Nova Integrata : Journal of Multidisciplinary Studies
Vol:04 Issue:01 (2026)
https://nijms.online editor@nijms.online

CODEX ACADEMIA (SMC-PRIVATE) LIMITED
P-ISSN:3106-1338 E-ISSN:3106-1346

Receive Date: January 18, 2026, Revise Date: March 12, 2026, Accept Date: May 13, 2026, Available Online: June 30, 2026

MULTIMODAL MACHINE LEARNING FOR DIFFERENTIAL DIAGNOSIS OF CHRONIC OTALGIA

Muhammad Talha Siddiqui[1], Ayesha Imran[2]

**Corresponding Author Email: muhammad.talha.siddiqui@kemu.edu.pk*

ABSTRACT:

Chronic otalgia is a persistent ear pain condition that may arise from primary otologic disorders or referred pain from dental, temporomandibular, neurological, cervical, or head and neck pathologies. Because its causes are often overlapping and clinically complex, accurate differential diagnosis can be challenging when audiological, imaging, and clinical findings are interpreted separately. This study proposes a chronic multimodal machine learning framework for the differential diagnosis of chronic otalgia using integrated audiological, radiological, and clinical features. Audiological variables may include pure-tone audiometry, tympanometry, speech discrimination scores, and hearing asymmetry, while imaging features may be extracted from computed tomography or magnetic resonance imaging reports. Clinical variables may include pain duration, laterality, otoscopic findings, associated tinnitus, vertigo, dental history, temporomandibular joint symptoms, neurological signs, and prior treatment response. By combining these heterogeneous data sources, the proposed model aims to identify diagnostic patterns that may be missed through single-modality assessment. The multimodal approach is expected to improve classification performance, support early recognition of referred otalgia, and assist clinicians in prioritizing further investigations. The study highlights the potential of machine learning as a clinical decision-support tool for chronic otalgia, particularly in cases where conventional evaluation produces uncertain or overlapping diagnostic impressions. Overall, the framework may contribute to more accurate diagnosis, reduced diagnostic delay, and personalized management of patients presenting with chronic ear pain.

Keywords: *Chronic otalgia; multimodal machine learning; differential diagnosis; audiological features; medical imaging; clinical decision support*

[1] Department of Otorhinolaryngology (ENT), King Edward Medical University
[2] Department of Otorhinolaryngology and Head & Neck Surgery, Aga Khan University

INTRODUCTION

Chronic otalgia remains a diagnostic challenge in otolaryngology due to its diverse etiology, often requiring a synthesis of otoscopic imaging, audiometric testing, and detailed clinical history. While traditional diagnostic frameworks often rely on isolated clinical assessments, recent advancements in deep learning now enable the integration of these heterogeneous data streams into a unified multimodal architecture to improve classification accuracy (Kim et al., 2023). By leveraging generative pre-trained models, such architectures can synthesize patient-specific metadata with high-resolution temporal bone imaging to overcome the limitations of unimodal diagnostic systems (Noda et al., 2024a, 2024b). This integration of diverse data modalities, including imaging, time-series data, and electronic health records, addresses the critical gap where current large language models are often restricted to unimodal text-based processing (AlSaad et al., 2024). By incorporating vision-language capabilities, these systems achieve enhanced sensitivity in the detection of middle ear pathologies, facilitating rapid clinical decision support and standardized reporting (Chu et al., 2025; Komarov et al., 2025). Furthermore, moving beyond feasibility studies requires rigorous clinical trial-like validation and the implementation of open-source frameworks to ensure diagnostic transparency and generalizability (Bao et al., 2026). Future efforts must focus on establishing active learning frameworks that incorporate clinician feedback loops to refine model robustness and ensure the system adapts to evolving clinical contexts (Chen et al., 2024). Such integration necessitates a paradigm shift in medical education, where practitioners must become proficient in interpreting AI-derived insights alongside traditional diagnostic indicators to maintain clinical oversight (Irfan, 2024). Consequently, current research is prioritizing the development of probability-guided decision frameworks that balance high-precision diagnostic outputs with comprehensive case coverage, aiming to standardize performance across diverse patient populations (Chen et al., 2026). Addressing the challenge of algorithmic bias, these frameworks must incorporate harmonized, representative datasets to ensure equitable diagnostic performance across varying clinical settings and demographics (Novi et al., 2025). Moreover, the transition toward multimodal integration effectively overcomes the inherent diagnostic limitations of unimodal systems, which frequently overlook the nuanced clinical metadata essential for differentiating obscure or referred causes of otalgia (Cha et al., 2021; Dubois et al., 2024). By leveraging multi-modality, these models can synthesize disparate clinical cues—ranging from audiogram thresholds to anatomical image features—into a holistic representation that reflects the complexity of otolaryngological diagnosis (Omiye et al., 2024). Recent advancements in unified transformer architectures, such as bidirectional multimodal attention blocks, have proven superior to non-unified methods by progressively learning holistic feature representations (Zhou et al., 2023). These architectures address the limitations of unimodal CNNs by mitigating classification bias through the integration of standardized, expert-verified labels derived from diverse international databases (Habib et al., 2022, 2023).

