ORIGINAL RESEARCH

Image-Based Search in Radiology: Identification of Brain Tumor Subtypes within Databases using MRI-Based Radiomic Features

Marc von Reppert*, Saahil Chadha*, Klara Willms, Arman Avesta, Nazanin Maleki, Tal Zeevi, Jan Lost, Niklas Tillmanns, Leon Jekel, Sara Merkaj, IBSR Consortium, MingDe Lin, Karl-Titus Hoffmann , Sanjay Aneja, Mariam S. Aboian
IBSR Consortium: Manpreet Kaur, Divya Ramakrishnan, Khaled Bousabarah, Nicolas Linder, Sarah M. Jacobs, Francesca Conti, Gabriel Cassinelli Petersen, Ujjwal Baid, Ichiro Ikuta, Fatima Memon, Claudia Kirsch, Melissa Davis, Spyridon Bakas

ABSTRACT

BACKGROUND AND PURPOSE: Existing neuroradiology reference materials do not cover the full range of primary brain tumor presentations, and text-based medical image search engines are limited by the lack of consistent structure in radiology reports. To address this, an image-based search approach is introduced here, leveraging an institutional database to find reference MRIs visually similar to presented query cases.

MATERIALS AND METHODS: 295 patients (mean age and SD, 51 ± 20 years) with primary brain tumors who underwent surgical and/or radiotherapeutic treatment between 2000 and 2021 were included in this retrospective study. Semi-automated convolutional neural network-based tumor segmentation was performed, and radiomic features were extracted. The dataset was split into reference and query subsets, and dimensionality reduction was applied to cluster reference cases. Radiomic features extracted from each query case were projected onto the clustered reference cases, and nearest neighbors were retrieved. Retrieval performance was evaluated using mean average precision at k, and the best-performing dimensionality reduction technique was identified. Expert readers independently rated visual similarity using a five-point Likert scale.

RESULTS: t-Distributed Stochastic Neighbor Embedding with six components was the highest-performing dimensionality reduction technique, with mean average precision at 5 ranging from 78% to 100% by tumor type. The top five retrieved reference cases showed high visual similarity Likert scores with corresponding query cases (76% 'similar' or 'very similar').

CONCLUSIONS: We introduce an image-based search method for exploring historical MR images of primary brain tumors and fetching reference cases closely resembling queried ones. Assessment involving comparison of tumor types and visual similarity Likert scoring by expert neuroradiologists validates the effectiveness of this method.

ABBREVIATIONS: PCA = Principal Component Analysis; t-SNE = t-Distributed Stochastic Neighbor Embedding; UMAP = Uniform Manifold Approximation and Projection; PHATE = Potential of Heat-Diffusion for Affinity-Based Trajectory Embedding; G/A = Glioblastoma and Astrocytoma CNS World Health Organization Grade 4; A/O = Astrocytoma and Oligodendroglioma CNS World Health Organization Grades 2-3; PA = Pilocytic Astrocytoma; MEN = Meningioma; mAP@k = Mean Average Precision at k; CNN = Convolutional Neural Network

Received month day, year; accepted after revision month day, year.
From the University of Leipzig, Leipzig, Germany (M.V.R., K.L., K.T.F.); Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA (S.C., A.A.); Center for Outcomes Research and Evaluation, Yale School of Medicine, New Haven, CT, USA (S.C., S.A.); Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA (A.A.); Department of Radiology, Children's Hospital of Philadelphia, Philadelphia, PA, USA (N.M., M.S.A.); Center for Translational Imaging Analysis and Machine Learning, Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA (T.Z.); Department of Neurosurgery, Heinrich-Heine-University, Moorenstrasse 5, 40225 Duesseldorf, Germany (J.L.); Department of Diagnostic and Interventional Radiology, Medical Faculty, University Dusseldorf, 40225, Dusseldorf, Germany (N.T.); University of Duisburg-Essen, 45147 Essen, Germany (L.J.); DKFZ Division of Translational Neuro-oncology at the WTZ, German Cancer Consortium, DKTK Partner Site, University Hospital Essen, 45147 Essen, Germany (L.J.); University of Ulm, Ulm, Germany (S.M.); Visage Imaging, Inc., San Diego, CA, USA (M.D.L); Department of Therapeutic Radiology, Yale School of Medicine, New Haven, CT, USA (S.A.); IBSR consortium: Manpreet Kaur (Ludwig Maximilian University, Munich, Germany), Divya Ramakrishnan (Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA), Khaled Bousabarah (Visage Imaging GmbH, Berlin, Germany), Nicolas Linder (University of Leipzig, Leipzig, Germany), Sarah M. Jacobs (Center for Image Sciences, University Medical Center Utrecht, Utrecht, The Netherlands), Francesca Conti (Università di Roma - La Sapienza), Gabriel Cassinelli Petersen (University of Göttingen, Göttingen, Germany), Ujjwal Baid (Indiana University School of Medicine, Indianapolis, IN, USA), Ichiro Ikuta (Department of Radiology, Mayo Clinic Comprehensive Cancer Center, Phoenix, AZ), Fatima Memon (Carolina Radiology, Myrtle Beach, SC, USA), Claudia Kirsch (Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA), Melissa Davis (Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA), Spyridon Bakas (Indiana University School of Medicine, Indianapolis, IN, USA)

* These authors contributed equally.

Author Claudia F.E. Kirsch discloses receiving royalties from Primal Pictures and Pharma Intelligence, a Core Curriculum Grant from the American Society of Head and Neck Radiology, grant funding from the National Institutes of Health (R21NS122389, R01CA202911), research funding from the Yale Department of Radiology & Biomedical Imaging, and serving as a consultant for Karamedica Inc., with a stock option in the company.
Author MingDe Lin is an employee and stockholder of Visage Imaging, Inc. Author MingDe Lin receives funding from NIH/NCI R01 CA275188 and NIH/NCI R01 CA206180, and is a board member of Tau Beta Pi Engineering Honor Society.

Please address correspondence to Mariam S. Aboian, MD, PhD Department of Radiology, Children's Hospital of Philadelphia, 3401 Civic Center Blvd,

Philadelphia, PA 19104, USA; aboianm@chop.edu.

