Nature and Science of Sleep

Open Access Full Text Article

REVIEW

# From Big Data to AI-Driven Decisions in Obstructive Sleep Apnea: A Narrative Review Integrating the DDPP Framework

Mengying Wu[1,2,*], Kexin Wang[2,*], Huai Huang[2,*], Xiaodan Wu[2], Zilong Liu[2], Shanqun Li[2]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, People's Republic of China; [2]Zhongshan Hospital, Fudan University, Shanghai, 200032, People's Republic of China

*These authors contributed equally to this work

Correspondence: Shanqun Li; Zilong Liu, Zhongshan Hospital, Fudan University, Shanghai, 200032, People's Republic of China, Email li.shanqun@zs-hospital.sh.cn; yjbbdx1988@126.com

**Abstract:** Obstructive sleep apnea (OSA) remains underdiagnosed and inadequately managed despite an explosion in multimodal data and swift progress in artificial intelligence (AI). To elucidate the extent of AI techniques utilized in OSA data resources, we conducted a comprehensive literature search in PubMed, Web of Science, Scopus, and IEEE Xplore from 1 April 2020 to 1 April 2025. Search terms related to AI were combined with "obstructive sleep apnea", and 575 original studies were found after de-duplication and exclusion. We employed the DDPP analytics model (Descriptive, Diagnostic, Predictive, and Prescriptive), derived from the business domain, to structure reported clinical applications. The study indicates a significant gap between available data and current AI: most research focuses on sleep monitoring signals, whereas patient-reported outcomes, electronic health records, and environmental data (both social and natural) are largely underutilized. In clinical practice, applications typically concentrate on Descriptive and Diagnostic phases, while Prescriptive analytics for personalized therapy is scarce. This is the first review to assess AI projects from the perspective of OSA data resources, and the first to apply the DDPP framework for sleep medicine analytics. We call on researchers to mine OSA-related data from multiple dimensions and to select suitable AI technologies based on the data characteristics, thereby enhancing clinical decision-making.

**Keywords:** obstructive sleep apnea, big data, artificial intelligence, data analysis framework



## Introduction

Obstructive Sleep Apnea (OSA) is a common sleep-related breathing disorder with high prevalence, low diagnostic rates, poor treatment adherence, and significant adverse impacts on individuals and society. According to research estimates, among individuals aged 30 to 69, approximately 936 million are affected by mild to severe OSA, while 425 million are affected by moderate to severe OSA.[1] The disease is characterized by intermittent hypoxemia and sleep fragmentation, which trigger a series of pathophysiological responses,[2] leading to cardiovascular diseases,[3,4] metabolic disorders,[5] neurological impairments,[6] and psychosomatic disorders.[7,8] However, as a heterogeneous disease, it exhibits a relatively uniform set of diagnostic and treatment measures with few available options, exacerbating the dilemma of clinical management.[9]

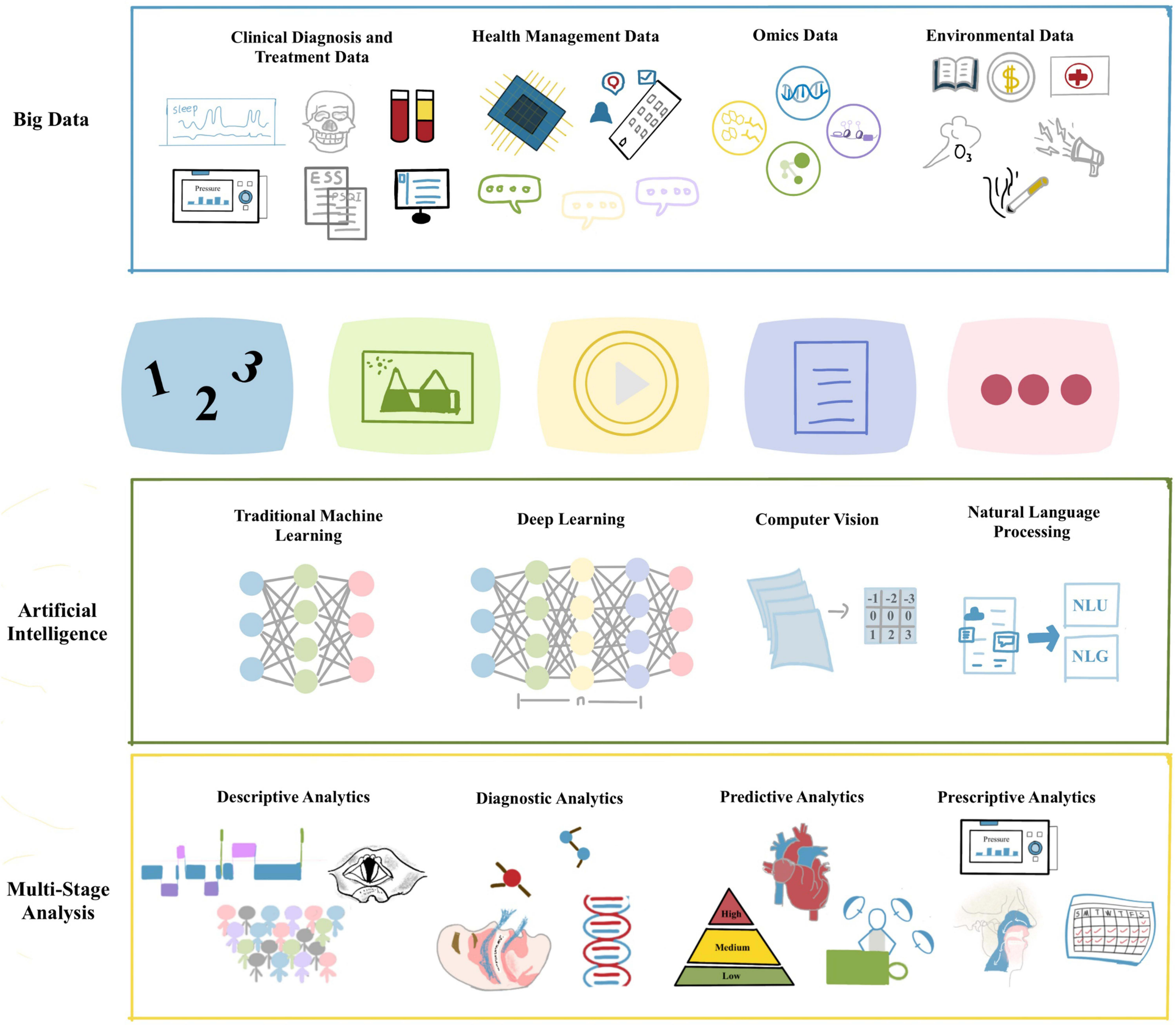
The advent of Big Data (BD) and Artificial Intelligence (AI) marked a new era in precision medicine. Data-driven AI techniques have made notable progress in rheumatology,[10] ophthalmology,[11] oncology,[12] and psychiatry,[13] while their applications in OSA are relatively underexplored. Following the wide implementation of monitoring tools in clinical practice, patients with OSA are generating large-scale, complex, and multidimensional datasets. In theory, concurrent advancements in the "three pillars" of AI—data, models, and computational power—should accelerate clinical translation. Nevertheless, the field of OSA still lacks a clear paradigm for elucidating anatomical and physiological mechanisms, defining meaningful phenotypes and endotypes, and developing personalized treatment models.

Based on the above observations, we hypothesize that an information mismatch exists between the two domains: AI researchers may be unaware of the existence and value of OSA-related data, while clinical researchers may be unfamiliar with the potential and constraints of AI. To address this issue, this narrative review adopts a problem-oriented approach,



Received: 3 June 2025
Accepted: 26 July 2025
Published: 21 August 2025

**Graphical Abstract**




structured around three key sections: data, technology, and analytical framework. The underlying logic and flow are outlined in Figure 1. In Section 3, we introduce a modified version of the DDPP model (Descriptive, Diagnostic, Predictive, Prescriptive), tailored to reflect the staged nature of clinical decision-making in AI-driven applications. Rather than using it as a strict evaluative criterion, we employ it as a conceptual tool to organize and interpret the reviewed studies.

# Methods

## Data Classification

Based on clinical and research knowledge, we first classified OSA-related information into four main types: clinical diagnosis and treatment, health management, omics, and environmental data. This classification provided the foundation for Section 1 and supported the analysis of how different AI technologies are matched to distinct data modalities, as presented in the Theoretical Perspective of Section 2.