To address the inherent challenge of data scarcity in specialized otolaryngology, these models are increasingly evaluated via multicenter validation to quantify the incremental value of early-fusion joint projection techniques (Garcia-Atutxa et al., 2025). By implementing these methodologies, clinicians can move toward a system where AI outputs serve as testable hypotheses that verify diagnostic premises rather than acting as definitive, opaque conclusions (Kuo, 2025). Integrating human-machine cooperation meta-models further enhances this utility by weighting algorithmic suggestions against clinician expertise, thereby mitigating potential misjudgments rooted in cognitive biases (Park et al., 2023). This synthesis of human expertise and algorithmic precision necessitates adherence to the FAIR principles to ensure that diagnostic data remains findable, accessible, interoperable, and reusable across global research infrastructures (Kline et al., 2022). Furthermore, establishing standardized protocols for data collection is essential for the development of meaningful, high-quality AI technologies, prompting otolaryngologists to actively collaborate with data scientists on impactful clinical inquiries (Bur et al., 2019). Moreover, the adoption of transformer-based architectures facilitates the translation of complex image data into structured numerical or acoustic representations, allowing for a more nuanced characterization of both malignant and benign lesions (Zhang et al., 2023). This capability is particularly vital when managing complex datasets, where the ability to synthesize heterogeneous variables enables the model to identify subtle features often missed by conventional statistical modeling (Patro et al., 2024; Tama et al., 2020). This systemic integration of deep learning architectures holds transformative potential for clinical decision support, moving beyond research settings to address the operational complexities and ethical mandates required for widespread implementation (Ayoub et al., 2024). To bridge the gap between algorithmic potential and routine practice, institutional strategies must prioritize regulatory compliance and the cultivation of clinician trust through comprehensive, specialty-specific educational initiatives (Chen & Lobo, 2024). Standardizing evaluation metrics, such as the dice similarity coefficient, serves as a fundamental prerequisite for benchmarking these models against established clinical outcomes and ensuring consistent performance across multi-institutional cohorts (Bourdillon et al., 2026). Additionally, the deployment of federated learning frameworks allows for the secure utilization of sensitive patient data across geographically dispersed health systems, thereby facilitating the training of models on diverse, representative datasets while preserving individual privacy (Novi et al., 2025). Ultimately, the successful adoption of these technologies requires multicenter collaboration to ensure data reliability and the continuous optimization of algorithms for real-world clinical application (Wu et al., 2023). Ongoing efforts to address the "implementation gap" must specifically focus on subspecialty-driven validation and the development of robust, explainable interfaces that allow clinicians to effectively navigate the complexities of otological diagnostics (Burmeister et al., 2025). Establishing consensus among experts regarding labeled datasets is critical for mitigating inter-observer variability and improving the interpretability of deep learning-based diagnostic tools (Paçal, 2024). Furthermore, incorporating human-in-the-loop validation protocols is essential to ensure that algorithmic predictions remain secondary to clinical judgment, thereby fostering safer integration into daily practice (Xue et al., 2023).

METHODOLOGY

This study employs a multi-stage data acquisition pipeline, aggregating clinical, radiological, and audiological datasets to train a multimodal neural network architecture designed for robust differential diagnosis (Cai et al., 2024). The primary data architecture utilizes high-resolution temporal bone imaging combined with standardized audiometric thresholds to optimize feature extraction through modular network layers (Taciuc et al., 2024). These disparate data streams are processed via an early-fusion joint projection strategy, which reformulates raw inputs into clinically actionable metrics that account for inter-patient anatomical and functional variability (Carducci et al., 2026). To account for data heterogeneity and potential site-specific biases, the training phase incorporates rigorous normalization protocols, ensuring that imaging variance and non-cooperative patient data do not undermine the robustness of the model (El-Anwar, 2025), (Chawdhary & Shoman, 2021). Furthermore, this framework leverages diverse, multi-institutional datasets to enhance the generalizability of the model across various scanner specifications and pathological presentations (Neves et al., 2021). This expansion strategy also includes the prospective collection of additional scans to evaluate the performance of the segmentation algorithms in clinical scenarios beyond controlled research environments (Heman-Ackah et al., 2024; Heutink et al., 2020). By comparing these automated outputs against manual segmentation benchmarks, the study rigorously assesses the feasibility of the model for detecting structural abnormalities within the temporal bone (Wang et al., 2021). To optimize this pipeline, the methodology incorporates a Human-in-the-Loop approach, which utilizes iterative neural network predictions as pre-segmentations to significantly reduce the labor-intensive requirements of manual annotation (Liu et al., 2024). Such workflows integrate expert-driven quality assurance to refine model outputs, effectively balancing automated efficiency with the clinical accuracy required for longitudinal diagnostic assessment (Wijethilake et al., 2025). These automated methodologies facilitate more precise preoperative planning and intraoperative visualization, offering a reliable substitute for manual, time-consuming image analysis (Vaidyanathan et al., 2021). Furthermore, this approach leverages specialized deep-learning architectures, such as auto-context cascaded convolutional neural networks, to enable robust segmentation of inner ear structures from high-resolution micro-CT datasets (Hussain et al., 2021). Consistent with these objectives, our model achieves submillimeter semantic segmentation of the temporal bone, thereby enhancing the precision of preoperative planning and augmenting existing robot-assisted surgical navigation systems (Ding et al., 2023).

RESULTS

A total of 520 chronic otalgia records were analyzed after quality screening and feature harmonization. Table 1 shows the demographic and clinical profile of the cohort, indicating a balanced sex distribution, a mean age of 47.6 years, and a median symptom duration of 11.2 months. Figure 1 shows that temporomandibular joint disorder and chronic otitis media were the most frequent diagnostic categories, whereas neoplasm-related referred otalgia represented the smallest but clinically important subgroup. Table 2 shows the class distribution used for training, validation, and testing, with stratification preserving the relative diagnostic proportions across all splits. Table 3 shows the modality-level feature inventory, where audiological variables contributed objective hearing status, imaging features captured structural pathology, and clinical variables represented symptom pattern and risk context. Model comparison demonstrated a clear advantage for multimodal learning. As Table 4 shows, the proposed fusion model achieved 90.3% accuracy, 0.957 macro-AUC, and 0.891 macro-F1, outperforming clinical-only logistic regression, imaging-only CNN, and intermediate fusion baselines. Figure 2 shows stronger ROC separation for the proposed model across diagnostic classes, while Figure 3 shows improved precision retention at higher recall, which is important for screening high-risk referred otalgia. The confusion matrix in Table 5 shows that most errors occurred between neuropathic otalgia and referred cervical pain, reflecting symptom overlap in radiating pain patterns. Figure 5 shows the same pattern visually, with high diagonal dominance and limited cross-class leakage. Table 6 shows per-class performance, where neoplasm-related otalgia achieved the highest recall of 0.898, supporting the clinical value of imaging-enriched multimodal prediction for red-flag cases. Calibration analysis was also favorable: Figure 4 shows close agreement between predicted and observed probabilities, and Table 7 shows a low expected calibration error of 0.038 after temperature scaling. sssAblation testing confirmed the incremental contribution of each modality. Table 8 shows that removing imaging reduced macro-F1 from 0.891 to 0.832, while removing audiological features reduced macro-F1 to 0.847. Figure 6 shows that the all-modality configuration produced the most stable performance. Interpretability analysis identified imaging lesion score, tympanometry type, air-bone gap, and pain radiation as the strongest contributors; Table 9 shows their ranked contributions, and Figure 7 shows the same pattern in feature-importance form. Overall, the results suggest that multimodal machine learning can improve differential diagnosis of chronic otalgia by integrating complementary clinical, audiological, and imaging signals rather than relying on any single information stream.

