the datasets with SNOMED codes. This followed a 30-minute training session introducing the extracted information, SNOMED CT, and demonstrating the annotation task. For all units where the two annotators do not have the same id, the disagreement was resolved by **[ZM]** and **[EZ]** concurrently. The whole process yielded in 232 annotated terms and 91 un-annotated terms.

5 Structured report templates use case scenario

Designing an NLG system to automatically generate structured radiology reports requires well-designed report templates. Templates differ according to the modality used and specific examination conducted. Generally, radiologists agree on the essential components to include [16,22]. However, variations exist across individual radiologists, departments and institutions with respect to content. The most consistent essential components of a radiology report include patient details, history, technique, findings, conclusion, and recommendations. Nevertheless, there is no discernible consensus when it comes to the rest of the elements. In the Zambian case, we have observed that radiologists adopt templates learnt from their respective past training institutions and pass them on to those training under them, irrespective of their suitability. Although such reports conform to internationally recommended standards to an extent, minor variations are evident. When authoring reports for a Brain CT scan, a hypothetical US-trained radiologist may find the template produced by the American Society of Neuroradiology and Radiological Society of North America[11] useful. However, the template does not specify that *Infarction*, *entrapment hydrocephalus*, *skull fracture*, *tonsillar herniation* must also be reported. Since *infarction*, brain tissue death as a result of reduced blood circulation to an area of the brain, is the more common cause of stroke, followed by a *haemorrhage* (a term that is included in the template), the resulting report would make it near impossible to confirm/exclude ischaemic stroke in an emergency CT. The dataset presented in this paper would solve such challenges as Zambian radiologists can design archetypes of what terms must be included in a report thus yield high report quality reports and remove guesswork.

6 Conclusion

We have created the first comprehensive list of radiology terminology that is not biased towards a specific region (e.g., North America). We demonstrated the utility of the dataset via a hypothetical use case, i.e., instead of using published structured reports that are partially suitable for Southern African countries, the dataset can be used to design archetypes for report templates thus potentially improving patient outcomes. Future work includes annotating existing useful radiology reports authoring by Zambian radiologists to enable the creation of AI-based assistive authoring tools.

Acknowledgements This study was conducted as part of a much larger "Enterprise Medical Imaging in Zambia" project, funded through generous grants from Google Research, Data Science Africa and The University of Zambia. We are very appreciative of this funding.

[11] `https://radreport.org/home/4/2009-12-01%2000:00:00`. Accessed: 12 July 2023

References

1. Bosmans, J.M.L., Peremans, L., Menni, M., De Schepper, A.M., Duyck, P.O., Parizel, P.M.: Structured reporting: if, why, when, how—and at what expense? results of a focus group meeting of radiology professionals from eight countries. Insights into Imaging **3**(3), 295–302 (Jun 2012)
2. Bwanga, O., Sichone, J.M., Sichone, P.N., Kazuma, Y.B.: Image interpretation and reporting by radiographers in Africa: Findings from the literature review and their application to Zambia. Medical Journal of Zambia **48**(2), 125–135 (2021)
3. Byamugisha, J.: Ontology verbalization in agglutinating Bantu languages: a study of Runyankore and its generalizability. Ph.D. thesis, Department of Computer Science, University of Cape Town, South Africa (2019)
4. Dale, R., Reiter, E.: Building Natural Language Generation Systems. Cambridge University Press (2000)
5. European Society of Radiology: Good practice for radiological reporting. guidelines from the European Society of Radiology (ESR). Insights into imaging **2**, 93–96 (2011)
6. Gatt, A., Portet, F., Reiter, E., Hunter, J., Mahamood, S., Moncur, W., Sripada, S.: From data to text in the neonatal intensive care unit: Using NLG technology for decision support and information management. AI Commun. **22**(3), 153–186 (2009). https://doi.org/10.3233/AIC-2009-0453, https://doi.org/10.3233/AIC-2009-0453
7. Hunter, J., Freer, Y., Gatt, A., Reiter, E., Sripada, S., Sykes, C.: Automatic generation of natural language nursing shift summaries in neonatal intensive care: Bt-nurse. Artif. Intell. Medicine **56**(3), 157–172 (2012). https://doi.org/10.1016/j.artmed.2012.09.002, https://doi.org/10.1016/j.artmed.2012.09.002
8. Kahn Jr, C.E., Langlotz, C.P., Burnside, E.S., Carrino, J.A., Channin, D.S., Hovsepian, D.M., Rubin, D.L.: Toward best practices in radiology reporting. Radiology **252**(3), 852–856 (2009)
9. Lovelace, J.R., Mortazavi, B.: Learning to generate clinically coherent chest X-Ray reports. In: Cohn, T., He, Y., Liu, Y. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020. Findings of ACL, vol. EMNLP 2020, pp. 1235–1243. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.findings-emnlp.110, https://doi.org/10.18653/v1/2020.findings-emnlp.110
10. Mahamood, S., Reiter, E.: Generating affective natural language for parents of neonatal infants. In: Gardent, C., Striegnitz, K. (eds.) ENLG 2011 - Proceedings of the 13th European Workshop on Natural Language Generation, 28-30 September 2011, Nancy, France. pp. 12–21. The Association for Computer Linguistics (2011), https://aclanthology.org/W11-2803/
11. Mahamood, S., Reiter, E.: Working with clinicians to improve a patient-information NLG system. In: Eugenio, B.D., McRoy, S., Gatt, A., Belz, A., Koller, A., Striegnitz, K. (eds.) INLG 2012 - Proceedings of the Seventh International Natural Language Generation Conference, 30 May 2012 - 1 June 2012, Starved Rock State Park, Utica, IL, USA. pp. 100–104. The Association for Computer Linguistics (2012), https://aclanthology.org/W12-1516/
12. Marcovici, P.A., Taylor, G.A.: Journal club: Structured radiology reports are more complete and more effective than unstructured reports. American journal of roentgenology **203**(6), 1265–1271 (2014)
13. Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. European Journal of Infor-