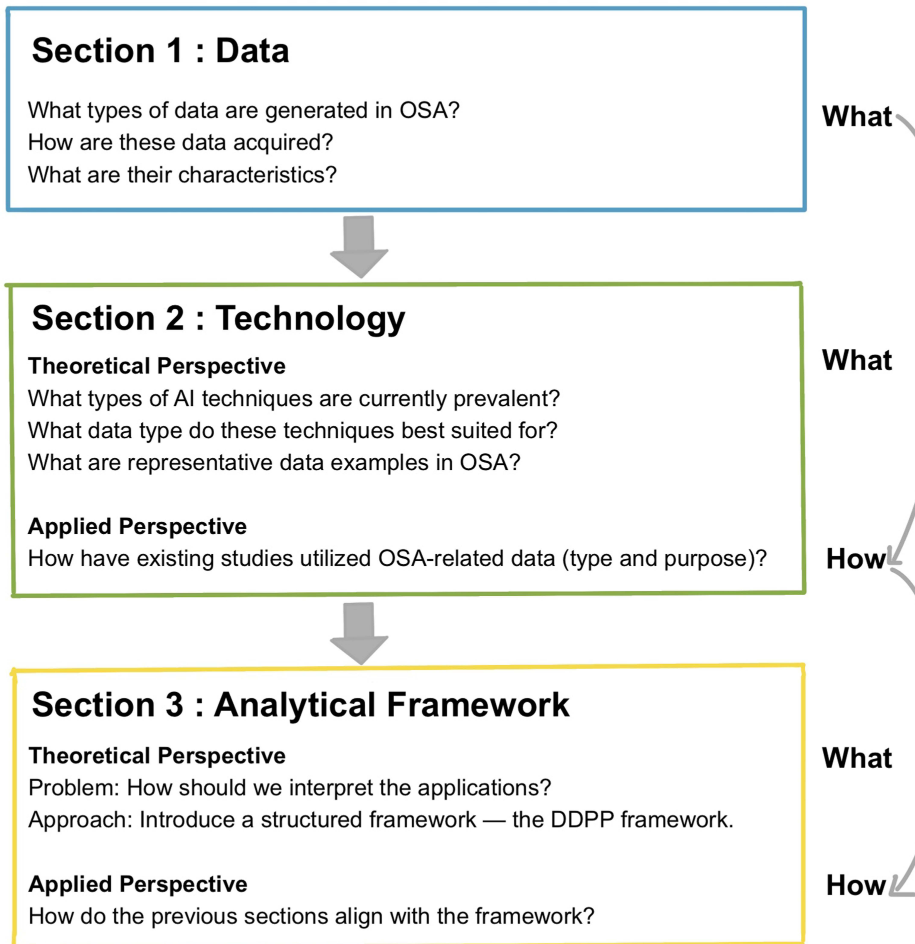



**Figure 1** The underlying logic and flow of this review.

## Literature Search

To provide a quantitative overview that informs Section 2, we conducted a literature search across four major databases, focusing on how AI technologies have been applied in OSA research. We restricted the search to the past five years because AI architectures, computational power, and sensor technologies have changed substantially since 2020, rendering older performance metrics less comparable to current clinical readiness. Titles and abstracts were screened by one researcher to identify studies involving AI applications in the context of OSA. Due to resource constraints, dual review was not conducted, which is acknowledged as a limitation while considered acceptable for this narrative review. All references were managed using EndNote 20. The overall search strategy is summarized in Table 1, detailed search strings are provided in Appendix 1, and the process is illustrated with a PRISMA flowchart in Figure 2.

## Framework Development

To facilitate a clinically oriented analysis, we developed a conceptual framework based on the DDPP model in Section 3. Although originally rooted in business analytics, its components were reinterpreted through semantic restructuring and cross-domain mapping to align with clinical decision-making. A detailed explanation is provided at the beginning of the Multi-Stage Mapping via DDPP.

**Table 1** Summary of Literature Search Strategy

| Item | Details |
|---|---|
| Databases | PubMed, Web of Science, Scopus, IEEE Xplore |
| Search Fields | Title/Abstract |
| Search Terms | #1 "Obstructive Sleep Apnea"<br>#2 "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Computer Vision" OR "Natural Language Processing" |
| Search Query | PubMed, Web of Science, Scopus: #1 AND #2<br>IEEE Xplore: #1 |
| Rationale for IEEE | Many engineering studies describe specific algorithms without general AI terminology in the title or abstract. |
| Timeframe | 1 April 2020 to 1 April 2025 |
| Language | English |
| Inclusion Criteria | - Original research articles<br>- Studies focusing on OSA |
| Exclusion Criteria | - Duplicate records<br>- Not original articles: review articles, bibliometric analyses, secondary data analyses, conference proceedings<br>- Not OSA-related: studies focusing solely on algorithm development, device improvement or not directly investigating OSA |

# Big Data In Osa

Four categories of data were identified and summarized, with their purpose, content, acquisition methods, and characteristics detailed in Table 2.

## Clinical Diagnosis and Treatment Data

### Sleep Monitoring Data

Sleep monitoring serves as the cornerstone of OSA diagnosis and treatment. Patients frequently exhibit symptoms such as daytime sleepiness, impaired concentration, fatigue, and nocturnal snoring, suggesting underlying physiological disturbances during sleep. To quantify core signals that reflect the interaction between sleep and respiration,[15] various monitoring devices have been developed. Monitoring can be conducted either in sleep laboratories using polysomnography (PSG; type I), or at home via home sleep apnea testing (HSAT; types II–IV), which differ in the number of monitored parameters and whether attended.[16] These data are synchronously acquired, present stage-specific and time-series characteristics, and are typically stored in EDF+ format,[17] which facilitates signal processing and algorithmic modeling.

Owing to its rich physiological information and standardized digital structure, sleep monitoring data have become a foundational input for AI applications. Nonetheless, several difficulties persist, including: (1) the high cost and limited accessibility of PSG examinations; (2) the susceptibility of signal quality to environmental interference; and (3) the reliability of manual annotation in sleep stages and events.

### Imaging Data

To visualize the anatomical structures and dynamic changes of the upper airway—features not captured by sleep monitoring—imaging techniques are added as complementary tools in OSA evaluation. Most modalities are static, including cephalometry, cone beam CT (CBCT), conventional multidetector CT (MDCT), and magnetic resonance imaging (MRI),[18] which allow for the assessment of airway anatomy and volume,[19,20] thereby informing etiological analysis and phenotypic classification. In contrast, dynamic techniques such as dynamic MRI,[21] drug-induced sleep endoscopy (DISE),[22] and videofluoroscopic swallowing study (VFSS),[23] provide real-time views of airway motion and swallowing function, aiding in surgical planning and functional assessment. These images are characterized by high spatial resolution and large data volumes, stored in standardized DICOM format.[24]

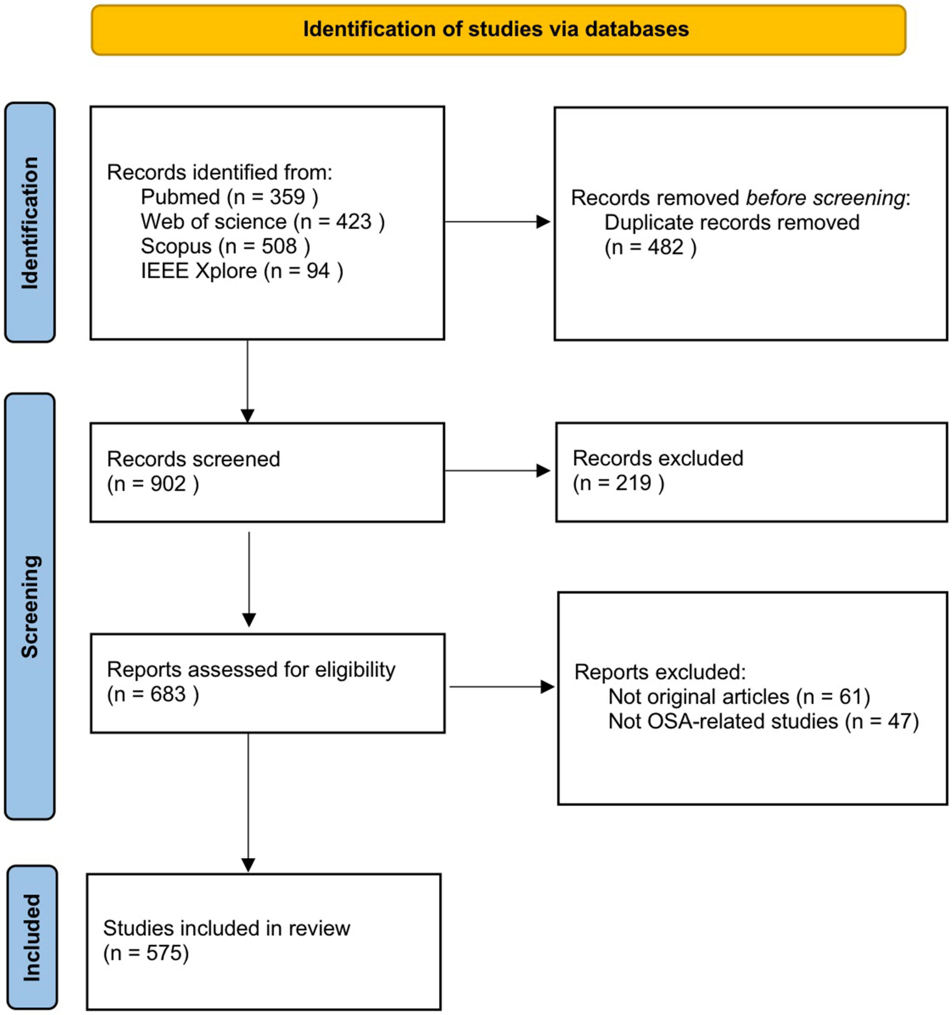



**Figure 2** PRISMA flowchart of search results at each step of the narrative review.
**Notes**: This figure was adapted from Page MJ, McKenzie JE, Bossuyt PM et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. Mar 29 2021;372:n71.[14]

Despite their strengths, the application of imaging data in AI faces some obstacles, including: (1) high computational requirements due to large image size; (2) inconsistency in image quality across scanners, limiting cross-center generalizability; and (3) the reliability of manual annotation in outlining upper airway contours.

## Laboratory Data

In OSA patients, intermittent hypoxia triggers a series of pathophysiological responses,[2] necessitating laboratory testing to identify systemic manifestations and potential risks. Commonly assessed parameters include blood gas analysis parameters, inflammation and oxidative stress markers, and metabolic or endocrine indicators. Some of these biomarkers—such as hs-CRP, IL-6, and urinary epinephrine sulfate—not only reflect general health status but are also associated with cardiovascular diseases and cognitive impairment.[25–27] These laboratory samples can be obtained from blood, urine, saliva, and exhaled breath condensate,[28] collected at single or multiple discrete time points, yielding low-dimensional, structured numerical data.