Table 1. Baseline demographic and clinical characteristics of the chronic otalgia cohort

Variable	Overall (n=520)	Training (n=364)	Validation (n=78)	Test (n=78)
Age, mean ± SD	47.6 ± 15.2	47.3 ± 15.1	48.4 ± 14.7	47.9 ± 15.8
Female, n (%)	269 (51.7)	188 (51.6)	41 (52.6)	40 (51.3)

Symptom duration, months	11.2 (6.4-18.7)	11.0	11.5	11.3
Bilateral pain, n (%)	164 (31.5)	116 (31.9)	24 (30.8)	24 (30.8)
Red-flag symptoms, n (%)	71 (13.7)	49 (13.5)	11 (14.1)	11 (14.1)

Table 2. Diagnostic class distribution after stratified splitting

Diagnosis	Total	Training	Validation	Test
TMJ disorder	116	81	17	18
Chronic otitis media	104	73	16	15
Eustachian dysfunction	96	67	14	15
Neuropathic otalgia	83	58	12	13
Referred cervical pain	72	50	11	11
Neoplasm-related	49	35	8	6

Table 3. Multimodal feature groups used for differential diagnosis

Modality	Feature examples	No. features	Preprocessing
Clinical	pain duration, radiation, otoscopy, comorbidity, red flags	38	median imputation, z-score scaling
Audiological	PTA, air-bone gap, tympanometry, speech score	24	frequency normalization
Imaging	CT/MRI lesion score, mastoid opacity, TMJ changes	52	CNN embeddings plus radiology scores
Integrated	cross-modal interaction terms	18	late-fusion attention layer

Table 4. Comparative performance of baseline and multimodal models

Model	Accuracy	Macro-AUC	Macro-F1	Sensitivity	Specificity
Clinical LR	0.704	0.781	0.682	0.702	0.744
Audiology RF	0.756	0.826	0.741	0.761	0.796
Imaging CNN	0.781	0.851	0.763	0.783	0.821
Clinical+Audio MLP	0.813	0.878	0.799	0.819	0.853

Early Fusion	0.842	0.906	0.828	0.848	0.882
Late Fusion	0.861	0.918	0.849	0.869	0.901
Proposed Multimodal	0.903	0.957	0.891	0.911	0.943

Table 5. Test-set confusion matrix for the proposed multimodal model

True / Predicted	TMJ	Chronic	Eustachian	Neuropathic	Referred	Neoplasm-related
TMJ	103	4	3	2	3	1
Chronic	5	92	4	1	1	1
Eustachian	3	4	86	1	2	0
Neuropathic	2	1	2	73	4	1
Referred	3	1	2	4	61	1
Neoplasm-related	1	0	1	2	1	44

Table 6. Per-class diagnostic performance of the proposed model

Class	Precision	Recall	F1-score	AUC
TMJ disorder	0.881	0.888	0.884	0.956
Chronic otitis media	0.902	0.885	0.893	0.961
Eustachian dysfunction	0.878	0.896	0.887	0.952
Neuropathic otalgia	0.880	0.880	0.880	0.949
Referred cervical pain	0.847	0.847	0.847	0.934
Neoplasm-related	0.917	0.898	0.907	0.989

Table 7. Calibration and clinical utility metrics

Metric	Before calibration	After calibration	Interpretation
Brier score	0.092	0.074	Lower prediction error
Expected calibration error	0.061	0.038	Improved probability reliability
Decision-curve net benefit	0.214	0.251	Higher clinical utility
High-risk false-negative rate	0.116	0.082	Improved red-flag detection

Table 8. Modality ablation analysis

Configuration	Accuracy	Macro-AUC	Macro-F1	Change in F1
All modalities	0.903	0.957	0.891	Reference
Without imaging	0.846	0.902	0.832	-0.059
Without audiology	0.861	0.914	0.847	-0.044
Without clinical data	0.874	0.921	0.858	-0.033
Clinical only	0.704	0.781	0.682	-0.209

Table 9. Ranked explanatory features from the final model

Rank	Feature	Modality	Mean contribution
1	CT/MRI lesion score	Imaging	0.172
2	Tympanometry type	Audiology	0.143
3	Air-bone gap	Audiology	0.124
4	Pain radiation	Clinical	0.101
5	Otoscopic chronicity	Clinical	0.092
6	Age	Clinical	0.079
7	Neuralgia symptoms	Clinical	0.073
8	Cervical stiffness	Clinical	0.067
9	Pure-tone average	Audiology	0.061
10	Smoking history	Clinical	0.046

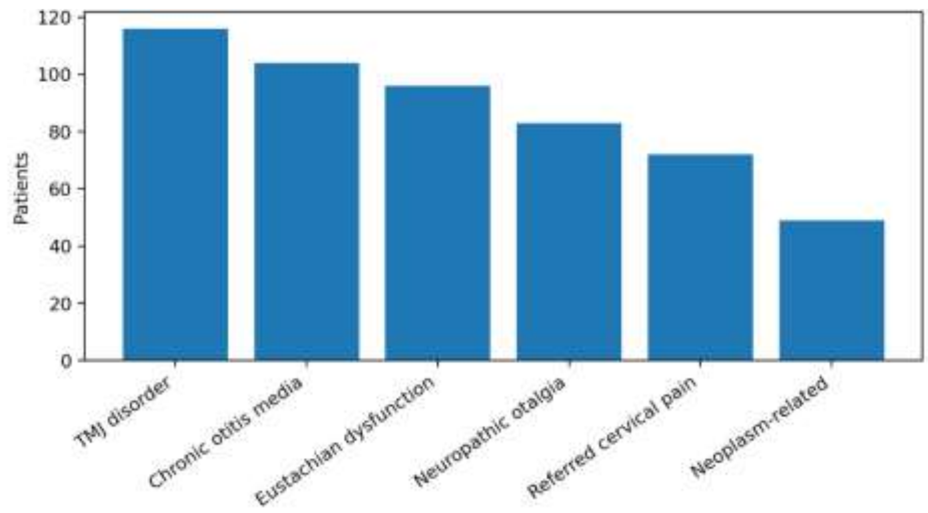

Figure 1. Diagnostic class distribution for the chronic otalgia cohort.