SUMMARY SECTION

PREVIOUS LITERATURE: Previous studies attempting image-based search for brain tumor MRI retrieval faced practical limitations. Techniques like visual bag of words models, image fusion, and transfer learning were leveraged but these earlier algorithms lacked practicality for clinical use as they required manual tumor delineation by expert neuroradiologists. The need for efficient, automatic methods to streamline image-based search has been emphasized, as reliance on traditional expert-driven segmentation impedes broad clinical application. Furthermore, the utility of retrieved search results was not validated by clinicians. There has been a gap in addressing brain tumor subtype classification with greater precision in retrieval.

KEY FINDINGS: This study introduces an algorithm combining deep learning-based automatic tumor segmentation with dimensionality reduction for image-based search. The method outperformed prior algorithms, achieving high retrieval scores across primary brain tumor entities. Its strong clinical validation offer significant potential for neuroradiology education and decision-making.

KNOWLEDGE ADVANCEMENT: This research advances knowledge by providing an efficient, automated method for image-based search, optimized for clinical use. Unlike previous methods, it eliminates the need for manual segmentation while demonstrating high retrieval accuracy. Its incorporation of radiomic feature clustering further enhances diagnostic capabilities, improving the integration of AI in neuroradiology practice.

INTRODUCTION

Existing neuroradiology reference materials largely feature classic imaging presentations of disease entities. Recent surveys reveal that the websites Radiopaedia, StatDx, and UpToDate are particularly popular among radiology trainees[1] and attendings.[2] However, these materials mainly offer curated textbook-style images and lack the full spectra of disease phenotypes. They also require prior knowledge of relevant search terms, rendering them ill-suited for more inexperienced users.

Though common reference materials include a limited scope of imaging features, institutional databases of historical medical imaging data can address this issue. Technological advancements have led healthcare centers to a tenfold increase in image acquisition between 1999 and 2010. [3] These databases encompass diverse imaging phenotypes with associated histopathologic diagnoses that are established and verified during the clinical care of patients, but most are currently underutilized for education and clinical decision-making.[4] Text-based search engines and machine learning methods are two popular methods for leveraging radiologic imaging archives, but both face concerns. Text-based search engines[5] may have more diversity in imaging pathology than classic reference materials, but they require well-crafted search terms. Such engines are familiar to most internet users, but despite their apparent uniquity, these engines often leverage controlled vocabularies, and standardizing free-text fields is complex and imperfect.[6] Moreover, like Radiopaedia and StatDx, text-based search methods rely on user's skills in crafting search terms. Machine learning offers a promising avenue to leverage archives; however, it faces interpretability and generalizability concerns. Despite the high performance of some primary brain tumor classifiers,[7,8] these models often lack explanations for diagnoses, inducing some apprehension among physicians about their integration at the point-of-care.[9] In addition, many models suffer from a significant drop in performance on out-of-distribution validation datasets from patient populations or healthcare institutions on which the models were not originally trained.[10]

There is a critical need for a method to retrieve historical medical images based on visual similarity to an image query. Similar methods have commonly been applied to face detection algorithms and other computer vision tasks outside of medicine.[11] And, previous work in this area[12–15] has been limited in clinical utility by requiring manual tumor segmentation. Here, we introduce an image-based search algorithm that utilizes a convolutional neural network (CNN) for semi-automatic tumor segmentation, radiomics for feature extraction, and dimensionality reduction for clustering. The algorithm automatically retrieves similar reference cases based on extracted query image features without any text input. This image-based search approach is evaluated on an institutional dataset of heterogenous primary brain tumors, while comparing dimensionality reduction techniques. Unlike previous image-based search algorithms, which grouped gliomas together,[12,13] this study is the first to separate glioma subtypes.[16] Performance is validated using established retrieval metrics and a visual similarity assessment by four board-certified neuroradiologists.

MATERIALS AND METHODS

Inclusion Criteria

Institutional review board approval was obtained, and a STROBE checklist was used for this retrospective observational study. 1,033 consecutive patients treated for primary brain tumors at the study institution between January 2000 and December 2021 were initially considered. Only patients who underwent surgical and/or radiotherapeutic treatment for histopathologically-confirmed Central Nervous System World Health Organization (CNS WHO)[16] glioblastoma and astrocytoma grade 4 (G/A), pilocytic astrocytoma (PA), astrocytoma and oligodendroglioma grades 2-3 (A/O), and meningioma (MEN) were included. Patient age was not limited. Pre-treatment MRIs from internal and external facilities were considered along with those from a pediatric low-grade glioma clinical trial (ClinicalTrials.gov: NCT01734512, PNOC001). Only MRI scans with T1-weighted, T1 contrast-enhanced (T1CE), and fluid attenuated inversion recovery (FLAIR) sequences were used. Patient scans with motion artifact or without hyperintensity on FLAIR were excluded. Overall, 295 patients were included.

Tumor Segmentation

Three volumes of interest (VOIs) were segmented: whole tumor on FLAIR as well as enhancing core and necrotic/cystic portions on T1CE. A pre-trained U-Net CNN-based automatic segmentation algorithm was utilized for adult-type diffuse gliomas[17] and manually corrected, as necessary. Pediatric-type gliomas and meningiomas were manually segmented. Segmentations were checked by a neuroradiologist with seven years of experience (M.S.A).

Pre-Processing and Radiomic Feature Extraction

PyRadiomics version 3.0.1 was used for image pre-processing and feature extraction.[18] Pre-processing included intensity Z-normalization, 3 mm isotropic voxel resampling, and image discretization. Default MRI parameters were used for feature extraction, which excluded sum average, maximal correlation coefficient, and neighboring gray tone difference matrix features due to redundancy. 14 shape-based features were extracted, and eight high and low-pass wavelet filters were applied to 18 first-order and 68 gray-level matrix features, resulting in 788 extracted features from each of the the two T1CE VOIs and one FLAIR VOI. The 2,364 features in total were utilized together for reference case retrieval.