Two key challenges hinder their use in AI: (1) missing values due to inconsistent test items across patients or institutions; and (2) sparse temporal sampling, limiting the ability to capture dynamic changes.

 https://doi.org/10.2147/NSS.S543091 

https://doi.org/10.2147/NSS.S543091


**Table 2** Four Data Categories in OSA

| Category | Subcategory | Why | What | How | Characteristics |
|---|---|---|---|---|---|
| Clinical Diagnosis and Treatment Data | Sleep Monitoring Data | To assess the physiological changes in sleep | - electrophysiological signals (eg, EEG, EOG, EMG)<br>- respiratory signals (eg, nasal airflow, thoracoabdominal respiratory effort, snoring)<br>- cardiovascular signals (eg, ECG, heart rate)<br>- oxygen saturation signals (eg, arterial oxygen saturation) | - PSG<br>- HSAT | - stage-specific<br>- time-series<br>- EDF+ |
| | Imaging Data | To display the anatomical structure and dynamic changes of the upper airway | - craniofacial skeletal structures (eg, anterior facial height, hyoid bone position, pharyngeal airway space)<br>- soft tissues surrounding the airway (eg, tongue, soft palate, parapharyngeal fat deposition).<br>- regional respiratory movement of the tongue<br>- upper airway collapse characteristics<br>- swallowing function | - cephalometry<br>- CBCT<br>- MDCT<br>- MRI<br>- dynamic MRI<br>- DISE<br>- VFSS | - tomographic or dynamic scanning<br>- high resolution<br>- large volumes<br>- DICOM |
| | Laboratory Data | To identify systemic manifestations and potential risks | - blood gas analysis parameters (eg, $PaO_2$, $PaCO_2$, $SaO_2$, Hb)<br>- inflammation and oxidative stress markers (eg, hs-CRP, IL-6, TNF-α, MDA, SOD)<br>- metabolic and endocrine indicators (eg, HbA1c, FPG, leptin) | - blood samples<br>- urine samples<br>- saliva samples<br>- exhaled breath condensate | - discrete time-point collection (single/ multiple)<br>- low-dimensional, structured numerical values |
| | PAP Data | To record objective therapeutic information | - sleep events<br>- airway pressure<br>- air leakage<br>- compliance | - PAP device | - time-series characteristics<br>- different modalities (eg, numerical parameters, waveform signals, and audio recordings) |
| | PROs | To reflect subjective health information | - sleep quality<br>- emotional state<br>- functional status<br>- quality of life | - self-designed questionnaires (eg, ESS, STOP-Bang, PSQI)<br>- validated scales | - numerical data<br>- textual data |
| | EHRs | To track disease progression | - medical history<br>- anthropometric measurements<br>- diagnostic information<br>- laboratory examinations<br>- imaging examinations<br>- medication records<br>- surgical records | - clinical record | - structured data<br>- unstructured data |

| | | | | | |
|---|---|---|---|---|---|
| Health Management Data | Wearable Devices | To acquire real-time physiological data | - pulse waves<br>- respiratory patterns<br>- blood oxygen saturation<br>- acoustic signals | - multiple sensors | - structured time-series data |
| | Mobile Health Applications | To track sleep and health-related behaviors | - lifestyle<br>- daily activities<br>- sleep logs<br>- PAP treatment adherence | - smartphone applications | - structured data<br>- unstructured data |
| | Social Media | To document the experiences shared by the patient | - disease management<br>- treatment feedback<br>- emotional changes | - social platforms<br>- health forums<br>- online patient communities | - largely unstructured text |
| Omics Data | Genomics/Epigenomics/ Transcriptomics/Proteomics/ Metabolomics/Emerging Omics | To provide comprehensive knowledge about cellular components and biomolecules | - DNA<br>- RNA<br>- proteins<br>- metabolites | - next-generation sequencing<br>- immunochemistry<br>- microarrays<br>- mass spectrometry<br>- nuclear magnetic resonance | - large-scale, high-dimensional, and sparse properties<br>- FASTQ (for sequencing)<br>- CEL (for microarrays)<br>- mzML (for mass spectrometry) |
| Environmental Data | Social Environment | To reflect patients' social resources and support systems | - personal-level social information (eg, income, occupation, educational level)<br>- societal-level healthcare information (eg, commercial insurance claims, public medical insurance claims, national patient registry) | - questionnaires<br>- databases | - large-scale and multi-source data<br>- structured and unstructured data |
| | Natural Environment | To assess the impact of external environmental factors | - heavy metal<br>- noise pollution<br>- light pollution<br>- second-hand smoke<br>- air pollution | - environmental monitoring agencies<br>- public health databases<br>- epidemiological surveys | - large-scale and multi-source data<br>- time-series and geospatial data |

**Abbreviations**: CBCT, cone beam CT; DICOM, digital imaging and communications in medicine; DISE, drug-induced sleep endoscopy; ECG, electrocardiography; EEG, electroencephalography; EMG, electromyography; EOG, electrooculography; ESS, Epworth sleepiness scale; HSAT, home sleep apnea testing; MDCT, conventional multidetector CT; MRI, magnetic resonance imaging; PAP, positive airway pressure; PSG, polysomnography; PSQI, Pittsburgh Sleep Quality Index; VFSS, videofluoroscopic swallowing study.




https://doi.org/10.2147/NSS.S543091

### Other Clinical Data

In addition to sleep monitoring, imaging, and laboratory information, a range of other clinical data sources contribute to an overall understanding of OSA. These include positive airway pressure (PAP), patient-reported outcomes (PROs), and electronic health records (EHRs).

Despite their clinical relevance, these data sources present distinct limitations in the implementation and interpretability of AI applications. PAP device data are often fragmented due to poor patient adherence and may be intermittently recorded; furthermore, data formats vary across manufacturers, limiting interoperability. PROs, though valuable for reflecting subjective health information, are constrained by linguistic ambiguity, different scale structures, and a scarcity of annotated datasets for training. EHRs encompass heterogeneous modalities with the information distributed in different subsystems. Moreover, variation in medical coding systems (eg, ICD, ICF) and unstructured documentation further hinders data mining. Extracting usable information often requires extensive preprocessing and entity-level normalization.

## Health Management Data

Health management data refers to various types of information generated outside clinical settings, including data from wearable devices, mobile health applications, and social media platforms.

Wearable devices and mobile health applications primarily support real-time acquisition of physiological signals and behavioral patterns. However, these data are limited in physiological scope and generally fall short of clinical quality standards. Social media content, derived from patient-shared experiences across platforms, is highly unstructured and linguistically diverse, introducing substantial semantic ambiguity.

## Omics Data

Multi-omics analysis provides comprehensive knowledge about cellular components and biomolecules,[29] including genomics, epigenomics, transcriptomics, proteomics, metabolomics, and emerging omics.[30] Advances in high-throughput technologies have enabled the simultaneous detection and integration of multi-layer information, which facilitates the investigation into disease pathogenesis.

In OSA research, notable progress has been achieved in identifying potential biomarkers, particularly through proteomics and metabolomics.[31] Multi-omics studies have provided insights into biomarkers associated with disease severity and treatment response.[32–34]

Characterized by large-scale, high-dimensional, and sparse properties, omics datasets often suffer from interpretability challenges. Additionally, technical variability across platforms and batches, and the integration across different omics layers complicate analysis.

## Environmental Data

Modern medicine emphasizes a holistic view of the patient. Beyond clinical, health management, and research data, environmental factors are gradually being recognized as potential factors in OSA severity. These can be classified into social and natural domains.

Social environment data cover both personal-level social information and societal-level healthcare information, reflecting the patients' social resources and support systems. These data, often integrated with clinical diagnosis and treatment records, are widely used to analyze the readmission rates and healthcare resource utilization among OSA patients;[35] to investigate the accuracy of claims-based algorithms in assessing PAP compliance;[36] and to evaluate the impact of the disease on the overall economic burden of the country.[37] The results of relevant studies provide data support for optimizing disease management strategies, improving insurance compensation mechanisms, and formulating public health policies.

Natural environmental factors associated with sleep include heavy metals, noise pollution, light exposure, second-hand smoke, and air pollution.[38] Associations between temperature, ozone levels, and AHI have also been reported.[39]

Such data are often challenging to obtain and integrate, largely due to the presence of numerous variables (eg, confounding, independent, and dependent), which require stratified analysis and rigorous analytical strategies when incorporating environmental factors into OSA research.

## Artificial Intelligence in OSA

Artificial intelligence, a modern approach,[40] covers four core areas: machine learning (ML), deep learning (DL), computer vision (CV), and natural language processing (NLP). Among them, ML and DL provide the theoretical foundations, while CV and NLP support practical implementations—rooted in their respective principles and increasingly enhanced by ML and DL algorithms. The internal relationships are visualized as a Venn diagram in Figure 3.

Although all four components are now being explored in OSA research, their development maturity, methodological robustness, and clinical applicability vary considerably. This section summarizes fundamental properties, key approaches, suitable data types, and typical applications, as presented in Table 3 and Table 4, and discusses patterns of deployment.

### Traditional Machine Learning

Traditional ML refers to algorithms that do not rely on deep neural networks and instead depend on manually designed features, which ensure strong interpretability. These models have low computational demands and are effective at handling structured data that are low-dimensional and small to moderate in scale, such as sleep monitoring metrics and anthropometric parameters.