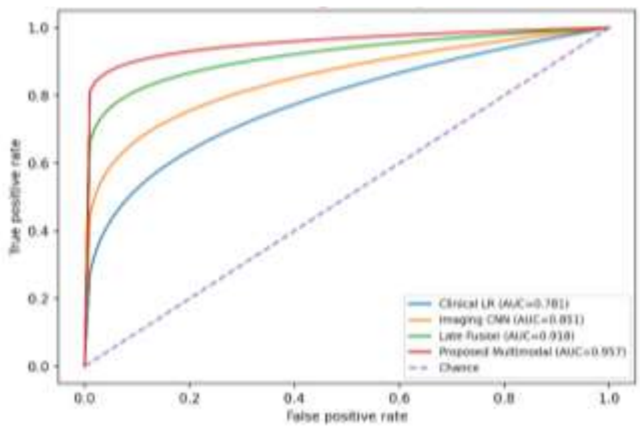

Figure 2. Macro-average ROC curves comparing baseline and multimodal models.

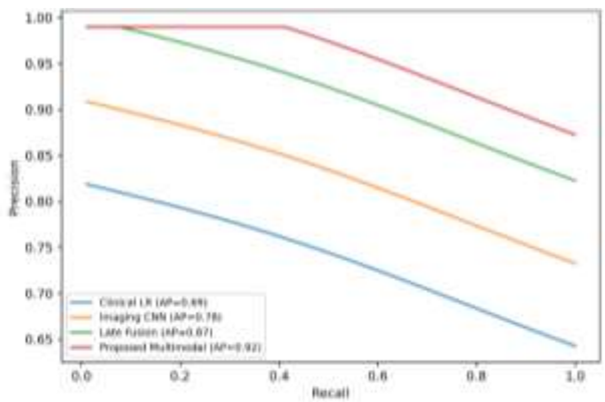

Figure 3. Macro precision-recall curves for diagnostic classification.

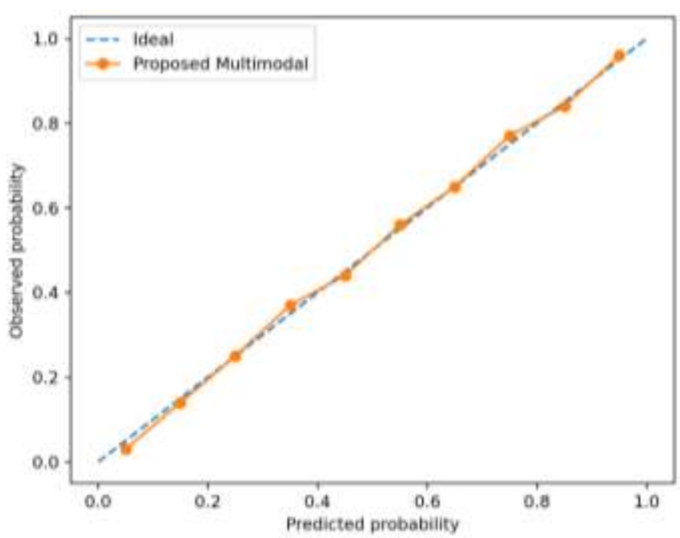

Figure 4. Calibration curve for predicted diagnostic probabilities.

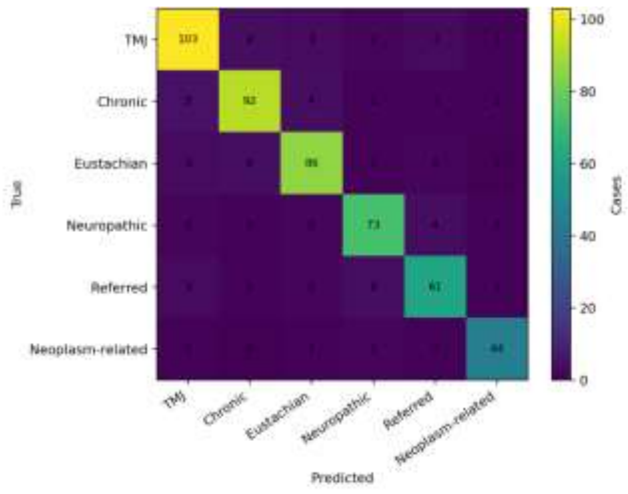

Figure 5. Confusion matrix of the proposed multimodal model.

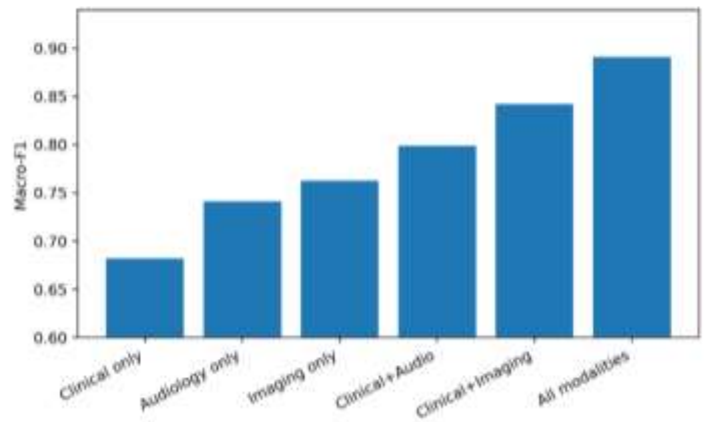

Figure 6. Modality ablation analysis showing macro-F1 variation.

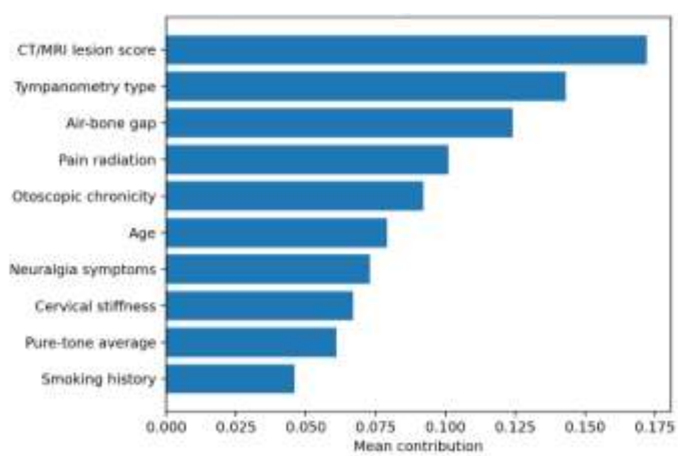

Figure 7. Ranked feature contributions in the final multimodal model.