Brain Tumor Clustering and Case Retrieval

Cases were randomly split into reference (85%) and query (15%) subsets. Reference cases' radiomic features were clustered onto the first *m* component vectors from dimensionality reduction. Four dimensionality reduction techniques were tested: principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP) and potential of heat-diffusion for affinity-based trajectory embedding (PHATE).[19–22] Two to ten included components vectors (*m*) were tested. Each query case's radiomic features were mapped onto the reduced feature space, and its most similar reference cases were retrieved through a nearest neighbor search. Figure 1 displays a workflow summary.

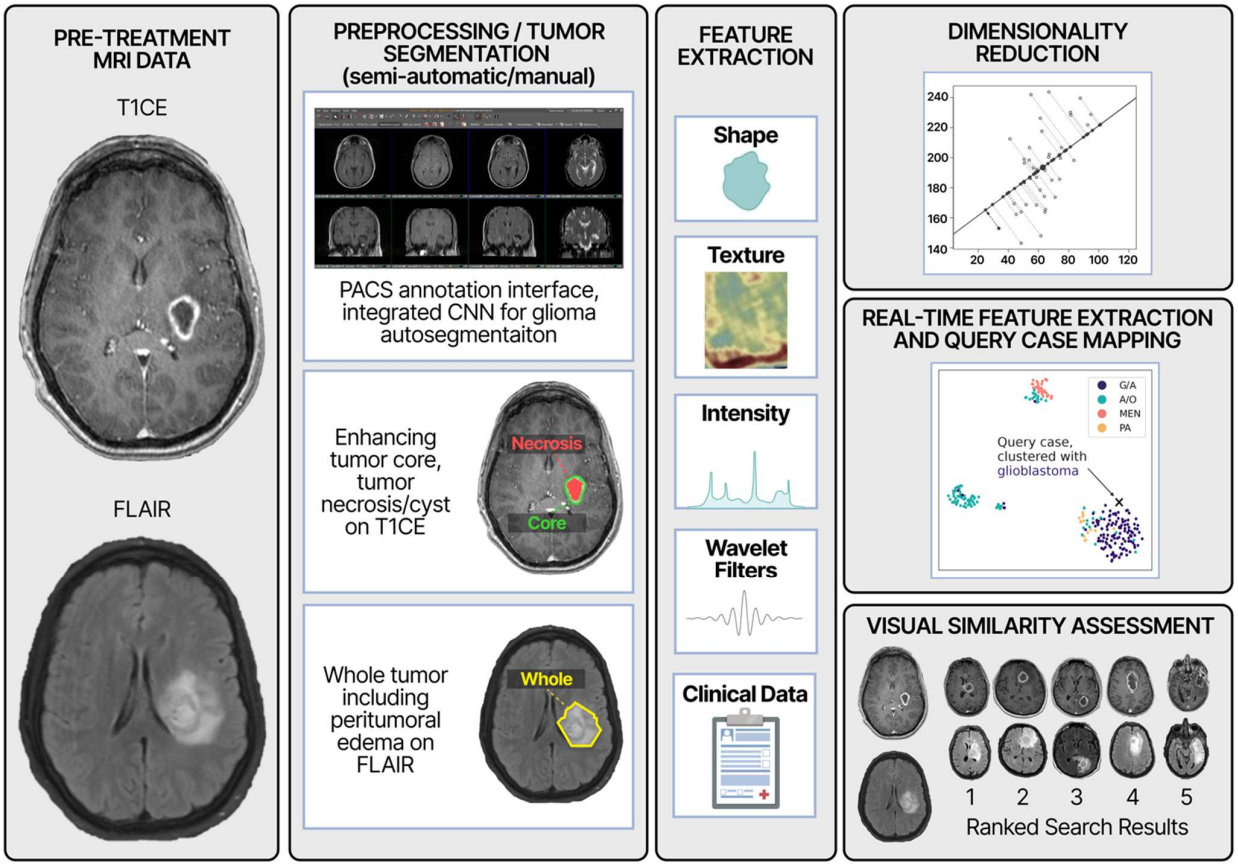

FIG 1. PACS-integrated Workflow for Volume of Interest Segmentation, Feature Extraction, Dimensionality Reduction and Clustering of Query Case. The inputs to our workflow are fluid attenuated inversion recovery (FLAIR) and T1-weighted contrast-enhanced (T1CE) images. Semi-automated segmentation was performed within our PACS to segment the whole, enhancing, and necrotic/cystic components of the tumor. Then, 788 radiomic features were extracted from the segmented tumor components. These extracted radiomic features additionally underwent wavelet transformation and were combined with clinical data, including age and sex. Dimensionality reduction was subsequently performed on these features. Reference cases were then projected into dimensionality-reduced space, clustering similar cases together. Finally, for a query case, the nearest reference cases in the dimensionality-reduced space were identified.

Assessment of Retrieval Performance

Retrieval performance was evaluated using mean average precision at k (mAP@k), a well-established measurement of the relevance of retrieved results to a search query[11] previously used in other medical image retrieval studies.[12,13]

Precision at k represents the proportion of accurate matches among the top k retrieved cases for a query. For r accurate matches among the top k retrieved cases:

$$\text{precision at } k = \frac{r}{k}$$

For n query cases each with at least k retrieved reference cases, mAP@k is defined as:

$$\text{mAP@}k = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{k}\sum_{j=1}^{k}\text{precision at } j \text{ for query } i$$

In this study, mAP@5 was the principal metric, as users are less likely to make use of search results beyond the top five. The technique and number of included components for dimensionality reduction were optimized by maximizing mAP@5.

Visual Similarity Assessment

Four expert neuroradiologists (M.S.A, F.M., C.K., and I.I) with a minimum of seven years of experience independently assessed visual similarity. They compared each of the top five retrieved reference MRI lesions to the corresponding query case on a 5-point scale (Table 1) as introduced previously.[23] Figure 2 shows representative images from this assessment.