In OSA research, traditional ML has been widely adopted for tasks involving identifying the presence of OSA (binary classification),[41] estimating disease severity (multi-class classification or regression),[32,33,42,43] and deriving phenotypic subgroups based on symptom profiles or physiological parameters (clustering).[44,45] Numerous studies have demonstrated good performance in these areas using classic algorithms like support vector machines (SVM), random forests (RFs), decision trees (DTs), and logistic regression (LR). However, the optimal algorithm varies according to data modality and feature set, even within a similar task. For example, in a screening study utilizing biochemical markers, LR outperformed three other algorithms, achieving the highest AUC.[56] In contrast, the OSA predictor tool employed XGBoost, selected as

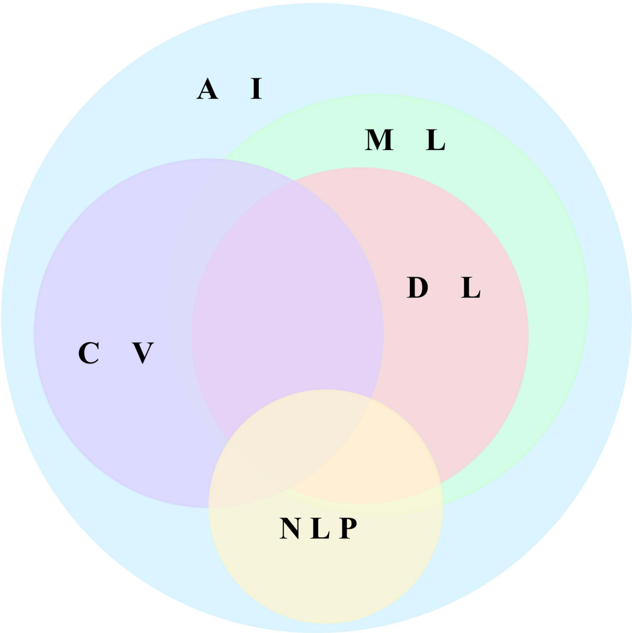



**Figure 3** Venn diagram of artificial intelligence.
**Abbreviations**: AI, artificial intelligence; CV, computer vision; DL, deep learning; ML, Machine Learning; NLP, Natural Language Processing.


https://doi.org/10.2147/NSS.S543091

**Table 3** Algorithmic Level Components in OSA

| Components | Characteristics | Core Tasks | Representative Algorithms | Applicable Data Types | OSA Data Examples | OSA Applications |
|---|---|---|---|---|---|---|
| Traditional ML | - non-DL methods<br>- hand-crafted<br>- strong interpretability<br>- low computational requirement | - classification<br>- regression<br>- clustering | - k-nearest neighbors<br>- support vector machines<br>- decision trees<br>- k-means | - structured data<br>- low-dimensional<br>- small-to-moderate scale | - sleep monitoring data (eg, sleep architecture, respiratory events, EEG signals, blood oxygen saturation, converted snoring signals)<br>- quantitative parameters extracted from imaging examinations (eg, cephalometric indicators, upper airway width)<br>- laboratory test data<br>- structured data from EHRs (eg, anthropometric parameters, diagnostic codes) | - automatic diagnosis[41]<br>- severity classification and prediction[32,33,42,43]<br>- derivation of phenotypic subgroups[44,45] |
| DL | - multi-layer neural networks and backpropagation<br>- end-to-end<br>- low interpretability<br>- high computational requirement | - classification<br>- regression<br>- clustering<br>- generation | - CNNs<br>- RNNs<br>- GNNs | - unstructured and structured data<br>- high-dimensional<br>- large scale | - imaging data<br>- PROs<br>- health management data collected in real-time (including waveform)<br>- omics data (eg, genomic sequences, protein structures)<br>- social environment data (eg, demographic data, lifestyle information) | - automatic diagnosis[46,47]<br>- precise disease assessment[48,49]<br>- personalized management[50] |

**Abbreviations**: CNNs, convolutional neural networks; EEG, electroencephalography; EHRs, electronic health records; GNNs, graph neural networks; PROs, patient-reported outcomes; RNNs, recurrent neural networks;


https://doi.org/10.2147/NSS.S543091

**Table 4** Application Level Components in OSA

| Components | Purposes | Core Techniques | Representative Models | Applicable Data Types | OSA Data Examples | OSA Applications |
|---|---|---|---|---|---|---|
| CV | - to endow computers with human-like visual comprehension capabilities | - image classification<br>- object detection<br>- object tracking<br>- instance segmentation<br>- semantic segmentation | - U-Net<br>- ResNet<br>- YOLO<br>- ViT | - image and video data (high-dimensional, unstructured) | - upper airway imaging<br>- DISE<br>- sleep video monitoring | - the accuracy of upper airway assessments[49]<br>- automated or semi-automated recognition of upper airway structures and soft tissue morphology[51,52]<br>- differentiation of certain OSA subtypes (eg, positional vs non-positional)[53] |
| NLP | - to develop computer systems capable of understanding and generating natural language | - embedding techniques<br>- temporal modeling architectures<br>- pre-trained language models | - Transformer<br>- BERT<br>- GPT | - textual data (unstructured) | - EHRs (eg, medical records, symptom descriptions, disease progression)<br>- PROs<br>- social media data (eg, sleep logs, medication feedback, forum discussions)<br>- medical literature | - OSA-related information extraction[54]<br>- sentiment analysis[55] |

**Abbreviations**: DISE, drug-induced sleep endoscopy; EHRs, electronic health records; PROs, patient-reported outcomes.


https://doi.org/10.2147/NSS.S543091

the optimal algorithm from eight machine learning methods, using seven routine clinical variables.[57] Beyond individual applications, a broader benchmarking analysis using the Sleep Health and Lifestyle Dataset reported Gradient Boosting as the most effective among 15 classifiers,[58] highlighting the influence of methodological choices.

Traditional ML retains practical value in specific OSA research contexts, though its role is gradually shifting—from serving as the primary analytic tool to acting as a complementary component in ensemble learning or a baseline for comparison against more advanced DL architectures.

## Deep Learning

DL, a branch of ML based on multi-layer neural networks and backpropagation, offers end-to-end automatic feature extraction that enhances representational power but weakens interpretability. It is notably adept at processing both unstructured and structured data that are high-dimensional, large-scale, or multimodal, all increasingly prevalent in OSA research, such as real-time health management data and omics data.

Building on these capabilities, DL enables new possibilities for diagnosis,[46,47] precise disease assessment,[48,49] and personalized management of OSA.[50] Depending on the data characteristics, convolutional neural networks (CNNs) are applied to image or video data, recurrent neural networks (RNNs) to sequential time-series inputs, and graph neural networks (GNNs) to graph structure information. For example, a scalogram-based CNN model was developed to detect OSA from single-lead ECG signals,[59] and the long short-term memory (LSTM) network—a type of RNNs—was applied to respiratory effort and airflow signals to identify apnea and hypopnea episodes.[60]

Unlike traditional ML, DL models are selected based on intrinsic data characteristics. As many biomedical signals exhibit multiple attributes simultaneously, hybrid architectures combining different neural networks are often required. A representative example is ECG signals, which possess both image-like and sequential properties, prompting the use of architectures such as CNN, CNN-LSTM, and CNN-GRU.[61] In a more recent study, Hossan et al proposed DeepApneaNet, a multiple CNN-Bi-LSTM framework, achieving improved accuracy in OSA detection.[62]

While DL provides advanced feature learning and strong tolerance to noise, its black-box nature limits clinical interpretability. Therefore, hybrid frameworks combining traditional ML with DL are increasingly adopted to balance automated representation learning and clinical transparency.[63]

## Computer Vision

CV supports the analysis of high-dimensional visual data in OSA research, including upper airway imaging, DISE recordings, and sleep video monitoring.

Most current studies focus on static image segmentation using semi-automated approaches to delineate upper airway structures from CBCT or MRI,[51,64] facilitating anatomical volume estimation and narrowing detection. With the integration of DL, several studies have developed fully automated segmentation pipelines. For instance, Bommineni et al implemented a U-Net architecture for precise segmentation of upper airway regions and quantification of anatomical risk factors, including the tongue-to-fat ratio.[52] The model achieved high accuracy and consistency, whereas its real-world clinical use was limited by sample size and imaging equipment variability.

Compared with static imaging, dynamic visual data analysis remains less explored but technically promising. Studies by Akbarian et al and Su et al employed optical flow–based deep learning models (ie, 3D-CNN) and object detection networks (ie, YOLOv4) to process infrared video and DISE video, respectively.[49,53] Although multicenter validation is lacking, these applications indicate the potential of DL-enhanced CV approaches for advancing noncontact monitoring and objective grading of airway obstruction.

## Natural Language Processing

NLP comprises a collection of computational techniques for the automatic analysis and representation of human language,[65] and is well-suited for processing textual data, such as EHRs, PROs, social media content, and medical literature.

The application of NLP in OSA is insufficient, with EHRs being the most commonly utilized data source. A typical pipeline involves applying traditional NLP techniques to preprocess unstructured clinical text—transforming it into structured feature variables such as symptoms, diagnostic codes, and key indicators—followed by large language models

(LLMs), an advanced NLP method augmented by DL, for information extraction and text classification. This approach enables the construction of high-quality disease cohorts and supports subsequent tasks such as model development, cohort comparison, and mortality risk analysis.[54,66,67]

Beyond EHRs, few studies have explored alternative data sources despite their promising potential. As an illustrative example, one study conducted sentiment analysis on social media posts describing experiences with modafinil—a wakefulness-promoting agent in OSA—yielding new insights into symptom improvement and perceived causality.[55]

In addition to inherent limitations of the above data types, a more prominent barrier is the scarcity of well-annotated databases. Manual labeling is labor-intensive and time-consuming, requiring substantial human effort to interpret nuanced language and customize tags according to specific study goals. Nevertheless, if appropriately developed, NLP techniques could unlock greater value by the identification of functional impairments from large volumes of clinical narratives, patient feedback and medical literature, and thus pinpoint critical concerns affecting OSA patients.

## Multi-Stage Mapping via DDPP

The DDPP model,[68,69] which includes four parts—Descriptive Analytics, Diagnostic Analytics, Predictive Analytics, and Prescriptive Analytics—has been widely adopted in business. In contrast, its application in the medical field is relatively scarce. Some scholars have attempted to apply it to bibliometric analysis,[70,71] demonstrating certain potential value. The concepts of "diagnosis" and "prediction" can be clarified in three interested fields.