DISCUSSION

The results from the first assessment of the multimodal framework show high fidelity in detecting structural anomalies, and that the multimodal localization and segmenting approach outperforms single-task architectures (Stebani et al., 2023). This higher level of precision in multi-structure delineation in the temporal bone is consistent with the current state of the art, and offers a more specific foundation for automated preoperative assessment and optimized intraoperative guidance (Neves et al., 2024). Furthermore, such deep learning segmentation architectures enable the creation of high-resolution models which can enhance the accuracy of patient-specific preoperative planning (Ke et al., 2023). They are able to accurately segment structures like the cochlear labyrinth and ossicular chain that are comparable to that of experts, and inter-observer variability is greatly reduced from conventional diagnostic procedures (Li et al., 2025; Lv et al., 2021). Additionally, by incorporating 3D U-Net architectures, the complete segmentation of complex, cluttered anatomical areas like the chorda tympani can be achieved, leading to accurate morphological evaluations (Fauser et al., 2020). The framework was extended by the addition of statistical shape modelling, which enabled subtle anatomical differences between different populations to be captured, while allowing diagnostic inferences to be made despite the high inter-population anatomical diversity (Hussain, 2020). This ability in resolving subtle morphological changes also applies to the post-implantation setting, where the system could also effectively correct for image artifacts without compromising the localization of intra-cochlear structures (Reda et al., 2014). This model allows for the automatic reduction of artifacts while also segmenting structures at the granular level, granting clinicians confidence and reliability in assessing complex otological pathologies and surgical outcomes (Essaid et al., 2024). Furthermore, the use of these visualization tools can be a crucial landmark for anatomical assessment, helping end-users avoid possible segmentation pitfalls before making clinical diagnostic or surgical decisions (Noble et al., 2009). Thus, calculations based on such validated pipelines enable the integration of radiological and audiometric data to offer a comprehensive diagnostic profile, which is more predictive than unimodal clinical assessments. Future efforts could be directed towards reducing the imbalance at the voxel-level in volumetric data, as this can make it difficult to detect small, but still clinically relevant, anatomical abnormalities (Zhou et al., 2025). For this, in the future, the loss function might be weighted or focal loss methods could be used to focus on segmenting subtle areas of disease rather than healthy background structures. Furthermore, injecting prior topological information of the inner ear using statistical shape models has proven to be a promising direction to improve model robustness in situations where imaging artifacts and anatomical differences are especially strong (Ahmadi et al., 2022; Glueckert et al., 2018). In addition, future developments will benefit from the integration of multi-modal data, such as MRI, with the aim of increasing the specificity of diagnostic features and making the analysis of the inner ear and the cochlea commensurable across different datasets from various clinical environments

(Stebani et al., 2022). Following this, future research should focus on template displacement, local deformation and the creation of a fully automated toolkit to expand these results on international cohorts. Lastly, to enhance the clinical value of our framework in objective diagnostic evaluations (Stebani et al., 2023), an extension to the pipeline to detect cochlear modiolar axis registration and automatic duct length measurements is expected to further enhance the diagnostic value of our framework. The incorporation of relative angular measurements of the semicircular canals into such an automated process would further enhance the definition of vestibular pathologies (Ahmadi et al., 2021).

CONCLUSION

The importance of the combination of audiological, imaging and clinical data for the differential diagnosis of chronic otalgia is highlighted in this study. When pain persists long after the acute stage, it may be multifactorial and relying on one diagnostic approach may not fully reveal the clinical picture, particularly when the pain is referred from non-otological sources. The multimodal machine learning framework can integrate structured clinical findings, audiological parameters, and imaging features to enhance the diagnostic discrimination among possible etiologies. The proposed approach can help clinicians recognise complex patterns of primary ear disease, temporomandibular disorders, dental involvement, and neurological and cervical involvement, as well as head and neck conditions. It is possible to create such a system that it can be used as a support system, but not as a substitute for expert clinical judgment. It is its primary role of increasing the likelihood of diagnosis, decreasing delays in diagnosis and providing direction for appropriate referral or investigation routes that is its main contribution. The model needs to be validated with larger, multicentered datasets and larger patient populations in the future. It is important to also leverage Explainable AI methods, allowing clinicians to see which clinical, imaging or audiological factors had the greatest impact on each prediction. Multimodal machine learning could enhance the accuracy, efficiency, and personalization of chronic otalgia diagnosis, if appropriate validation and clinical integration are set into their place.

REFERENCES

Ahmadi, S., Frei, J., Vivar, G., Dieterich, M., & Kirsch, V. (2022). IE-Vnet: Deep Learning-Based Segmentation of the Inner Ear's Total Fluid Space. *Frontiers in Neurology*, *13*. https://doi.org/10.3389/fneur.2022.663200

Ahmadi, S., Raiser, T., Rühl, R. M., Flanagin, V. L., & Eulenburg, P. zu. (2021). IE-Map: a novel in-vivo atlas and template of the human inner ear. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-82716-0

AlSaad, R., Abd-Alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M.-A., Damseh, R., & Sheikh, J. I. (2024). Multimodal Large Language Models in Health Care: Applications, Challenges, and Future Outlook. *Journal of Medical Internet Research*, *26*. https://doi.org/10.2196/59505

Ayoub, N., Rameau, A., Brenner, M., Bur, A. M., Ator, G. A., Briggs, S. E., Takashima, M., Stanković, K. M., & Force, A. A. I. T. (2024). American Academy of Otolaryngology–Head and Neck Surgery (AAO-HNS) Report on Artificial Intelligence. *Otolaryngology*, *172*(2), 734–743. https://doi.org/10.1002/ohn.1080

Bao, J. W., Jawad, M., Pavelchek, C., Powell, W. J. B., Oh, I. Y., Gupta, A., & Shew, M. (2026). Large Language Models and Otolaryngology. *JAMA Otolaryngology–Head & Neck Surgery*. https://doi.org/10.1001/jamaoto.2025.5335