mation Systems **22**(3), 336–359 (2013). https://doi.org/10.1057/ejis.2012.26, https://doi.org/10.1057/ejis.2012.26
14. Nobel, J.M., van Geel, K., Robben, S.G.F.: Structured reporting in radiology: a systematic review to explore its potential. European Radiology **32**(4), 2837–2854 (Apr 2022)
15. Nobel, J.M., Kok, E.M., Robben, S.G.F.: Redefining the structure of structured reporting in radiology. Insights into Imaging **11**(1), 10 (Feb 2020)
16. Pahadia, M., Khurana, S., Geha, H., Deahl, S.T.I.: Radiology report writing skills: A linguistic and technical guide for early-career oral and maxillofacial radiologists. Imaging Science in Dentistry **50**(3), 269 (2020). https://doi.org/10.5624/isd.2020.50.3.269
17. Park, C., You, S.C., Jeon, H., Jeong, C.W., Choi, J.W., Park, R.W.: Development and validation of the radiology common data model (R-CDM) for the international standardization of medical imaging data. Yonsei Medical Journal **63**(Suppl), S74–S83 (Jan 2022)
18. Reiner, B.I., Knight, N., Siegel, E.L.: Radiology reporting, past, present, and future: The radiologist's perspective. Journal of the American College of Radiology **4**(5), 313–319 (2007)
19. Roca-Espiau, M., Valero-Tena, E., Ereño-Ealo, M.J., Giraldo, P.: Structured bone marrow report as an assessment tool in patients with hematopoietic disorders. Quant Imaging Med Surg **12**(7), 3717–3724 (Jul 2022)
20. Royal College of Radiologists: Standards for interpretation and reporting of imaging investigations. Tech. Rep. BFCR(18)1, Royal College of Radiologists (2018)
21. Tang, A., Tam, R., Cadrin-Chênevert, A., Guest, W., Chong, J., Barfett, J., Chepelev, L., Cairns, R., Mitchell, J.R., Cicero, M.D., Poudrette, M.G., Jaremko, J.L., Reinhold, C., Gallix, B., Gray, B., Geis, R., for the Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group: Canadian association of radiologists white paper on artificial intelligence in radiology. Canadian Association of Radiologists Journal **69**(2), 120–135 (2018). https://doi.org/10.1016/j.carj.2018.02.002
22. Wallis, A., McCoubrie, P.: The radiology report — are we getting the message across? Clinical Radiology **66**(11), 1015–1022 (Nov 2011). https://doi.org/10.1016/j.crad.2011.05.013