(1) In business data analysis, "diagnostic analytics" focuses on investigating the underlying causes of a given phenomenon, while "predictive analytics" centers on forecasting possible future situations and trends.
(2) At the intersection of medicine and data science, "clinical prediction models" can be divided into diagnostic prediction models (which assess whether a disease is present) and prognostic prediction models (which assess the likelihood of developing a specific health outcome in the future).[72]
(3) In clinical practice, "diagnosis" refers to "whether a disease is present" and "how severe it is".

Therefore, to better express medical concepts and facilitate academic communication, this study seeks to integrate three critical elements—the DDPP model, clinical prediction models, and clinical experience—into a systematic analytical framework. The logic and application of this adaptive framework are depicted in Figure 4.

### Descriptive Analytics

Descriptive analytics mainly answers "What happened?", emphasizing the "explicit features" of the data. It is appropriate for analyzing baseline characteristics in epidemiological investigations. Compared with traditional descriptive analytics, which focuses on basic statistical characteristics, modern descriptive analytics, leveraging BD and AI, not only captures more precise medical information (eg, shorter-epoch sleep architecture, epiglottic obstruction ratio),[48,49] but also identifies subgroups of patients with similar physiological or clinical characteristics through techniques.[44,45] Modern descriptive analytics is conducive to understanding the heterogeneity of OSA and proposing etiological hypotheses based on the observed phenomena.

Cluster analysis, as an unsupervised machine learning method, plays a pivotal role in identifying phenotypic subgroups by integrating and analyzing multidimensional health data under unlabeled conditions.[73] A common strategy is to use PSG data as the core, supplemented by multi-source data (eg, anthropometric measurements, comorbidity profiles, PROs, and social environment information), then employ one or more ML algorithms for model training, and finally validate the model. Tondo et al utilized hierarchical clustering to classify 402 OSA patients into three phenotypes.[45] Ferreira-Santos et al applied k-modes to perform cluster analysis on 13 variables and then visualized their clinical features through heatmaps and radar charts.[74] These studies help provide clinicians with clearer clinical information regarding OSA patients, and serve as a preparatory step for predictive analytics.


https://doi.org/10.2147/NSS.S543091

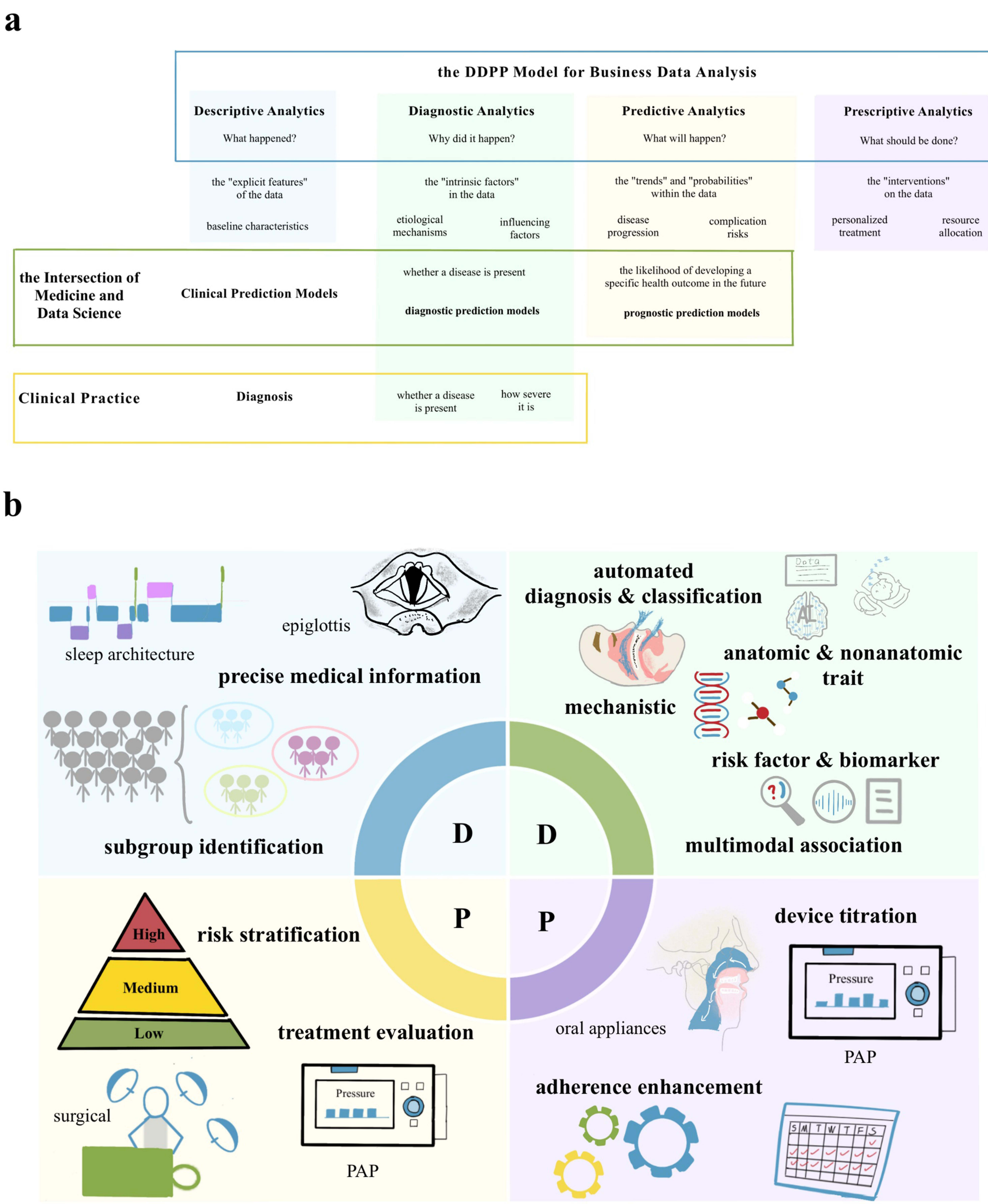



**Figure 4** Conceptual diagram of the DDPP model in OSA. (**a**) illustrates the mapping process of the DDPP framework in OSA through colored outlines and filled boxes. The blue outline shows the commercial interpretation of the DDPP model. The green outline shows the clinical prediction models at the intersection of medicine and data science. The yellow outline shows the concept of "diagnosis" in clinical practice. The blue filled box represents the scope of Descriptive Analytics in OSA, derived from answering "What happened?" to explore the "explicit features" (ie, baseline characteristics). The green filled box represents the scope of Diagnostic Analytics in OSA, which contains three parts: (1) derived from answering "Why did it happen?" to explore the "intrinsic factors" (ie, etiological mechanisms and influencing factors); (2) the diagnostic prediction models correspond to this analytics; (3) the concept of "diagnosis" corresponds to this analytics (ie, determining disease presence and severity). The yellow filled box represents the scope of Predictive Analytics in OSA, which comprises two aspects: (1) derived from answering "What will happen?" to explore the "trends" and "probabilities" (ie, disease process and the complication risks); (2) the prognostic prediction models correspond to this analytics. The purple filled box represents the scope of Prescriptive Analytics in OSA, derived from answering "What should be done?" to explore the "interventions" (ie, personalized treatment and resource allocation); (**b**) Description of what is contained in the second panel demonstrates the specific application of DDPP in the field of OSA.

## Diagnostic Analytics

Diagnostic analytics primarily answers “Why did it happen?”, emphasizing the “intrinsic factors” in the data. It is suitable for the elucidation of etiological mechanisms and the identification of influencing factors. In OSA research, the objectives of such analytics can encompass the following: (1) automated diagnosis and classification; (2) analysis of anatomic and nonanatomic traits around the PALM model;[75] (3) mechanistic exploration by mapping multi-omics data onto specific molecular pathways; (4) risk factor mining and biomarker identification based on cohort data; (5) association studies leveraging multimodal datasets. Diagnostic analytics often necessitates support from large-scale, multi-source heterogeneous data and AI technologies centered on DL.

Existing systematic reviews have shown that ML has been successfully applied to the screening and diagnosis of OSA, demonstrating good performance when utilizing easily obtained features such as ECG, pulse oximetry, and sound signals.[76] Among these ML methods, LR is the most frequently used, followed by linear regression, SVM, and neural networks.[77] Integrating DL with other AI techniques is expected to enhance performance to a greater extent. Levy et al developed a DL model named OxiNet, which uses a dual-branch network architecture for the estimation of the AHI from single-channel oximetry data. Results showed that its missed diagnosis rate for moderate-to-severe OSA was only 0.2%, which was significantly better than the best benchmark’s 21%.[46] Additionally, some scholars have searched for more innovative data sources (eg, radar signals, ultrasound images of dynamic tongue movements, infrared video data) to broaden the scope of applications.[47,53,78] These studies can help reduce missed OSA cases and provide a more efficient alternative when PSG availability is constrained.

In terms of etiological research, Cederberg et al systematically explored biomarkers associated with the presence, severity, and treatment response of OSA by analyzing data from three cohorts, which included plasma proteome (5000 proteins), PSG (AHI index), and questionnaire (anthropometric and symptom indicators). They employed a regularized LR model (a supervised machine learning approach) combined with 5-fold cross-validation.[33] The study revealed that 84 proteins were associated with the AHI, among which PAI-1, tPA, and sE-Selectin were identified as pivotal biomarkers. These findings suggest the pathophysiological mechanisms that involve endothelial dysfunction and coagulation-fibrinolysis imbalance, and offer omics evidence to elucidate the comorbidity between OSA and cardiovascular diseases.

## Predictive Analytics

Predictive analytics chiefly answers “What will happen?”, emphasizing the “trends” and “probabilities” within the data. It is applicable to assess disease progression and complication risks. Predictive analytics requires various modeling approaches, including survival analysis (eg, random survival forest, Cox proportional hazards model) and ensemble learning (eg, RFs, XGBoost). In OSA research, predictive analytics can be used for risk stratification (eg, disease susceptibility, comorbidity potential, mortality risk) and treatment evaluation (eg, surgical efficacy, PAP compliance).

