

Assessment of Olfactory Disorders Using Olfactometry and Artificial Intelligence

Yakubdjanov D. D.¹, Rajabov D. B.¹, Vokhidov U. N.²

¹Department of Otorhinolaryngology, The Tashkent State Medical University, Tashkent, Uzbekistan

²The Republican Specialized Scientific and Practical Medical Center Otolaryngology and Head and Neck Diseases, Tashkent, Uzbekistan

Abstract Artificial Intelligence in Olfactory Disorder Diagnosis: A Narrative Review. AI in Olfactory Diagnosis. Background: Olfactory dysfunction is a prevalent clinical problem significantly impairing quality of life. Conventional olfactometric methods remain subjective and poorly reproducible. The objective of this study was evaluated using artificial intelligence-based olfactory diagnostics. Materials and Methods: This narrative review analyzed published studies from 2019–2024 involving adult patients with olfactory dysfunction. Data were obtained from PubMed, Scopus, and Web of Science. Diagnostic tools included Sniffin' Sticks, UPSIT, MRI, EEG, and CT. Descriptive statistics were applied using SPSS version 26.0. A significance level of $p < 0.050$ was adopted. Results: AI-based diagnostic models integrating olfactometric, neuroimaging, and electrophysiological data demonstrated higher accuracy than conventional tests. Convolutional neural networks achieved diagnostic accuracy up to 94.600%, while machine learning models reached 87.000–92.000%. AI approaches improved early detection of subclinical dysfunction 1.500 times compared with standard methods ($p = 0.031$). Variability of repeated measurements decreased significantly ($p = 0.004$). Conclusion: Artificial intelligence significantly enhances diagnostic accuracy, objectivity, and reproducibility in olfactory dysfunction assessment and represents a promising tool for early detection and personalized patient monitoring.

Keywords Olfaction, Anosmia, Hyposmia, Artificial Intelligence, Machine Learning

1. Introduction

Olfaction is a fundamental sensory modality essential for environmental adaptation, nutrition, hazard detection, and emotional processing. Olfactory dysfunction significantly impairs quality of life, contributing to appetite disturbances, depression, and social maladaptation. Epidemiological data indicate a prevalence of 15–25% in the general population, increasing to over 40% among individuals older than 60 years. Postviral olfactory disorders, particularly following COVID-19, have emerged as a major public health concern.

Despite the availability of standardized diagnostic tools such as the Sniffin' Sticks test and the University of Pennsylvania Smell Identification Test (UPSIT), clinical assessment remains challenging due to subjectivity and limited reproducibility. Variability of repeated measurements may reach 20–30%, reducing diagnostic reliability. Traditional olfactometry does not fully account for structural and functional alterations of olfactory pathways. Neuroimaging studies have demonstrated correlations between olfactory dysfunction and reduced olfactory bulb volume, tract integrity loss, and cortical changes. However, integration of these parameters into routine diagnostics remains limited.

Artificial intelligence (AI) has gained increasing attention as a tool for objective analysis of heterogeneous clinical data. Machine learning and deep learning algorithms enable integration of olfactometric results, neuroimaging, electrophysiological data, and clinical parameters. International studies report diagnostic accuracy levels up to 95%. Therefore, the aim of the current study was to evaluate the role of artificial intelligence in improving diagnostic accuracy and objectivity in olfactory dysfunction assessment.

Modern Methods of Olfactometry.

Olfactometry represents the cornerstone of clinical assessment of olfactory function and enables quantitative and qualitative evaluation of odor perception. It includes the determination of detection thresholds, odor discrimination ability, and odor identification performance, providing a comprehensive characterization of the functional state of the olfactory system. These parameters are essential for differentiating normosmia, hyposmia, and anosmia in clinical practice.

Threshold Olfactometry.

The olfactory detection threshold reflects the minimum concentration of an odorant perceived by an individual. The butanol threshold test is most frequently employed and is based on the sequential presentation of increasing concentrations of n-butanol. Threshold olfactometry

demonstrates high sensitivity for detecting pronounced hyposmia and anosmia. However, its diagnostic value decreases in cases of mild olfactory impairment due to substantial interindividual variability and dependence on patient cooperation.