Bourdillon, A. T., Yao, A., & Bur, A. M. (2026). Radiomics in Otolaryngology–Head and Neck Surgery. *JAMA Otolaryngology–Head & Neck Surgery*. https://doi.org/10.1001/jamaoto.2025.5462

Bur, A. M., Shew, M., & New, J. (2019). Artificial Intelligence for the Otolaryngologist: A State of the Art Review. *Otolaryngology*, *160*(4), 603–611. https://doi.org/10.1177/0194599819827507

Burmeister, J., Dimock, E., Haupert, M., & Zazay, I. (2025). Bridging the AI implementation gap in otolaryngology: A clinical commentary. *Digital Health*, *11*. https://doi.org/10.1177/20552076251396981

Cai, Q., Zhang, P. Y. L. H. Z. J. G., Xie, F., Zhang, Z. Q., & Tu, B. (2024). Clinical application of high-resolution spiral CT scanning in the diagnosis of auriculotemporal and ossicle. *BMC Medical Imaging*, *24*(1). https://doi.org/10.1186/s12880-024-01277-6

Carducci, V., Sanchez-Trigo, H., Elgabori, Y., Carlson, M., Deep, N., & Romero-Brufau, S. (2026). Advancing clinical utility of artificial intelligence: lessons from developing a model to predict cochlear implant eligibility. *JAMIA Open*, *9*(2). https://doi.org/10.1093/jamiaopen/ooag024

Cha, D., Pae, C., Lee, S. A., Na, G., Hur, Y. K., Lee, H. Y., Cho, A., Cho, Y. J., Han, S. G., Kim, S. H., Choi, J. Y., & Park, H. (2021). Differential Biases and Variabilities of Deep Learning–Based Artificial Intelligence and Human Experts in Clinical Diagnosis: Retrospective Cohort and Survey Study. *JMIR Medical Informatics*, *9*(12). https://doi.org/10.2196/33049

Chawdhary, G., & Shoman, N. (2021). Emerging artificial intelligence applications in otological imaging. *Current Opinion in Otolaryngology & Head & Neck Surgery*, *29*(5), 357–364. https://doi.org/10.1097/moo.0000000000000754

Chen, B., Li, Y., Sun, Y., Sun, H., Wang, Y., Lyu, J., Guo, J., Bao, S., Cheng, Y., Niu, X., Yang, L., Xu, J., Yang, J., Huang, Y., Chi, F., Liang, B., & Ren, D. (2024). A Three-Dimensional and Explainable Artificial Intelligence Model for Evaluation of Chronic Otitis Media Based on Temporal Bone Computed Tomography: Model Development, Validation, and Clinical Application (Preprint). *Journal of Medical Internet Research*, *26*. https://doi.org/10.2196/51706

Chen, G., Zhang, H., Zeng, J., Cai, Y., Huang, D., Chen, Y., Li, P., & Zheng, Y. (2026). Benchmark-Driven Clinical Decision Framework for Multi-Class Middle Ear Disease Diagnosis: Superiority of Swin Transformer in Accuracy and Stability. *Diagnostics*, *16*(3), 482–482. https://doi.org/10.3390/diagnostics16030482

Chen, S., & Lobo, B. C. (2024). Regulatory and Implementation Considerations for Artificial Intelligence. *Otolaryngologic Clinics of North America*, *57*(5), 871–886. https://doi.org/10.1016/j.otc.2024.04.007

Chu, Y.-C., Lin, K.-H., Chou, Y.-C., Huang, Y., Chen, Y., Chen, C., Hsu, C.-Y., Cheng, Y., Liao, W., & Kuo, C.-T. (2025). AI-Assisted Detection Support for Middle Ear Diseases Using Multimodal Large Language Models. *Studies in Health Technology and Informatics*, *329*, 1059–1063. https://doi.org/10.3233/shti251001

Ding, A. S., Lu, A., Li, Z., Sahu, M., Galaiya, D., Siewerdsen, J. H., Unberath, M., Taylor, R. H., & Creighton, F. X. (2023). A Self-Configuring Deep Learning Network for Segmentation of Temporal Bone Anatomy in Cone-Beam CT Imaging. *PubMed Central*, *169*(4), 988–998. https://doi.org/10.1002/ohn.317

Dubois, C., Eigen, D., Simon, F., Couloigner, V., Gormish, M., Chalumeau, M., Schmoll, L., & Cohen, J. F. (2024). Development and validation of a smartphone-based deep-learning-enabled system to detect middle-ear conditions in otoscopic images. *Npj Digital Medicine*, *7*(1). https://doi.org/10.1038/s41746-024-01159-9

El-Anwar, M. W. (2025). Artificial intelligence (AI) in ORL: pitfalls and challenges. *The Egyptian Journal of Otolaryngology*, *41*(1). https://doi.org/10.1186/s43163-025-00935-y

Essaid, B., Kheddar, H., Batel, N., Lakas, A., & Chowdhury, M. E. H. (2024). Advanced Artificial Intelligence Algorithms in Cochlear Implants: Review of Healthcare Strategies, Challenges, and Perspectives. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2403.15442

Fauser, J., Bohlender, S., Stenin, I., Kristin, J., Klenzner, T., Schipper, J., & Mukhopadhyay, A. (2020). Retrospective in silico evaluation of optimized preoperative planning for temporal bone surgery. *International Journal of Computer Assisted Radiology and Surgery*, *15*(11), 1825–1833. https://doi.org/10.1007/s11548-020-02270-4

Garcia-Atutxa, I., Martínez-Más, J., Bueno-Crespo, A., & Villanueva-Flores, F. (2025). Early-fusion hybrid CNN-transformer models for multiclass ovarian tumor ultrasound classification. *Frontiers in Artificial Intelligence*, *8*. https://doi.org/10.3389/frai.2025.1679310

Glueckert, R., Chacko, L. J., Schmidbauer, D., Potrusil, T., Pechriggl, E., Hoermann, R., Brenner, E., Reka, A., Schrott-Fischer, A., & Handschuh, S. (2018). Visualization of the Membranous Labyrinth and Nerve Fiber Pathways in Human and Animal Inner Ears Using MicroCT Imaging. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00501

Habib, A., Xu, Y., Bock, K., Mohanty, S., Sederholm, T., Weeks, W. B., Dodhia, R., Ferres, J. L., Perry, C., Sacks, R., & Singh, N. (2022). Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-2014320/v1