Zhang et al developed and validated a metabolite index using LASSO regression based on data from two cross-race /ethnicity cohorts, that assesses an individual’s risk of moderate-to-severe OSA.[32] The research found that each 1 SD (standard deviation) of OSA metabolite index in the two cohorts was associated with a 50% and 55% increase in the risk, respectively. Eguchi et al adopted a method that combines LR and learning to rank to predict whether patients will exhibit poor adherence during the subsequent 12-week treatment period by analyzing PAP log data.[79] Tondo et al utilized Cox regression and other approaches to analyze long-term follow-up data from patients who are classified into three phenotypes, achieving mortality risk stratification and identifying key risk factors.[45] These studies assist clinicians in assessing patients’ conditions and selecting appropriate strategies for early intervention, such as enhancing management for those likely to show poor treatment adherence.

## Prescriptive Analytics

Prescriptive analytics principally answers “What should be done?”, emphasizing the “interventions” on the data. It is ideal for personalized treatment and resource allocation. Based on the degree of automation, we attempted to classify them into two types: “decision support assistance” and “intelligent closed-loop decision system”. The “decision support assistance” usually relies on descriptive analytics, diagnostic analytics, and predictive analytics to provide data-driven intervention recommendations for clinical practice. In the management of OSA, it can be reflected in the titration of medical devices (eg, PAP, oral appliances) and the enhancement of treatment adherence.[50,80] The "intelligent closed-loop decision system" largely leverages approaches such as reinforcement learning to achieve real-time decision optimization through dynamic interaction with the environment. Although neither type can replace human involvement, the latter's dynamic feature is more helpful in the treatment and management of the disease. The "intelligent closed-loop decision system" is still in the exploratory stage in the field of OSA, while its application in diabetes has been reported.[81,82]

## Limitations, Challenges and Future Directions

### Limitations

This narrative review has two main limitations. First, it does not include quantitative analysis or quality assessment, and keyword coverage may be incomplete, providing only a general overview of existing studies. Second, some less typical combinations of AI technologies and data types were not specifically emphasized. To address these limitations, we have clearly outlined the research strategy in the *Methods*, presented a mapping diagram in the *Multi-Stage Mapping via DDPP*, and provided an indication chart in the *Future Directions* to guide subsequent research.

### Challenges

BD and AI have shown great potential in the descriptive, diagnostic, predictive, and prescriptive analytics of OSA. However, a substantial amount of data remains underexplored, whether in literature resources or in various datasets of patients with comorbidities. Moreover, most AI applications in clinical practice are concentrated in Descriptive and Diagnostic phases, while Prescriptive analytics for personalized therapy is limited. These call for more in-depth and systematic research designs to fully elucidate the patterns of disease progression.

Currently, three major elements need attention to achieve precision medicine: "human", "data", and "technology".

First, human factors. Interdisciplinary collaboration between clinicians, data scientists, and engineers is indispensable. When confronting massive datasets and complex models, it is essential for experts to continue cultivating a keen sense to achieve a balance between local optima and global optima in the real world.

Second, at the data level, issues such as low quality, insufficient quantity, and privacy concerns urgently have to be addressed. Variability across device types, manufacturers, and institutional standards leads to inconsistencies in data formats, undermining data integration. In addition, sleep-related data are collected during unconscious states and may retain identifiable characteristics even after anonymization. The security of patient data across institutions necessitates

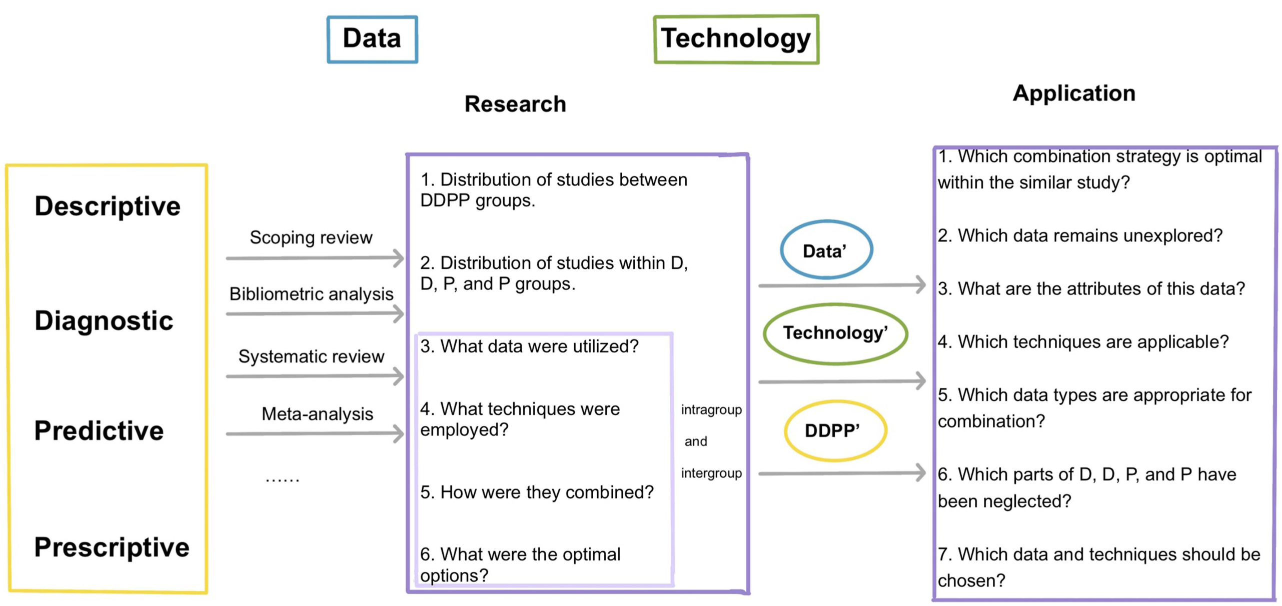



**Figure 5** Future Directions for OSA Research Integrating BD, AI, and the DDPP Framework.

strict compliance with data protection regulations and exploration of privacy-preserving techniques such as federated learning, though its clinical implementation in OSA remains rare.

Third, technology selection requires careful evaluation of data characteristics, task complexity, and resource conditions in order to balance algorithm performance and model interpretability.

### Future Directions

In future research, both forward-thinking and backward-thinking strategies can be employed based on the proposed framework, integrating data resources and AI technologies. Specifically, forward-thinking refers to systematically investigating the distribution of research content, the application of different data types, and the combinations of various data and technologies across and within DDPP stages. Conversely, backward-thinking focuses on identifying gaps such as unexplored research areas, untapped data resources, and insufficient technology utilization. These two perspectives are briefly illustrated in Figure 5. If applied in practice, this structured approach could serve as a methodological toolkit for OSA research, aiding the selection of suitable solutions when addressing clinical or scientific problems.

## Conclusion

This article identifies four major categories of data in OSA research and explores the role of four core AI components in their applications. Additionally, it introduces and optimizes a commercial analytics framework to support closed-loop analysis and enhance disease understanding. These insights provide a systematic foundation for future interdisciplinary research and data-driven AI techniques in clinical decision-making about OSA.

## Abbreviations

AHI, apnea-hypopnea index; AI, artificial intelligence; BD, big data; CBCT, cone beam CT; CNNs, convolutional neural networks; CV, computer vision; DICOM, digital imaging and communications in medicine; DISE, drug-induced sleep endoscopy; DL, deep learning; DTs, decision trees; ECG, electrocardiography; EDF+, European data format "plus"; EEG, electroencephalography; EHRs, electronic health records; EMG, electromyography; EOG, electrooculography; ESS, Epworth sleepiness scale; GNNs, graph neural networks; HSAT, home sleep apnea testing; ICD, international classification of diseases; ICF, international classification of functioning, disability and health; LLMs, large language models; LR, logistic regression; LSTM, long short-term memory; MDCT, conventional multidetector CT; ML, Machine Learning; MRI, magnetic resonance imaging; NLP, Natural Language Processing; OSA, Obstructive Sleep Apnea; PAP, positive airway pressure; PROs, patient-reported outcomes; PSG, polysomnography; PSQI, Pittsburgh Sleep Quality Index; RFs, random forests; RNNs, recurrent neural networks; SD, Standard Deviation; SVM, support vector machines; VFSS, videofluoroscopic swallowing study.

## Data Sharing Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## Author Contributions

Mengying Wu: Conceptualization, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review and editing, Validation. Kexin Wang: Formal analysis, Investigation, Methodology, Writing - review and editing, Validation. Huai Huang: Formal analysis, Investigation, Methodology, Writing - review and editing, Validation. Xiaodan Wu: Formal analysis, Methodology, Writing - review and editing, Supervision, Validation, Funding acquisition. Zilong Liu: Formal analysis, Methodology, Writing - review and editing, Supervision, Validation, Funding acquisition. Shanqun Li: Project administration, Formal analysis, Methodology, Writing - review and editing, Supervision, Validation, Funding acquisition. All authors gave final approval of the version to be published and agree to be accountable for all aspects of the work. All authors agreed on the journal for submission.

## Funding

This work was supported by the National Natural Science Foundation of China (Grant No.82370088), the National Natural Science Foundation of China (Grant No.82470089), and the National Natural Science Foundation of China (Grant No.82470090).


## Disclosure

The authors report no conflicts of interest in this work.