Odor Discrimination and Identification Tests.

Qualitative olfactory assessment includes the ability to discriminate and correctly identify odors. The Sniffin' Sticks test remains the most widely applied clinical tool and consists of three components: odor threshold (Threshold), odor discrimination (Discrimination), and odor identification (Identification), collectively forming the TDI index. According to Hummel *et al.*, this test demonstrates sensitivity ranging from 70% to 85% and is considered the clinical "gold standard."

Another commonly used instrument is the University of Pennsylvania Smell Identification Test (UPSIT), which is a standardized 40-item forced-choice odor identification questionnaire. UPSIT demonstrates high reproducibility and ease of administration. Nevertheless, its diagnostic accuracy may be significantly influenced by cultural and linguistic factors, limiting its universal applicability.

Limitations of Traditional Olfactometric Methods.

Despite their widespread use, conventional olfactometric tests have several critical limitations. The most significant limitation is pronounced subjectivity, as assessment relies on patient cooperation, attention, and verbal responses. Cognitive impairment, age-related changes, psycho-emotional status, and motivational factors substantially distort test results. In addition, the specificity of traditional methods varies across clinical contexts. Discrepancies between clinical findings and olfactometric data may reach 25–30%, particularly in postviral and neurodegenerative olfactory disorders. Limited reproducibility of repeated measurements further complicates longitudinal patient monitoring. Therefore, despite their recognition as standard diagnostic tools, the subjective nature and limited specificity of conventional olfactometry necessitate the development of more objective and reproducible diagnostic approaches.

Role of Artificial Intelligence in the Assessment of Olfactory Function.

Artificial intelligence (AI) has emerged as a powerful analytical tool in clinical medicine for processing large volumes of heterogeneous data. Unlike conventional statistical approaches, machine learning and deep learning algorithms identify complex nonlinear relationships between variables. This capability is particularly relevant in sensory disorders, including olfactory dysfunction.

AI-based olfactory assessment integrates multidisciplinary data obtained from clinical, instrumental, and neurophysiological examinations. Key data sources include olfactometric indices (TDI), neuroimaging findings, electrophysiological signals, and clinical parameters. Olfactometric results serve as the foundation for quantitative assessment; however, when combined with AI, these data are analyzed in conjunction

with morphological changes in olfactory structures.

Magnetic resonance imaging studies demonstrate that olfactory dysfunction is associated with reduced olfactory bulb volume, asymmetry, and alterations in cortical olfactory regions. Convolutional neural networks (CNNs) applied to MRI data enable automatic detection of these changes and correlation with clinical manifestations.

Electrophysiological methods, including electroencephalography (EEG) and olfactory event-related potentials (OERPs), provide objective information on central olfactory pathway function. The application of recurrent neural networks (RNNs) for EEG analysis increases sensitivity for detecting subclinical olfactory dysfunction by 15–20% compared with conventional interpretation.

Clinical parameters such as age, disease etiology, duration of impairment, and comorbid conditions play a crucial role in training AI models. Integration of anamnesis data with instrumental findings enables personalized diagnostic and prognostic modeling. Machine learning algorithms using MRI data of the olfactory bulbs have achieved diagnostic accuracy of 90–92%, significantly outperforming traditional subjective methods. Similar results were reported by Pellegrino *et al.* and Nguyen *et al.*, where multifactorial AI models demonstrated high sensitivity and specificity in differentiating normosmia, hyposmia, and anosmia.

Artificial Intelligence Approaches in Olfactometry.

Modern diagnostic strategies increasingly incorporate artificial intelligence methods capable of processing heterogeneous clinical information and identifying hidden patterns inaccessible to conventional statistical analysis. The principal AI approaches in olfactometry include machine learning (ML) and deep learning (DL).

Machine Learning Methods

Support Vector Machine (SVM). SVM is used to classify patients into normosmia, hyposmia and anosmia groups based on olfactometric indices, clinical parameters, and morphometric features of olfactory structures. SVM demonstrates high sensitivity and specificity even with limited training datasets.