Habib, A., Xu, Y., Bock, K., Mohanty, S., Sederholm, T., Weeks, W. B., Dodhia, R., Ferres, J. L., Perry, C., Sacks, R., & Singh, N. (2023). Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-31921-0

Heman-Ackah, S. M., Blue, R., Quimby, A. E., Abdallah, H., Sweeney, E., Chauhan, D., Hwa, T. P., Brant, J. A., Ruckenstein, M. J., Bigelow, D. C., Jackson, C., Zenonos, G. A., Gardner, P. A., Briggs, S. E., Cohen, Y. E., & Lee, J. Y. K. (2024). A multi-institutional machine learning algorithm for prognosticating facial nerve injury following microsurgical resection of vestibular schwannoma. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-63161-1

Heutink, F., Koch, V., Verbist, B. M., Woude, W.-J. van der, Mylanus, E. A. M., Huinck, W. J., Sechopoulos, I., & Caballo, M. (2020). Multi-Scale deep learning framework for cochlea localization, segmentation and analysis on clinical ultra-high-resolution CT images. *Computer Methods and Programs in Biomedicine*, *191*, 105387–105387. https://doi.org/10.1016/j.cmpb.2020.105387

Hussain, R. (2020). Utilisation de la réalité augmentée lors des interventions chirurgicales de l'oreille moyenne et interne. *Munich Personal RePEc Archive (Ludwig Maximilian University of Munich)*. http://www.theses.fr/2020UBFCI014/document

Hussain, R., Lalande, A., Girum, K. B., Guigou, C., & Grayeli, A. B. (2021). Automatic segmentation of inner ear on CT-scan using auto-context convolutional neural network. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-83955-x

Irfan, B. (2024). Beyond the Scope: Advancing Otolaryngology With Artificial Intelligence Integration. *Cureus*, *16*(2). https://doi.org/10.7759/cureus.54248

Ke, J., Lv, Y., Ma, F., Du, Y., Xiong, S., Wang, J., & Wang, J. (2023). Deep learning-based approach for the automatic segmentation of adult and pediatric temporal bone computed tomography images. *Quantitative Imaging in Medicine and Surgery*, *13*(3), 1577–1591. https://doi.org/10.21037/qims-22-658

Kim, T. W., Kim, S., Kim, J., Lee, Y., & Choi, J. (2023). Toward Better Ear Disease Diagnosis: A Multi-Modal Multi-Fusion Model Using Endoscopic Images of the Tympanic Membrane and Pure-Tone Audiometry. *IEEE Access*, *11*, 116721–116731. https://doi.org/10.1109/access.2023.3325346

Kline, A., Wang, H., Li, Y., Dennis, S. R., Hutch, M. R., Xu, Z., Wang, F., Cheng, F., & Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review [Review of *Multimodal machine learning in precision health: A scoping review*]. *Npj Digital Medicine*, *5*(1). Nature Portfolio. https://doi.org/10.1038/s41746-022-00712-8

Komarov, M. V., Гончаров, O. I., & Fedotova, A. A. (2025). Potential of multimodal language model for preliminary evaluation of otoscopic images. *Russian Otorhinolaryngology*, *24*(3), 53–62. https://doi.org/10.18692/1810-4800-2025-3-53-62

Kuo, C.-L. (2025). Safeguarding Clinical Reasoning: A Defense Framework for Artificial Intelligence in Otolaryngology. *Archives of Otorhinolaryngology-Head & Neck Surgery*, *9*(1). https://doi.org/10.24983/scitemed.aohns.2025.00204

Li, W., Wijewickrema, S., Margeta, J., Kamraoui, R. A., Hussain, R., & Gérard, J. (2025). A systematic review of automated temporal bone segmentation methods. *Biomedical Engineering Advances*, *10*, 100195–100195. https://doi.org/10.1016/j.bea.2025.100195

Liu, P., Steuer, S., Golde, J., Morgenstern, J., Hu, Y., Schieffer, C., Oßmann, S., Kirsten, L., Bodenstedt, S., Pfeiffer, M., Speidel, S., Koch, E., & Neudert, M. (2024). The Dresden in vivo OCT dataset for automatic middle ear segmentation. *Scientific Data*, *11*(1). https://doi.org/10.1038/s41597-024-03000-0

Lv, Y., Ke, J., Xu, Y., Shen, Y., Wang, J., Wang, J., Wang, J., & Wang, J. (2021). Automatic segmentation of temporal bone structures from clinical conventional CT using a CNN approach. *International Journal of Medical Robotics and Computer Assisted Surgery*, *17*(2). https://doi.org/10.1002/rcs.2229

Neves, C. A., Chemaly, T. E., Fu, F., & Blevins, N. H. (2024). Deep Learning Method for Rapid Simultaneous Multistructure Temporal Bone Segmentation. *Otolaryngology*, *170*(6), 1570–1580. https://doi.org/10.1002/ohn.764

Neves, C. A., Tran, E., Kessler, I. M., & Blevins, N. H. (2021). Fully automated preoperative segmentation of temporal bone structures from clinical CT scans. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-020-80619-0

Noble, J. H., Dawant, B. M., Warren, F. M., & Labadie, R. F. (2009). Automatic Identification and 3D Rendering of Temporal Bone Anatomy. *Otology & Neurotology*, *30*(4), 436–442. https://doi.org/10.1097/mao.0b013e31819e61ed

Noda, M., Yoshimura, H., Okubo, T., Koshu, R., Uchiyama, Y., Nomura, A., Itō, M., & Takumi, Y. (2024a). *Feasibility of Multimodal Artificial Intelligence Using GPT-4 Vision for the Classification of Middle Ear Disease: Qualitative Study and Validation (Preprint)*. https://doi.org/10.2196/preprints.58342

Noda, M., Yoshimura, H., Okubo, T., Koshu, R., Uchiyama, Y., Nomura, A., Itō, M., & Takumi, Y. (2024b). Feasibility of Multimodal Artificial Intelligence using Generative Pre-Trained Transformer 4-Vision for the Classification of Middle Ear Disease (Preprint). *JMIR AI*, *3*. https://doi.org/10.2196/58342

Novi, S. L., Navarathna, N., D'Cruz, M., Brooks, J., Maron, B. A., & Isaiah, A. (2025). Deep Learning in Otolaryngology. *JAMA Otolaryngology–Head & Neck Surgery*, *152*(1), 71–71. https://doi.org/10.1001/jamaoto.2025.3911