## References


1. Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687–698. doi:10.1016/S2213-2600(19)30198-5
2. Lv R, Liu X, Zhang Y, et al. Pathophysiological mechanisms and therapeutic approaches in obstructive sleep apnea syndrome. *Signal Transduct Target Ther*. 2023;8(1):218. doi:10.1038/s41392-023-01496-3
3. Javaheri S, Javaheri S, Somers VK, et al. Interactions of obstructive sleep apnea with the pathophysiology of cardiovascular disease, part 1: JACC state-of-the-art review. *J Am Coll Cardiol*. 2024;84(13):1208–1223. doi:10.1016/j.jacc.2024.02.059
4. Javaheri S, Javaheri S, Gozal D, et al. Treatment of OSA and its impact on cardiovascular disease, part 2: JACC state-of-the-art review. *J Am Coll Cardiol*. 2024;84(13):1224–1240. doi:10.1016/j.jacc.2024.07.024
5. Giampá SQC, Lorenzi-Filho G, Drager LF. Obstructive sleep apnea and metabolic syndrome. *Obesity*. 2023;31(4):900–911. doi:10.1002/oby.23679
6. Pase MP, Harrison S, Misialek JR, et al. Sleep architecture, obstructive sleep apnea, and cognitive function in adults. *JAMA Network Open*. 2023;6 (7):e2325152. doi:10.1001/jamanetworkopen.2023.25152
7. Gharsalli H, Harizi C, Zaouche R, et al. Prevalence of depression and anxiety in obstructive sleep apnea. *Tunis Med*. 2022;100(7):525–533.
8. Edwards C, Almeida OP, Ford AH. Obstructive sleep apnea and depression: a systematic review and meta-analysis. *Maturitas*. 2020;142:45–54. doi:10.1016/j.maturitas.2020.06.002
9. MA M-G, Campos-Rodriguez F, Barbé F, Gozal D, Agustí A. Precision medicine in obstructive sleep apnoea. *Lancet Respir Med*. 2019;7 (5):456–464. doi:10.1016/S2213-2600(19)30044-X
10. Stoel BC, Staring M, Reijnierse M, van der Helm-van Mil AHM. Deep learning in rheumatological image interpretation. *Nat Rev Rheumatol*. 2024;20(3):182–195. doi:10.1038/s41584-023-01074-5
11. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retinal Eye Res*. 2019;72:100759. doi:10.1016/j.preteyeres.2019.04.003
12. Wu X, Li W, Tu H. Big data and artificial intelligence in cancer research. *Trends Cancer*. 2024;10(2):147–160. doi:10.1016/j.trecan.2023.10.006
13. Washington P, Park N, Srivastava P, et al. Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. *Biol Psych*. 2020;5(8):759–769. doi:10.1016/j.bpsc.2019.11.015
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021:372:n71. doi:10.1136/bmj.n71
15. Chokroverty S, Bhat S. Polysomnographic recording technique. In: Thomas RJ, Bhat S, Chokroverty S, editors. *Atlas of Sleep Medicine*. Springer International Publishing; 2023:17–41.
16. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American academy of sleep medicine clinical practice guideline. *J Clin Sleep Med*. 2017;13(3):479–504. doi:10.5664/jcsm.6506
17. Kemp B, Olivan J. European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data. *Clin Neurophysiol*. 2003;114(9):1755–1761. doi:10.1016/s1388-2457(03)00123-8
18. Whyte A, Gibson D. Imaging of adult obstructive sleep apnoea. *Eur J Radiol*. 2018;102:176–187. doi:10.1016/j.ejrad.2018.03.010
19. Jugé L, Olsza I, FL K, et al. Effect of upper airway fat on tongue dilation during inspiration in awake people with obstructive sleep apnea. *Sleep*. 2021;44(12):zsab192. doi:10.1093/sleep/zsab192
20. Neelapu BC, Kharbanda OP, Sardana HK, et al. Craniofacial and upper airway morphology in adult obstructive sleep apnea patients: a systematic review and meta-analysis of cephalometric studies. *Sleep Med Rev*. 2017;31:79–90. doi:10.1016/j.smrv.2016.01.007
21. Jugé L, Knapman FL, Burke PGR, et al. Regional respiratory movement of the tongue is coordinated during wakefulness and is larger in severe obstructive sleep apnoea. *J Physiol*. 2020;598(3):581–597. doi:10.1113/JP278769
22. Viana A, Estevão D, Zhao C. The clinical application progress and potential of drug-induced sleep endoscopy in obstructive sleep apnea. *Ann Med*. 2022;54(1):2909–2920. doi:10.1080/07853890.2022.2134586
23. Valarelli LP, Corradi AMB, Grechi TH, et al. Cephalometric, muscular and swallowing changes in patients with OSAS. *J Oral Rehabil*. 2018;45 (9):692–701. doi:10.1111/joor.12666
24. Mantri M, Taran S, Sunder G. DICOM integration libraries for medical image interoperability: a technical review. *IEEE Rev Biomed Engine*. 2022;15:247–259. doi:10.1109/RBME.2020.3042642
25. Lebkuchen A, Freitas LS, Cardozo KHM, Drager LF. Advances and challenges in pursuing biomarkers for obstructive sleep apnea: implications for the cardiovascular risk. *Trend Cardiovasc Med*. 2021;31(4):242–249. doi:10.1016/j.tcm.2020.04.003
26. Wang Y, Wang X, Luo J, Qiu B, Huang R, Xiao Y. Urinary epinephrine sulfate can predict cardiovascular risk in moderate-to-severe OSA: a metabolomics-based study. *Nat Sci Sleep*. 2024;16:1153–1168. doi:10.2147/NSS.S470154
27. Baril A-A, Carrier J, Lafrenière A, et al. Biomarkers of dementia in obstructive sleep apnea. *Sleep Med Rev*. 2018;42:139–148. doi:10.1016/j.smrv.2018.08.001
28. Gaspar LS, Santos-Carvalho A, Santos B, et al. Peripheral biomarkers to diagnose obstructive sleep apnea in adults: a systematic review and meta-analysis. *Sleep Med Rev*. 2022;64:101659. doi:10.1016/j.smrv.2022.101659

29. Li W, Shao C, Zhou H, et al. Multi-omics research strategies in ischemic stroke: a multidimensional perspective. *Ageing Res Rev*. 2022;81:101730. doi:10.1016/j.arr.2022.101730
30. He X, Liu X, Zuo F, Shi H, Jing J. Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. *Semi Cancer Biol*. 2023;88:187–200. doi:10.1016/j.semcancer.2022.12.009
31. Conte L, Greco M, DM T, Arigliani M, Maffia M, Benedetto MD. A review of the "OMICS" for management of patients with obstructive sleep apnoea. *ACTA Otorhinolaryngologica Italica*. 2020;40(3):164–172. doi:10.14639/0392-100X-N0409
32. Zhang Y, Ngo D, Yu B, et al. Development and validation of a metabolite index for obstructive sleep apnea across race/ethnicities. *Sci Rep*. 2022;12 (1):21805. doi:10.1038/s41598-022-26321-9
33. Cederberg KLJ, Hanif U, Peris Sempere V, et al. Proteomic biomarkers of the apnea hypopnea index and obstructive sleep apnea: insights into the pathophysiology of presence, severity, and treatment response. *Int J Mol Sci*. 2022;23(14):7983. doi:10.3390/ijms23147983
34. Pinilla L, Benítez ID, Santamaria-Martos F, et al. Plasma profiling reveals a blood-based metabolic fingerprint of obstructive sleep apnea. *Biomed Pharmacothe*. 2022;145:112425. doi:10.1016/j.biopha.2021.112425
35. Agrawal R, Jones MB, Spiegelman AM, Bandi VD, Hirshkowitz M, Sharafkhaneh A. Presence of obstructive sleep apnea is associated with higher future readmissions and outpatient visits-a nationwide administrative dataset study. *Sleep Med*. 2022;89:60–64. doi:10.1016/j.sleep.2021.10.034
36. Alpert N, Cole KV, Dexter RB, Sterling KL, Wickwire EM. Performance of claims-based algorithms for adherence to positive airway pressure therapy in commercially insured patients with OSA. *Chest*. 2024;165(5):1228–1238. doi:10.1016/j.chest.2024.01.020
37. Streatfeild J, Smith J, Mansfield D, Pezzullo L, Hillman D. The social and economic cost of sleep disorders. *Sleep*. 2021;44(11):zsab132. doi:10.1093/sleep/zsab132
38. Liu J, Ghastine L, Um P, Rovit E, Wu T. Environmental exposures and sleep outcomes: a review of evidence, potential mechanisms, and implications. *Environ Res*. 2021;196:110406. doi:10.1016/j.envres.2020.110406
39. Weinreich G, Wessendorf TE, Pundt N, et al. Association of short-term ozone and temperature with sleep disordered breathing. *Europ resp J*. 2015;46(5):1361–1369. doi:10.1183/13993003.02255-2014
40. Russell SJ, Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. Pearson; 2020:1115.
41. Gutiérrez-Tobal GC, Álvarez D, Vaquerizo-Villar F, et al. Conventional machine learning methods applied to the automatic diagnosis of sleep apnea. In: Penzel T, Hornero R, editors. *Advances in the Diagnosis and Treatment of Sleep Apnea: Filling the Gap Between Physicians and Engineers*. Springer International Publishing; 2022. 131–146.
42. Kim D, Park JY, Song YW, Kim E, Kim S, Joo EY. Machine-learning-based classification of obstructive sleep apnea using 19-channel sleep EEG data. *Sleep Med*. 2024;124:323–330. doi:10.1016/j.sleep.2024.09.041
43. Han H, Oh J. Application of various machine learning techniques to predict obstructive sleep apnea syndrome severity. *Sci Rep*. 2023;13(1):6379. doi:10.1038/s41598-023-33170-7
44. Zhao X, Wang X, Yang T, et al. Classification of sleep apnea based on EEG sub-band signal characteristics. *Sci Rep*. 2021;11(1):5824. doi:10.1038/s41598-021-85138-0
45. Tondo P, Scioscia G, Bailly S, et al. Exploring phenotypes to improve long-term mortality risk stratification in obstructive sleep apnea through a machine learning approach: an observational cohort study. *Eur J Internal Med*. 2024;S0953-6205(24). doi:10.1016/j.ejim.2024.12.015
46. Levy J, Álvarez D, Del Campo F, Behar JA. Deep learning for obstructive sleep apnea diagnosis based on single channel oximetry. *Nat Commun*. 2023;14(1):4881. doi:10.1038/s41467-023-40604-3
47. Choi JW, Koo DL, Kim DH, et al. A novel deep learning model for obstructive sleep apnea diagnosis: hybrid CNN-transformer approach for radar-based detection of apnea-hypopnea events. *Sleep*. 2024;47(12):zsae184. doi:10.1093/sleep/zsae184
48. Korkalainen H, Leppanen T, Duce B, et al. Detailed assessment of sleep architecture with deep learning and shorter epoch-to-epoch duration reveals sleep fragmentation of patients with obstructive sleep apnea. *IEEE J Biomed Health Inform*. 2021;25(7):2567–2574. doi:10.1109/JBHI.2020.3043507
49. Su HH, Lu CP. Development of a deep learning-based epiglottis obstruction ratio calculation system. *Sensors*. 2023;23(18):7669. doi:10.3390/s23187669
50. Turino C, Benítez ID, Rafael-Palou X, et al. Management and treatment of patients with obstructive sleep apnea using an intelligent monitoring system based on machine learning aiming to improve continuous positive airway pressure treatment compliance: randomized controlled trial. *J Med Internet Res*. 2021;23(10):e24072. doi:10.2196/24072
51. Chen H, Lv T, Luo Q, et al. Reliability and accuracy of a semi-automatic segmentation protocol of the nasal cavity using cone beam computed tomography in patients with sleep apnea. *Clinical Oral Investigations*. 2023;27(11):6813–6821. doi:10.1007/s00784-023-05295-6
52. Bommineni VL, Erus G, Doshi J, et al. Automatic segmentation and quantification of upper airway anatomic risk factors for obstructive sleep apnea on unprocessed magnetic resonance images. *Acad Radiol*. 2023;30(3):421–430. doi:10.1016/j.acra.2022.04.023
53. Akbarian S, Ghahjaverestan NM, Yadollahi A, Taati B. Noncontact sleep monitoring with infrared video data to estimate sleep apnea severity and distinguish between positional and nonpositional sleep apnea: model development and experimental validation. *J Med Internet Res*. 2021;23(11):e26524. doi:10.2196/26524
54. Azarian M, Ramezani A, Sharafkhaneh A, et al. The association between all-cause mortality and obstructive sleep apnea in adults: a U-shaped curve. *Ann Am Thorac Soc*. 2025;22(4):581–590. doi:10.1513/AnnalsATS.202407-755OC
55. Walsh J, Cave J, Griffiths F. Combining topic modeling, sentiment analysis, and corpus linguistics to analyze unstructured web-based patient experience data: case study of modafinil experiences. *J Med Internet Res*. 2024;26:e54321. doi:10.2196/54321
56. Huang J, Zhuang J, Zheng H, et al. A machine learning prediction model of adult obstructive sleep apnea based on systematically evaluated common clinical biochemical indicators. *Nat Sci Sleep*. 2024;16:413–428. doi:10.2147/nss.S453794
57. Talukder A, Li Y, Yeung D, et al. OSApredictor: a tool for prediction of moderate to severe obstructive sleep apnea-hypopnea using readily available patient characteristics. *Comput Biol Med*. 2024;178:108777. doi:10.1016/j.compbiomed.2024.108777
58. Rahman MA, Jahan I, Islam M, et al. Improving sleep disorder diagnosis through optimized machine learning approaches. *IEEE Access*. 2025;13:20989–21004. doi:10.1109/ACCESS.2025.3535535
59. Mashrur FR, Islam MS, Saha DK, Islam SMR, Moni MA. SCNN: scalogram-based convolutional neural network to detect obstructive sleep apnea using single-lead electrocardiogram signals. *Comput Biol Med*. 2021;134:104532. doi:10.1016/j.compbiomed.2021.104532
60. Drzazga J, Cyganek B. An LSTM network for apnea and hypopnea episodes detection in respiratory signals. *Sensors*. 2021;21(17):5858. doi:10.3390/s21175858