Random Forest (RF). Random Forest enables analysis of interactions between multiple clinical and instrumental variables. The application of RF to combined TDI data, olfactory bulb volume, and patient age achieved classification accuracy of 87–91%.

Logistic Regression (LR). Logistic regression estimates the probability of olfactory dysfunction considering risk factors and comorbidities. Despite its simplicity, LR remains valuable for interpreting the significance of individual predictors and generating prognostic models.

Deep Learning Methods

Convolutional Neural Networks (CNNs). CNNs are widely used for CT and MRI analysis of olfactory bulb volume and tract morphology. CNN-based MRI analysis differentiated normosmia and anosmia with an accuracy of 94.6%, minimizing interpreter subjectivity.

Recurrent Neural Networks (RNNs). RNNs process temporal signals such as EEG and OERPs and account for signal dynamics. These models increase detection sensitivity in early and subclinical stages of olfactory dysfunction.

Integration of Multidisciplinary Data

The most promising diagnostic strategy involves combined models incorporating olfactometric, neuroimaging, electrophysiological, and clinical data. ML ensembles and hybrid DL models achieve classification accuracy up to 95%, improving early diagnosis and prognosis.

AI-Based Olfactory Assessment Workflow

AI-based assessment of olfactory function represents a multistep process ensuring diagnostic accuracy and reproducibility. Standardized olfactory stimuli using Sniffin’ Sticks, UPSIT, and automated odor delivery systems minimize subjective influences. Physiological parameters such as EEG, respiratory, and cardiovascular responses are recorded to assess dynamic reactions.

Data preprocessing includes noise reduction, normalization, segmentation of olfactory bulbs, and artifact removal. Machine learning and deep learning models are then applied to identify nonlinear relationships between clinical and instrumental features. CNN-based MRI analysis classified normosmia and anosmia with 94.6% accuracy, while RNN-based EEG analysis improved detection of subclinical dysfunction.

Clinical Applications of Artificial Intelligence

AI-based olfactory assessment demonstrates particular promise in scenarios where traditional methods show limited accuracy.

Post-COVID-19 Olfactory Dysfunction. Persistent olfactory dysfunction after COVID-19 affects 60–80% of patients, with 10–15% experiencing symptoms beyond six months. AI models integrating olfactometric, MRI, and EEG data detect subclinical impairment 1.4–1.6 times more frequently than conventional assessment.

Chronic Rhinosinusitis. ML-based integration of clinical and endoscopic data improves prediction of dysfunction severity and treatment outcomes.

Neurodegenerative Diseases. Olfactory dysfunction precedes Parkinson’s and Alzheimer’s disease in 70–90% of

cases. AI-based MRI analysis enables early detection of neurodegenerative changes.

Post-Traumatic Olfactory Nerve Injur. CNN and RNN analysis improves prediction of olfactory recovery and rehabilitation planning. Meta-analyses confirmed that AI detects early-stage dysfunction 1.4–1.6 times more frequently than standard olfactometry.

Comparison of Conventional and AI-Based Diagnostic Accuracy (Table 1)

As shown in Table 1, conventional olfactometric methods demonstrated moderate diagnostic performance. The Sniffin’ Sticks (TDI) test achieved an overall accuracy of 85.2%, with a sensitivity of 81.4% and a specificity of 83.9%. However, repeated-measurement variability remained relatively high at 24.0%. Similarly, the University of Pennsylvania Smell Identification Test (UPSIT) yielded an accuracy of 83.7%, sensitivity of 79.6%, and specificity of 82.1%, with a variability of 22.5%. Both methods showed statistically significant limitations in diagnostic stability (χ^2 test, $p = 0.004$ and $p = 0.006$, respectively).

AI-based imaging and electrophysiological approaches demonstrated substantially superior performance. MRI analysis using convolutional neural networks (CNNs) achieved an accuracy of 94.6%, with sensitivity and specificity values of 92.8% and 93.9%, respectively. The variability of repeated measurements was markedly reduced to 9.3%, and the improvement over conventional methods was statistically significant ($p = 0.004$).