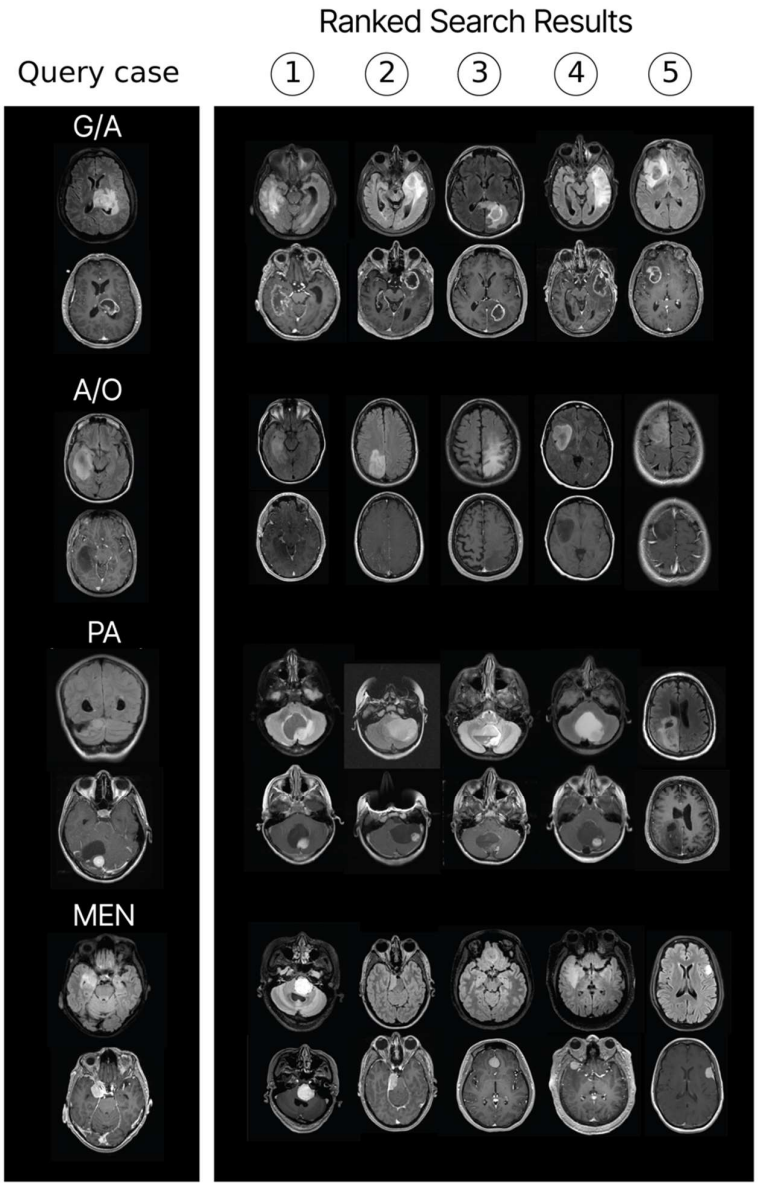

FIG 2. Axial FLAIR and T1 Contrast-Enhanced MRIs of Example Query Cases with Corresponding Five Top-Retrieved results. G/A indicates glioblastoma and astrocytoma CNS World Health Organization grade 4; A/O, astrocytoma and oligodendroglioma CNS

World Health Organization grades 2-3; MEN, meningioma; PA, pilocytic astrocytoma.

Statistical Analysis

Inter-rater reliability was assessed using Intraclass Correlation Coefficient (3, k) with Pingouin version 0.5.3.[24] Values were interpreted as previously reported: poor, < .5; moderate, .5 – .75; good, .75 – .9; and excellent, > .9.[25] A pairwise t-test was also performed to compare mAP scores between dimensionality reduction techniques. $P < .05$ was considered statistically significant.

Data Availability

Data used in this study are available from the authors upon request. Algorithm and analysis code are available through a GitHub repository (www.github.com/ImagineQuant/radiomics_clustering).

RESULTS

Patient Demographics and Data Split

Online Supplemental Data shows baseline demographic, clinical, and imaging data by tumor type. Of the 295 patients, 135 (46%) had G/A; A/O, 100 (34%); MEN, 40 (14%); PA, 20 (7%). Moreover, 103 patients' scans (35%) were acquired externally, and the remaining 192 patients' scans (65%) were acquired at the study institution. The reference subset contained 250 cases, and the query subset contained 45 cases.

Retrieval Performance

A visual representation of reference case clustering by PCA, t-SNE, UMAP, and PHATE can be found in Figure 3A. Maximum mAP@5 values for each dimensionality reduction technique were calculated across every number of included components (Figure 4, Table S1). t-SNE achieved the best performance across tumor types, though it was not statistically significantly better than other techniques ($P = .18$ [PHATE], .49 [UMAP], .89 [PCA]; Table S2). PCA and UMAP achieved comparable results to t-SNE for G/A, A/O, and MEN, but results differed for PA. PCA achieved a maximum mAP@5 of 73% for PA, and UMAP achieved 68%, compared to 78% for t-SNE. Across tumor types, the poorest performance was observed for PHATE.

Once t-SNE was identified as the highest-performing dimensionality reduction technique, the optimal number of included components was determined by comparing mAP@5 values by tumor type as the number of included components varied from two to ten (Figure 3B). Including additional t-SNE components beyond six did not result in performance improvement. Thus, the final retrieval model used in qualitative visual similarity scoring incorporated the top six t-SNE components.