61. Almutairi H, Hassan GM, Datta A. Detection of obstructive sleep apnoea by ECG signals using deep learning architectures. 2020 28th European Signal Processing Conference (EUSIPCO);2021;1382–1386;IEEE.
62. Hossan I, Dip MSS, Kabir S, Motin MA. Deep Apnea net: a multistage CNN-Bi-LSTM hybrid model for sleep apnea detection from single-lead ECG Signal. *IEEE Sens Lett*. 2025;9(5):6005804. doi:10.1109/LSENS.2025.3558675
63. Javeed A, Berglund JS, Dallora AL, Saleem MA, Anderberg P. Predictive power of XGBoost_BiLSTM model: a machine-learning approach for accurate sleep apnea detection using electronic health data. *Int J Comput Intell Syst*. 2023;16(1):188. doi:10.1007/s44196-023-00362-y
64. Shahid M, Mir J, Shaukat F, Saleem MK, Tariq M, Nouman A. Classification of pharynx from MRI using a visual analysis tool to study obstructive sleep apnea. *Curr Med Imaging*. 2021;17(5):613–622. doi:10.2174/1573405616666201118143935
65. Chowdhary KR, Chowdhary KR. Natural language processing. *Funda Artific Intellige* 2020:603–649. doi:10.1007/978-81-322-3972-7_19
66. Lipford MC, Ip W, Awasthi S, et al. Demographic characteristics and comorbidities of patients with narcolepsy: a propensity-matched cohort study. *Sleep Adv*. 2024;5(1):zpae067. doi:10.1093/sleepadvances/zpae067
67. Zhou J, Li X, Xia Q, Yu L. Innovations in otolaryngology using LLM for early detection of sleep-disordered breathing. *SLAS Technol*. 2025;32:100278. doi:10.1016/j.slast.2025.100278
68. Sun Z. Problem driven computing and analytics. *Inter J Future Computer Commun*. 2022;11:52–60. doi:10.18178/ijfcc.2022.11.3.588
69. Gupta NS, Kumar P. Perspective of artificial intelligence in healthcare data management: a journey towards precision medicine. *Comput Biol Med*. 2023;162:107051. doi:10.1016/j.compbiomed.2023.107051
70. Hung C-C, Tu M-Y, Chien T-W, Lin C-Y, Chow JC, Chou W. The model of descriptive, diagnostic, predictive, and prescriptive analytics on 100 top-cited articles of nasopharyngeal carcinoma from 2013 to 2022: bibliometric analysis. *Medicine*. 2023;102(6):e32824. doi:10.1097/MD.0000000000032824
71. Ho SY-C, Chien T-W, Tsai K-T, Chou W. Analysis of citation trends to identify articles on delirium worth reading using DDPP model with temporal heatmaps (THM): a bibliometric analysis. *Medicine*. 2023;102(8):e32955. doi:10.1097/MD.0000000000032955
72. van Smeden M, Reitsma JB, Riley RD, Collins GS, Moons KG. Clinical prediction models: diagnosis versus prognosis. *J Clin Epidemiol*. 2021;132:142–145. doi:10.1016/j.jclinepi.2021.01.009
73. Ghorvei M, Karhu T, Hietakoste S, et al. A comparative analysis of unsupervised machine-learning methods in PSG-related phenotyping. *J Sleep Res*. 2025;34(3):e14349. doi:10.1111/jsr.14349
74. Ferreira-Santos D, Rodrigues PP. Obstructive sleep apnea: a categorical cluster analysis and visualization. *Pulmonology*. 2023;29(3):207–213. doi:10.1016/j.pulmoe.2021.10.003
75. Eckert DJ, White DP, Jordan AS, Malhotra A, Wellman A. Defining phenotypic causes of obstructive sleep apnea. Identification of novel therapeutic targets. *Am J Respir Crit Care Med*. 2013;188(8):996–1004. doi:10.1164/rccm.201303-0448OC
76. Bazoukis G, Bollepalli SC, Chung CT, et al. Application of artificial intelligence in the diagnosis of sleep apnea. *J Clin Sleep Med*. 2023;19(7):1337–1363. doi:10.5664/jcsm.10532
77. Ferreira-Santos D, Amorim P, Silva Martins T, Monteiro-Soares M, Pereira Rodrigues P. Enabling early obstructive sleep apnea diagnosis with machine learning: systematic review. *J Med Internet Res*. 2022;24(9):e39452. doi:10.2196/39452
78. Ontimare Manlises C, Chen J-W, Huang -C-C. A gated recurrent unit model based on ultrasound images of dynamic tongue movement for determining the severity of obstructive sleep apnea. *Ultrasonics*. 2024;141:107320. doi:10.1016/j.ultras.2024.107320
79. Eguchi K, Yabuuchi T, Nambu M, et al. Investigation on factors related to poor CPAP adherence using machine learning: a pilot study. *Sci Rep*. 2022;12(1):19563. doi:10.1038/s41598-022-21932-8
80. Jl P, Pa C, C E, et al. Mandibular jaw movement automated analysis for oral appliance monitoring in obstructive sleep apnea: a prospective cohort study. *Ann Am Thoracic Soc*. 2024;21(5). doi:10.1513/AnnalsATS.202312-1077OC
81. Wang G, Liu X, Ying Z, et al. Optimized glycemic control of type 2 diabetes with reinforcement learning: a proof-of-concept trial. *Nature Med*. 2023;29(10):2633–2642. doi:10.1038/s41591-023-02552-9
82. Jafar A, Kobayati A, Tsoukas MA, Haidar A. Personalized insulin dosing using reinforcement learning for high-fat meals and aerobic exercises in type 1 diabetes: a proof-of-concept trial. *Nat Commun*. 2024;15(1):6585. doi:10.1038/s41467-024-50764-5