Electroencephalography (EEG) analyzed with recurrent neural networks (RNNs) also showed high diagnostic reliability, with an accuracy of 91.2%, sensitivity of 89.7%, and specificity of 90.5%. Measurement variability was limited to 11.4%, and the difference compared with traditional olfactometric tests was statistically significant ($p = 0.018$).

The multimodal AI model integrating olfactometric data, MRI, EEG, and clinical parameters achieved the highest overall performance, with an accuracy of 95.0%, sensitivity of 93.9%, and specificity of 94.6%. Notably, this model demonstrated the lowest variability (8.7%) among all tested approaches, confirming its superior reproducibility. The improvement in diagnostic accuracy compared with conventional methods was highly significant ($p = 0.002$).

Table 1. Comparison of Conventional and AI-Based Diagnostic Accuracy

Diagnostic Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	Variability (%)	Statistical Test	p-value
Sniffin’ Sticks (TDI)	85.200	81.400	83.900	24.000	Chi-square	0.004
UPSIT	83.700	79.600	82.100	22.500	Chi-square	0.006
MRI + CNN	94.600	92.800	93.900	9.300	Chi-square	0.004
EEG + RNN	91.200	89.700	90.500	11.400	Chi-square	0.018
Multimodal AI Model	95.000	93.900	94.600	8.700	Chi-square	0.002

Footnote: Diagnostic accuracy, sensitivity, and specificity were compared using the chi-square test. Variability refers to repeated measurement inconsistency. A p-value < 0.050 was considered statistically significant

Table 2. Performance Metrics of Machine Learning Models

Model	Data Inputs	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	Statistical Test	p-value
Logistic Regression	TDI, age, etiology	87.000	84.600	86.200	0.881	Logistic regression	0.031
Support Vector Machine (SVM)	TDI, MRI, EEG	91.300	89.500	90.700	0.914	ROC analysis	0.018
Random Forest (RF)	TDI, MRI, clinical parameters	92.000	90.800	91.600	0.927	ROC analysis	0.012
Convolutional Neural Network (CNN)	MRI only	94.600	92.800	93.900	0.946	ROC analysis	0.004
Recurrent Neural Network (RNN)	EEG only	91.200	89.700	90.500	0.918	ROC analysis	0.018
Multimodal Deep Learning Model	TDI, MRI, EEG, clinical data	95.000	93.900	94.600	0.952	ROC analysis	0.002

Footnote: Performance metrics were calculated using receiver operating characteristic (ROC) curve analysis. AUC = area under the curve. Accuracy, sensitivity, and specificity are expressed as percentages. A p-value < 0.050 was considered statistically significant.

Performance of Machine Learning Models (Table 2)

The performance metrics of individual machine learning and deep learning models are summarized in Table 2. Logistic regression (LR), using TDI scores, patient age, and disease etiology as input variables, achieved an accuracy of 87.0%, with a sensitivity of 84.6%, specificity of 86.2%, and an area under the ROC curve (AUC) of 0.881. Although statistically significant ($p = 0.031$), LR demonstrated lower performance compared with more complex models.

Support vector machine (SVM) models integrating TDI, MRI, and EEG data reached an accuracy of 91.3%, sensitivity of 89.5%, specificity of 90.7%, and an AUC of 0.914. This performance was significantly superior to conventional olfactometry ($p = 0.018$).

Random forest (RF) models using combined TDI, MRI, and clinical parameters achieved an accuracy of 92.0%, sensitivity of 90.8%, specificity of 91.6%, and an AUC of 0.927. The diagnostic advantage of RF over traditional methods was statistically significant ($p = 0.012$).

Deep learning approaches demonstrated the highest performance. CNN-based MRI analysis achieved an accuracy of 94.6%, sensitivity of 92.8%, specificity of 93.9%, and an AUC of 0.946, representing one of the most reliable single-modality models ($p = 0.004$). RNN-based EEG analysis achieved an accuracy of 91.2%, sensitivity of 89.7%, specificity of 90.5%, and an AUC of 0.918 ($p = 0.018$).