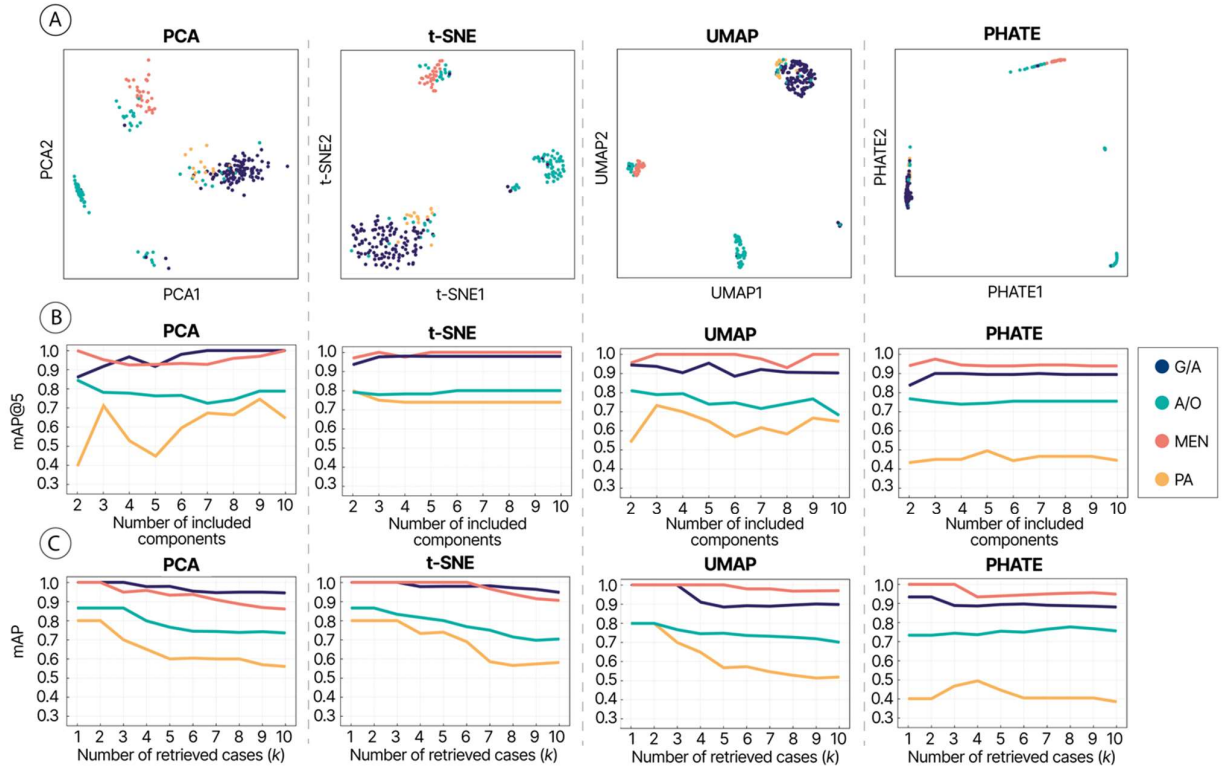

FIG 3. Retrieval Case Clustering and Performance. A, Retrieval case clustering of the first two components. B, mean average precision at 5 by varying number of components. C, mean average precision at k for six included components and varying number

of retrieved cases (k). t-SNE indicates t-Distributed Stochastic Neighbor Embedding; PCA, Principal Component Analysis; UMAP, Uniform Manifold Approximation and Projection; PHATE, Potential of Heat-Diffusion for Affinity-Based Trajectory Embedding. G/A indicates glioblastoma and astrocytoma CNS World Health Organization grade 4; A/O, astrocytoma and oligodendroglioma CNS World Health Organization grades 2-3; MEN, meningioma; PA, pilocytic astrocytoma.

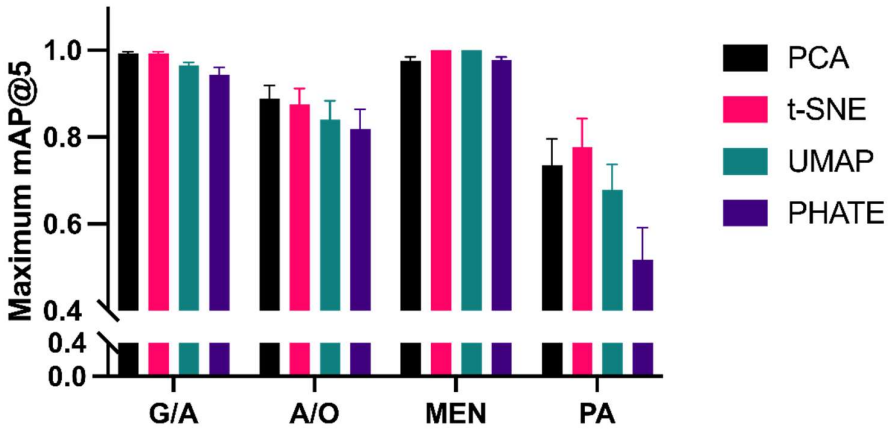

FIG 4. Maximum Mean Average Precision of Top 5 Retrieved Cases by Dimensionality Reduction Technique and Tumor Type. Error bars represent standard errors of the mean. PCA indicates Principal Component Analysis; t-SNE, t-Distributed Stochastic Neighbor Embedding; UMAP, Uniform Manifold Approximation and Projection; PHATE, Potential of Heat-Diffusion for Affinity-Based Trajectory Embedding. G/A indicates glioblastoma and astrocytoma CNS World Health Organization grade 4; A/O, astrocytoma and oligodendroglioma CNS World Health Organization grades 2-3; MEN, meningioma; PA, pilocytic astrocytoma.

Visual Similarity Assessment by Four Neuroradiologists

The top-ranked MRI search results demonstrated a mean visual similarity score of 4.27 ± .82 overall, with 32% 'similar' and 51% 'very similar' (Table 1). This score slightly diminished down the rankings, with a mean visual similarity score of 4.11 ± .96 for the top five results, with 28% 'similar' and 48% 'very similar' (Figure 5, Tables S3-4). The inter-rater reliability of the overall data demonstrated a good level of agreement, with an ICC (3, k) of .78 (95% CI [.65, .87], $P < .001$), providing a good level of reliability between raters. Overall, the top five retrieved reference lesions demonstrated high visual similarity with their corresponding query lesions.

Analysis across the four tumor groups exhibited high agreement levels for G/A and PA, each demonstrating an ICC (3, *k*) of .95 (G/A: $P < .001$, 95% CI [.90, .98]; PA: $P < .001$, 95% CI [.84, .99]). A/O tumors had similar ICC (3, k) values of .92 ($P < .001$, 95% CI [.83, .96]). While there was a moderate level of agreement among neuroradiologists in their similarity assessments of meningiomas (ICC [3, k], .55), this level of agreement was not statistically significant ($P = .05$; 95% CI [-.19, .87]).