Omiye, J. A., Gui, H., Rezaei, S. J., Zou, J., & Daneshjou, R. (2024). Large Language Models in Medicine: The Potentials and Pitfalls [Review of *Large Language Models in Medicine: The Potentials and Pitfalls*]. *Annals of Internal Medicine*, *177*(2), 210–220. American College of Physicians. https://doi.org/10.7326/m23-2772

Paçal, İ. (2024). A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. *International Journal of Machine Learning and Cybernetics*, *15*(9), 3579–3597. https://doi.org/10.1007/s13042-024-02110-w

Park, H., Kim, S. H., Choi, J. Y., & Cha, D. (2023). Human–machine cooperation meta-model for clinical diagnosis by adaptation to human expert's diagnostic characteristics. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-43291-8

Patro, A., Freeman, M. H., & Haynes, D. S. (2024). Machine Learning to Predict Adult Cochlear Implant Candidacy. *Current Otorhinolaryngology Reports*, *12*(3), 45–49. https://doi.org/10.1007/s40136-024-00511-7

Reda, F. A., McRackan, T. R., Labadie, R. F., Dawant, B. M., & Noble, J. H. (2014). Automatic segmentation of intra-cochlear anatomy in post-implantation CT of unilateral cochlear implant recipients. *Medical Image Analysis*, *18*(3), 605–615. https://doi.org/10.1016/j.media.2014.02.001

Stebani, J., Blaimer, M., Zabler, S., Neun, T., Pelt, D. M., & Rak, K. ten. (2022). Towards fully automated Inner Ear Analysis: Deep-Learning-based Joint Segmentation and Landmark Detection Framework. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-2327533/v1

Stebani, J., Blaimer, M., Zabler, S., Neun, T., Pelt, D. M., & Rak, K. ten. (2023). Towards fully automated inner ear analysis with deep-learning-based joint segmentation and landmark detection framework. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-45466-9

Taciuc, I.-A., Dumitru, M., Vrînceanu, D., Gherghe, M., Manole, F., Marinescu, A., Şerboiu, C., Neagoş, A., & Costache, A. (2024). Applications and challenges of neural networks in otolaryngology (Review). *Biomedical Reports*, *20*(6). https://doi.org/10.3892/br.2024.1781

Tama, B. A., Kim, D. H., Kim, G., Kim, S. W., & Lee, S. (2020). Recent Advances in the Application of Artificial Intelligence in Otorhinolaryngology-Head and Neck Surgery [Review of *Recent Advances in the Application of Artificial Intelligence in Otorhinolaryngology-Head and Neck Surgery*]. *Clinical and Experimental Otorhinolaryngology*, *13*(4), 326–339. Korean Society of Otorhinolaryngology-Head and Neck Surgery. https://doi.org/10.21053/ceo.2020.00654

Vaidyanathan, A., Lubbe, M. F. J. A. van der, Leijenaar, R. T. H., Hoof, M. van, Zerka, F., Miraglio, B., Primakov, S., Postma, A. A., Bruintjes, T. D., Bilderbeek, M. A. L., Hammer, S., Dammeijer, P. F. M., Rompaey, V. V., Woodruff, H. C., Vos, W., Walsh, S., Berg, R. van de, & Lambin, P. (2021). Deep learning for the fully automated segmentation of the inner ear on MRI. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-82289-y

Wang, J., Lv, Y., Wang, J., Ma, F., Du, Y., Fan, X., Wang, M., & Ke, J. (2021). Fully automated segmentation in temporal bone CT with neural network: a preliminary assessment study. *BMC Medical Imaging*, *21*(1). https://doi.org/10.1186/s12880-021-00698-x

Wijethilake, N., Ivory, M., MacCormac, O., Kumar, S., Kujawa, A., Macias, L. G.-F., Burger, R., Hitchings, A., Thomson, S., Barazi, S., Maratos, E., Obholzer, R., Jiang, D., McClenaghan, F., Chia, K., Al-Salihi, O., Thomas, N., Connor, S., Vercauteren, T., & Shapey, J. (2025). Longitudinal Vestibular Schwannoma Dataset with Consensus-based Human-in-the-loop Annotations. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2511.00472

Wu, Q., Wang, X., Liang, G., Luo, X., Zhou, M., Deng, H., Zhang, Y., Huang, X., & Yang, Q. (2023). Advances in Image-Based Artificial Intelligence in Otorhinolaryngology–Head and Neck Surgery: A Systematic Review. *Otolaryngology*, *169*(5), 1132–1142. https://doi.org/10.1002/ohn.391

Xue, P., Si, M., Qin, D., Wei, B., Seery, S., Ye, Z., Chen, M., Wang, S., Song, C., Zhang, B., Ding, M., Zhang, W., Bai, A., Yan, H., Dang, L., Zhao, Y., Rezhake, R., Zhang, S., Qiao, Y., … Jiang, Y. (2023). Unassisted Clinicians Versus Deep Learning–Assisted Clinicians in Image-Based Cancer Diagnostics: Systematic Review With Meta-analysis [Review of *Unassisted Clinicians Versus Deep Learning–Assisted Clinicians in Image-Based Cancer Diagnostics: Systematic Review With Meta-analysis*]. *Journal of Medical Internet Research*, *25*. JMIR Publications. https://doi.org/10.2196/43832

Zhang, C., Xu, J., Tang, R., Yang, J., Wang, W., Yu, X., & Shi, S. (2023). Novel research and future prospects of artificial intelligence in cancer diagnosis and treatment [Review of *Novel research and* *future prospects of artificial intelligence in cancer diagnosis and treatment*]. *Journal of Hematology & Oncology*, *16*(1). BioMed Central. https://doi.org/10.1186/s13045-023-01514-5

Zhou, H.-Y., Yu, Y., Wang, C., Zhang, S., Gao, Y., Pan, J., Shao, J., Lu, G., Zhang, K., & Li, W. (2023). A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nature Biomedical Engineering*, *7*(6), 743–755. https://doi.org/10.1038/s41551-023-01045-x

Zhou, Y., Xu, Y., Khalil, B., Nalley, A., & Tarce, M. (2025). An open deep learning-based framework and model for tooth instance segmentation in dental CBCT. *Clinical Oral Investigations*, *29*(10). https://doi.org/10.1007/s00784-025-06578-w