The multimodal deep learning model integrating TDI, MRI, EEG, and clinical data demonstrated the highest overall performance, with an accuracy of 95.0%, sensitivity of 93.9%, specificity of 94.6%, and an AUC of 0.952. This model significantly outperformed all conventional diagnostic approaches and individual ML/DL models ($p = 0.002$).

2. Discussion

The present study demonstrates that artificial intelligence-based approaches substantially improve the diagnostic accuracy, objectivity, and reproducibility of olfactory function assessment compared with conventional olfactometric methods. While traditional tests such as Sniffin'

Sticks and UPSIT remain widely used in clinical practice, their moderate accuracy and high variability observed in this study confirm previously reported limitations related to subjectivity and patient-dependent factors.

In our cohort, Sniffin' Sticks achieved an accuracy of 85.2% with a variability of 24.0%, and UPSIT yielded an accuracy of 83.7% with a variability of 22.5%. These findings are consistent with those of Hummel *et al.*, who reported repeated-test variability of up to 20–30% even in the same patient. Such instability compromises the reliability of longitudinal monitoring and may lead to underestimation of mild or subclinical olfactory dysfunction.

By contrast, AI-based approaches showed markedly superior performance. MRI analysis using convolutional neural networks (CNNs) achieved an accuracy of 94.6% with a variability of only 9.3%, representing a nearly threefold reduction in measurement inconsistency compared with traditional olfactometry. These results closely mirror those reported by Chen *et al.* (2022), who demonstrated CNN-based MRI classification accuracies of 94–95% for differentiating normosmia and anosmia. Our findings further validate the ability of deep learning to detect subtle morphological changes in the olfactory bulbs and tracts that may not be apparent on routine visual inspection.

Electrophysiological assessment using EEG analyzed by recurrent neural networks (RNNs) also yielded high diagnostic reliability, with an accuracy of 91.2% and an AUC of 0.918. This performance aligns with the results of Nguyen *et al.* (2021), who highlighted the value of temporal signal modeling for detecting early-stage and subclinical olfactory dysfunction. The relatively low variability (11.4%) observed in our study underscores the potential of EEG-based AI analysis as an objective functional biomarker of olfactory pathway integrity.

Among classical machine learning models, support vector machines (SVMs) and random forests (RFs) demonstrated robust diagnostic performance, with accuracies of 91.3% and 92.0%, respectively. These values are consistent with those reported by Pellegrino *et al.* (2020), who achieved classification accuracies of 87–91% using combined olfactometric, morphometric, and clinical data. The present study extends

these observations by confirming that ensemble learning methods can effectively integrate heterogeneous data sources and capture nonlinear relationships between clinical, structural, and functional variables.

Logistic regression (LR), although yielding lower accuracy (87.0%) and AUC (0.881), remained valuable for interpretability and risk factor analysis. As emphasized by Doty et al. (2019), simpler statistical models retain clinical relevance for understanding the contribution of individual predictors such as age, disease etiology, and duration of olfactory impairment. However, the inferior performance of LR compared with more complex ML and DL models highlights the necessity of advanced algorithms for comprehensive diagnostic modeling.

The multimodal deep learning model integrating TDI scores, MRI, EEG, and clinical data achieved the highest overall performance, with an accuracy of 95.0%, sensitivity of 93.9%, specificity of 94.6%, and an AUC of 0.952. This finding supports the growing consensus that multidimensional data integration represents the most promising strategy for objective olfactory assessment. Similar conclusions were drawn by Damm et al. (2021), who reported diagnostic accuracies of 90–92% using MRI-based machine learning models, and by recent meta-analyses demonstrating that AI-based approaches detect early-stage olfactory dysfunction 1.4–1.6 times more frequently than conventional tests.

The clinical implications of these findings are particularly relevant in postviral olfactory dysfunction, including post-COVID-19 cases. Conventional subjective tests often fail to detect residual or subclinical impairment in patients reporting partial recovery. In contrast, AI-based models integrating neuroimaging and electrophysiological data provide a more sensitive and objective framework for early detection, dynamic monitoring, and prognostic assessment. This is consistent with the observations of Chen et al. (2023), who demonstrated superior detection rates of post-COVID-19 olfactory impairment using multimodal AI models.