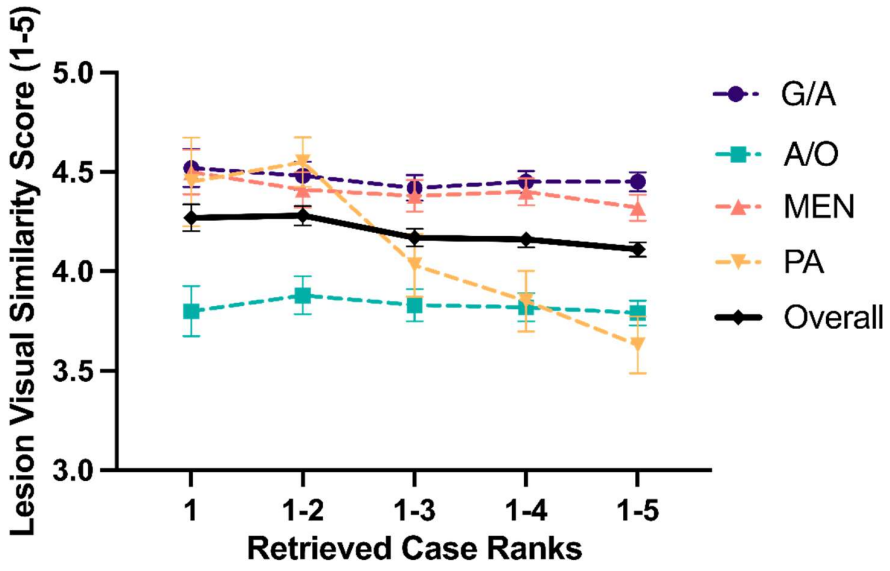

FIG 5. Likert Lesion Visual Similarity Scores of Top 1-5 Retrieved Cases by Tumor Type. Visual similarity scoring was conducted on a scale of 1 (very dissimilar) to 5 (very similar) based on the appearance of each retrieved reference lesion to the query lesion. The similarity scores for the top five cases were iteratively pooled in descending order to represent use in clinical practice. The percentage of cases rated 'very dissimilar' or 'dissimilar' are along the left axes, the percentage rated 'similar' or 'very similar' are along the right axes, and the percentage rated 'undecided' are along the center axis of each plot. G/A indicates glioblastoma

and astrocytoma CNS World Health Organization grade 4; A/O, astrocytoma and oligodendroglioma CNS World Health Organization grades 2-3; MEN, meningioma; PA, pilocytic astrocytoma.

DISCUSSION

Many radiologists and trainees rely on reference materials that focus on classic imaging features and require user expertise; the full spectrum of brain tumor imaging appearance is only learned over years of clinical practice. To address these challenges, we proposed an image-based search technique to retrieve primary brain tumor reference cases based on imaging appearance. This method has potential applications in differential diagnosis formulation, clinical and research database development, characterization of brain tumor subtype imaging features, and educational initiatives.

We demonstrated that dimensionality reduction and radiomic feature clustering is a promising method for performing image-based search of radiologic databases. t-SNE was the best-performing dimensionality reduction technique, achieving near-perfect mAP@5 scores for glioblastoma/astrocytoma grade 4 and meningiomas. Performance on pilocytic astrocytomas was comparatively lower at 78%, attributable to fewer cases and overlapping contrast enhancement features between PAs and glioblastomas/meningiomas. Queries for astrocytoma/oligodendroglioma grades 2-3 also had a comparatively lower mAP@5 score of 88%, likely due to overlapping imaging features on FLAIR. Expert neuroradiologists' visual similarity assessment demonstrated overall high performance of our approach, with a mean visual similarity score of 4.11 ± .99 out of 5 (76% 'similar' or 'very similar') for the top five retrieved results. This highlights the promising utility of image-based search in clinical practice.

Prior studies[12–15] attempted to perform image-based search for retrieval of primary brain tumor MRIs but were ill-suited for clinical implementation due to practical constraints. Though in practice our algorithm retrieves all reference cases in a ranked list, we utilized the mAP@5 metric to compare performance with this previous work. Retrieval performance of these earlier algorithms was fair; two algorithms used visual bag of words models[12,13], and others used image fusion[14] and transfer learning.[15] These achieved overall maximum mAP@5 values of 92% for meningiomas and 98% for gliomas. Our image-based search algorithm outperformed all these previous algorithms for meningioma retrieval (mAP@5, 100%), and average weighted performance on glioma subtypes was comparable (mean mAP@5, 93%). Crucially, however, all the earlier-mentioned algorithms required manual tumor segmentation from expert neuroradiologists, a known time-consuming and subjective process.[26] Clinical implementation of such algorithms would be limited by this bottleneck. We depart from these earlier, less practical algorithms by incorporating a CNN-based model for automatic tumor segmentation. Together with dimensionality reduction to minimize computational load, our image-based search algorithm is directly optimized for clinical utility. Furthermore, unlike earlier work, we clinically validate our search algorithm using expert neuroradiologist assessment.

There are a few limitations to this study. The introduced algorithm cannot distinguish visually similar primary brain tumor subtypes due to their grouping (e.g. low-grade astrocytoma/oligodendroglioma), a common challenge in image-based search focused on maximizing signal-to-noise ratios. Notably, to the authors' knowledge, this is the first study to perform image-based search of primary brain tumors while separating glioma subtypes. Future work will provide imaging characteristics to be used in our proposed pipeline to distinguish between tumors such as grade 2 and grade 3 IDH-mutant astrocytomas. Additionally, our algorithm does not account for anatomical location, yet retrieval performance remains strong regardless. Moreover, visual similarity assessment involved only four expert raters, but high ICC values indicate reliability of these results. Finally, although this study utilized a large, heterogenous dataset, multi-institutional testing will be used to further validate the proposed algorithm and determine the optimal degree of generalization and customization for scaling.

Through image-based search, neuroradiologists and trainees can identify reference cases that share visual similarities with their patients' scans. The performance of this approach has been validated through both quantitative and qualitative assessments, and it has shown promise for clinical application. Image-based search can provide valuable guidance in neuroradiology education, clinical decision-making, and research.

Table 1: Definition of the 5-point visual similarity scoring system.

Score	Definition
5: Very similar	The retrieved case is almost identical to the query case.
4: Similar	The retrieved case is almost identical, although there are differences in imaging features. However, it is subjectively classified to be the same disease.
3: Undecided	Some of the imaging features are similar, while some are different. Radiologically, the likelihood of the same disease is approximately half.
2: Dissimilar	The retrieved case has some similarities to the search case, but still looks different.
1: Very dissimilar	The retrieved case is very different from the search case and appears to be a different tumor etiology.

CONCLUSIONS

We propose a method to search historical MR images of primary brain tumors and retrieve reference cases highly similar to presented query cases. Evaluation through tumor type comparison and visual similarity scoring by expert neuroradiologists corroborated the high

performance of this approach. If translated into a clinical setting, image-based search will offer a reference material that encompasses the full diversity of imaging pathology.

ACKNOWLEDGMENTS

We would like to thank Dr. XP who was instrumental in project conception and design. This research was supported by the Richard K. Gershon Endowed Medical Student Research Fellowship and Yale School of Medicine Fellowship for Medical Student Research.

REFERENCES

1. Derakhshani A, Ding J, Vijayasarathi A. On-call radiology 2020: Where trainees look for help in a high stakes and time sensitive environment. *Clinical Imaging* 2021;77:219–23.
2. Niederhauser BD, Liaw K, McDonald RobertJ, et al. Pick Up a Book or "Google It?" A Survey of Radiologist and Trainee-Preferred References and Resources. *J Digit Imaging* 2014;27:26–32.
3. McDonald RJ, Schwartz KM, Eckel LJ, et al. The Effects of Changes in Utilization and Technological Advancements of Cross-Sectional Imaging on Radiologist Workload. *Academic Radiology* 2015;22:1191–8.
4. Akgül CB, Rubin DL, Napel S, et al. Content-Based Image Retrieval in Radiology: Current Status and Future Directions. *J Digit Imaging* 2011;24:208–22.
5. Demner-Fushman D, Kohli MD, Rosenman MB, et al. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association* 2016;23:304–10.
6. Sotomayor CG, Mendoza M, Castañeda V, et al. Content-Based Medical Image Retrieval and Intelligent Interactive Visual Browser for Medical Education, Research and Care. *Diagnostics (Basel)* 2021;11:1470.
7. Deepak S, Ameer PM. Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology and Medicine* 2019;111:103345.
8. Kokkalla S, Kakarla J, Venkateswarlu IB, et al. Three-class brain tumor classification using deep dense inception residual network. *Soft Comput* 2021;25:8721–9.
9. Geis JR, Brady AP, Wu CC, et al. Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. *Radiology* 2019;293:436–40.
10. Eche T, Schwartz LH, Mokrane F-Z, et al. Toward Generalizability in the Deployment of Artificial Intelligence in Radiology: Role of Computation Stress Testing to Overcome Underspecification. *Radiology: Artificial Intelligence* 2021;3:e210097.
11. Latif A, Rasheed A, Sajid U, et al. Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review. *Mathematical Problems in Engineering* 2019;2019:1–21.
12. Huang M, Yang W, Wu Y, et al. Content-Based Image Retrieval Using Spatial Layout Information in Brain Tumor T1-Weighted Contrast-Enhanced MR Images. Chen K, ed. *PLoS ONE* 2014;9:e102754.
13. Huang M, Yang W, Yu M, et al. Retrieval of Brain Tumors with Region-Specific Bag-of-Visual-Words Representations in Contrast-Enhanced MRI Images. *Computational and Mathematical Methods in Medicine* 2012;2012:e280538.
14. Dube S, El-Saden S, Cloughesy TF, et al. Content based image retrieval for MR image studies of brain tumors. *Conf Proc IEEE Eng Med Biol Soc* 2006;2006:3337–40.
15. Swati ZNK, Zhao Q, Kabir M, et al. Content-Based Brain Tumor Retrieval for MR Images Using Transfer Learning. *IEEE Access* 2019;7:17809–22.
16. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* 2021;23:1231–51.
17. Aboian M, Bousabarah K, Kazarian E, et al. Development of a workflow efficient PACS based automated brain tumor segmentation and radiomic feature extraction for clinical implementation (N2.003). *Neurology* 2022;98.
18. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research* 2017;77:e104–7.
19. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 1901;2:559–72.
20. Maaten L van der, Hinton G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 2008;9:2579–605.
21. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:180203426* 2018. [Epub ahead of print].
22. Moon KR, van Dijk D, Wang Z, et al. Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol* 2019;37:1482–92.
23. Hwang HJ, Seo JB, Lee SM, et al. Content-Based Image Retrieval of Chest CT with Convolutional Neural Network for Diffuse Interstitial Lung Disease: Performance Assessment in Three Major Idiopathic Interstitial Pneumonias. *Korean Journal of Radiology* 2021;22:281.
24. Vallat R. Pingouin: statistics in Python. *Journal of Open Source Software* 2018;3:1026.
25. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 2016;15:155–63.
26. Multi-Institutional Target Delineation in Oncology Group. Human-computer interaction in radiotherapy target volume delineation: a prospective, multi-institutional comparison of user input devices. *J Digit Imaging* 2011;24:794–803.

SUPPLEMENTAL FILES

Online Supplemental Data: Patient Demographics, Tumor Volumetrics, and Image Acquisition Details.