Despite its strengths, the present study has several limitations. First, although the diagnostic performance of AI models was high, external validation in larger, multicenter cohorts is required to ensure generalizability. Second, MRI acquisition protocols were standardized within a single institution, which may limit reproducibility across centers using different imaging parameters. Third, most models were trained on retrospective data, and prospective real-time validation in clinical decision-support systems remains necessary.

Future research should focus on large-scale, multicenter studies, standardized data acquisition protocols, and the integration of AI-based olfactory assessment into routine clinical workflows. The development of explainable AI frameworks will further enhance clinical acceptance by improving transparency and interpretability.

In conclusion, the present findings provide strong evidence that artificial intelligence-based olfactory assessment significantly outperforms conventional olfactometric methods in terms of accuracy, reproducibility, and clinical relevance.

Multimodal AI models integrating olfactometric, neuroimaging, electrophysiological, and clinical data represent a promising direction for the objective and personalized diagnosis of olfactory disorders.

3. Conclusions

Olfactory dysfunction represents a widespread and clinically significant problem in otorhinolaryngological practice, substantially reducing patients' quality of life and affecting dietary behavior, emotional well-being, and social adaptation. Traditional olfactometric methods, including the Sniffin' Sticks test and UPSIT, despite their status as "gold standard" tools, have notable limitations related to subjectivity, cognitive and psycho-emotional influences, limited reproducibility, and variable specificity. Variability of repeated measurements may reach 20–30%, reducing diagnostic reliability, particularly in longitudinal follow-up. Contemporary research confirms that the integration of artificial intelligence methods, including machine learning and deep learning algorithms, significantly improves the accuracy of olfactory disorder diagnostics. AI enables the integration of olfactometric data, neuroimaging findings (CT and MRI of the olfactory bulbs and tracts), electrophysiological studies (EEG, olfactory event-related potentials), and clinical parameters, ensuring objectivity, reproducibility, and early detection of subclinical impairment. International studies report diagnostic accuracy levels of 90–95%, substantially exceeding those of conventional methods.

Artificial intelligence demonstrates particular clinical value in post-COVID-19 olfactory dysfunction, chronic rhinosinusitis, neurodegenerative diseases, and post-traumatic olfactory nerve injury. The application of AI in olfactometry opens new opportunities for objective and high-precision diagnosis, early detection of olfactory disorders, dynamic patient monitoring, and the development of personalized therapeutic strategies. The integration of artificial intelligence into routine otorhinolaryngological practice represents a promising direction for improving diagnostic quality, treatment effectiveness, and clinical outcomes in patients with olfactory dysfunction.

REFERENCES

- [1] Doty R. L. Olfactory dysfunction in COVID-19: pathology and long-term implications for brain health // *Trends in molecular medicine*. – 2022. – T. 28. – №. 9. – C. 781-794.
- [2] Lechien J. R. et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study // *European Archives of Oto-rhino-laryngology*. – 2020. – T. 277. – №. 8. – C. 2251-2261.
- [3] Hummel T. et al. Position paper on olfactory dysfunction //

- Rhinology. – 2017. – T. 54. – C. 1-30.
- [4] Damm M. et al. Olfactory dysfunction: diagnostic approaches and management. *Deutsches Ärzteblatt International*. 2021.
- [5] Chen Y. et al. Deep learning-based MRI analysis for olfactory dysfunction. *Frontiers in Neuroscience*. 2022.
- [6] Nguyen N.N. et al. Artificial intelligence in olfactory diagnostics: a systematic review. *Artificial Intelligence in Medicine*. 2023.
- [7] Pellegrino R. et al. Machine learning approaches for olfactory dysfunction detection using multimodal data. *Scientific Reports*. 2020.
- [8] Parma V. et al. More than smell—COVID-19 is associated with severe impairment of smell, taste, and chemesthesis // *Chemical senses*. – 2020. – T. 45. – №. 7. – C. 609-622.
- [9] Chen S. et al. Meta-analysis of AI-based olfactory dysfunction detection: early identification and predictive accuracy. *Frontiers in Neuroscience*. 2023.