Characteristic	G/A	A/O	MEN	PA	Total
n	135 (46)	100 (34)	40 (14)	20 (7)	295
Age (years)†	62 ± 15	41 ± 15	59 ± 14	14 ± 13	51 ± 20
Sex					
Male	85 (63)	62 (62)	12 (30)	14 (70)	173 (59)
Female	50 (37)	38 (38)	28 (70)	6 (30)	122 (41)
Data Source					
Internal	89 (66)	57 (57)	39 (98)	7 (35)	192 (65)
External	46 (34)	43 (43)	1 (3)	13 (65)	103 (35)
Magnetic Field Strength					
3 T	65 (48)	43 (43)	24 (60)	7 (35)	141 (48)
1.5 T	68 (50)	55 (55)	12 (30)	12 (60)	149 (51)
Other	2 (1)	2 (2)	0 (0)	1 (5)	5 (2)
Presence of EC	130 (96)	37 (37)	35 (88)	20 (100)	227 (77)
Presence of N/C	132 (98)	25 (25)	0 (0)	20 (100)	182 (62)
WT Volume (cm3)†	135 ± 66	85 ± 80	46 ± 61	39 ± 29	100 ± 77
EC Volume (cm3)†	45 ± 26	24 ± 30	23 ± 29	5 ± 4	33 ± 29
N/C Volume (cm3)†	16 ± 12	11 ± 16	NA	18 ± 19	15 ± 14
Acquisition Parameters					
FLAIR TR†	8011 ± 1848	8116 ± 2046	7513 ± 1931	8524 ± 1639	8015 ± 1920
FLAIR TE†	161 ± 119	144 ± 103	201 ± 144	139 ± 82	159 ± 116
PGGE TR†	1527 ± 756	1182 ± 922	1773 ± 522	1326 ± 908	1530 ± 755
PGGE TE†	3.20 ± .64	3.77 ± 1.72	2.82 ± .39	3.07 ± .43	3.18 ± .88
PGSE TR†	637 ± 484	748 ± 634	503 ± 87	528 ± 94	652 ± 503
PGSE TE†	14.3 ± 5.32	16.2 ± 8.19	15.0 ± 4.24	15.5 ± 5.12	14.9 ± 6.18
Scanner					
GE Medical Systems	26 (19)	30 (30)	2 (5)	8 (40)	66 (22)
Signa Excite	8 (31)	10 (33)	1 (50)	2 (25)	21 (32)
Signa HDxt	10 (38)	10 (33)	0 (0)	1(13)	21 (32)
Other model	8 (31)	10 (33)	1 (50)	5 (63)	24 (36)
Siemens	99 (73)	62 (63)	38 (95)	11 (55)	210 (71)
Avanto	29 (29)	14 (23)	1 (3)	3 (27)	47 (22)
Verio	56 (57)	33 (53)	20 (53)	3 (27)	112 (53)
Other model	14 (14)	15 (24)	17 (45)	5 (45)	51 (24)
Other vendor	10 (7)	7 (7)	0 (0)	1 (5)	18 (6)

Note:—Numbers in parentheses are percentages. G/A indicates glioblastoma and astrocytoma CNS World Health Organization grade 4; A/O, astrocytoma and oligodendroglioma CNS World Health Organization grades 2-3; MEN, meningioma; PA, pilocytic astrocytoma; WT, whole tumor; EC, enhancing core; N/C, necrosis/cystic portion.

†Data are mean and standard deviation.

Table S1: Maximum Mean Average Precision of Top 5 Retrieved Cases by Dimensionality Reduction Technique and Tumor Type

Tumor Type	PCA	UMAP	PHATE	t-SNE	Number of Query Cases
G/A	99%	97%	94%	99%	15
A/O	89%	84%	82%	88%	15
MEN	98%	100%	98%	100%	10
PA	73%	68%	52%	78%	5

Note:– PCA indicates Principal Component Analysis; UMAP, Uniform Manifold Approximation and Projection; PHATE, Potential of Heat-Diffusion for Affinity-Based Trajectory Embedding; t-SNE, t-Distributed Stochastic Neighbor Embedding.

Table S2: Pairwise Performance Comparisons between Dimensionality Reduction Techniques

		T-Statistic	*P* Value
t-SNE	PCA	.14	.89
t-SNE	UMAP	.69	.49
t-SNE	PHATE	1.34	.18
PCA	UMAP	.56	.56
PCA	PHATE	1.27	.21
UMAP	PHATE	.70	.49

Note:– Pairwise t-tests were conducted to assess differences in mean average precision scores between the four techniques. A one-way ANOVA test yielded an F-statistic of .86 and *P* value of .46.

Table S3: Lesion Visual Similarity Scores of Top 1-5 Retrieved Results by Reader

Reader	Search Result Rank of Retrieved Lesions				
	1	1-2	1-3	1-4	1-5
1	4.71 ± .63	4.77 ± .58	4.70 ± .58	4.69 ± .67	4.64 ± .70
2	4.53 ± .76	4.56 ± .70	4.47 ± .81	4.46 ± .85	4.42 ± .88
3	4.00 ± 1.02	4.02 ± 1.07	3.89 ± 1.21	3.88 ± 1.24	3.79 ± 1.26
4	3.82 ± .89	3.76 ± .88	3.64 ± .94	3.62 ± .97	3.60 ± .99
Total	4.27 ± .91	4.28 ± .92	4.17 ± 1.02	4.16 ± 1.05	4.11 ± 1.07

Note:– Data are presented as the mean ± standard deviation. Visual similarity scoring was conducted on a scale of 1 (very dissimilar) to 5 (very similar) based on the appearance of each retrieved reference lesion to the query lesion. The similarity scores for the top five cases were iteratively pooled in descending order to represent use in clinical practice.

Table S4: Lesion Visual Similarity Scores of Top 1-5 Retrieved Results by Tumor Entity

Tumor Entity	Search Result Rank of Retrieved Lesions				
	1	1-2	1-3	1-4	1-5
G/A	4.52 ± .75	4.48 ± .78	4.42 ± .87	4.45 ± .83	4.45 ± .81

A/O	3.80 ± .97	3.88 ± 1.05	3.83 ± 1.09	3.82 ± 1.11	3.79 ± 1.10
MEN	4.50 ± .72	4.41 ± .79	4.38 ± .86	4.40 ± .85	4.32 ± .93
PA	4.45 ± 1.00	4.55 ± .78	4.03 ± 1.22	3.85 ± 1.37	3.63 ± 1.43
Overall	4.27 ± .91	4.28 ± .92	4.17 ± 1.02	4.16 ± 1.05	4.11 ± 1.07

Note:– Data are presented as the mean ± standard deviation. The similarity scores for the top 5 cases were iteratively pooled in descending order to represent use in clinical practice.